

Spatio-temporal Gait Characterization from a Wrist-Worn Inertial Measurement Unit in Daily Life

Aikaterini Athanasia - Deha Ay - Bhavana Channavajjala - Lucia Hainline - Anouck Rietveld
Columbia University Data Science Institute

In collaboration with Johnson & Johnson - Mentor: Rana Rehman

1 Abstract

This study investigates the use of wrist-worn inertial measurement units (IMUs) for spatio-temporal gait analysis, a key health indicator. While lower-back sensors provide reliable gait data, they lack practicality for continuous, daily use. Our approach integrates machine learning with advanced file processing methods, such as chunking and parallel processing, to handle extensive time-series data efficiently. Using lower-back sensor data as ground truth, we developed algorithms to classify gait patterns from wrist data, addressing non-walking movements through robust signal processing and self-supervised learning. Preliminary results achieved a classification accuracy of 50%, with weighted precision, recall, and F1 scores of 0.31, 0.50, and 0.37, respectively. These findings underscore the potential and limitations of wrist-worn IMUs for clinical gait monitoring, advancing wearable technology in health diagnostics. Next steps in this study include labeling walking segments for wrist-worn acceleration data, and using these time periods to analyze gait parameters of the participants.

2 Introduction

Gait is a crucial biomarker for assessing mobility, functional health, and overall quality of life, particularly in aging populations and those with neurological conditions. Accurate characterization of spatiotemporal gait parameters (e.g., walking speed, stride length, cadence) is essential for diagnosing mobility impairments, predicting health outcomes, and monitoring disease progression (Brand et al., 2022). Traditionally, gait analysis has relied on lower-back or lower-limb sensors, which provide high accuracy but pose practical challenges for long-term, real-world monitoring due to discomfort and adherence issues (Chan et al., 2022). In contrast, wrist-worn inertial mea-

surement units (IMUs), such as those embedded in smartwatches, offer a more convenient and widely accepted alternative, with potential for continuous monitoring in daily life (Brand et al., 2022) (Soltani et al., 2020).

However, extracting reliable spatio-temporal gait parameters from wrist-worn sensors is non-trivial due to the complex and varied movements of the arms during non-walking activities, which can degrade the accuracy of gait detection algorithms (Brand et al., 2022) (Soltani et al., 2020). This challenge has spurred the development of advanced machine learning and signal processing algorithms aimed at filtering out irrelevant wrist movements and improving gait detection (Brand et al., 2022). For example, previous studies have shown that combining self-supervised learning with deep learning models like convolutional neural networks (CNNs) can enhance the accuracy of gait detection in older adults (Brand et al., 2024). Further, another study by Yuan et al. (2024) leveraged self-supervised learning to exploit the UK Biobank accelerometer dataset and build models that can obtain a highly competitive activity recognition model. Additionally, autocorrelation and windowed-peak detection algorithms have proven successful in measuring step count from a wrist-worn tri-axial accelerometer signal for walking when compared to step count measured from video of participants walking (Femiano et al., 2022). Despite these advancements, few studies have thoroughly evaluated the spatio-temporal parameters extracted from wrist sensors against the established “gold-standard” of lower-back sensor data to characterize gait patterns.

The primary objective of this study is to develop and validate algorithms that accurately characterize spatio-temporal gait parameters using wrist-worn IMU data. To address the challenges associated with wrist sensor data, we will apply robust signal processing techniques and implement ma-

chine learning models. The study will begin by labeling the data from the lower back sensors with walking versus non-walking, which will serve as the ‘ground truth.’ By syncing this data with the wrist-worn data, key gait features will then be extracted from wrist sensor data and compared to the ground truth from the lower back sensors. Machine learning models will be employed to predict spatio-temporal gait parameters. By assessing the accuracy, reliability, and feasibility of wrist-worn IMUs for gait analysis in everyday life, this study aims to expand the use of wearable sensors in clinical gait assessments and remote participant monitoring.

3 Related Literature

3.1 Gait Detection

Recent studies on gait detection using wearable sensors, particularly wrist-worn devices, have employed a range of advanced methodologies to address the challenges of accurately monitoring gait in real-world environments. Traditional methods often use lower back sensors due to their proximity to the body’s center of mass, which provides reliable data for detecting walking patterns. However, wrist-worn sensors have gained attention for their practicality and high compliance in long-term monitoring applications. For instance, a study by Brand et al. (2022) applied a deep convolutional neural network (DCNN) for gait detection, utilizing labeled data from lower back sensors as a reference to train and validate its performance on wrist-worn data, demonstrating high precision in daily living scenarios. While many wrist-based algorithms focus primarily on healthy, younger adults, Brand et al. (2024) also implemented an anomaly detection algorithm to identify gait patterns from wrist-worn 3D accelerometer recordings in older adults, showing that daily living gait can be effectively quantified using wrist sensors. Additionally, ElderNet, a self-supervised learning (SSL) model, was developed to pre-train on large unlabelled datasets and then fine-tune on smaller, labeled datasets specific to older adults and individuals with Parkinson’s disease. This two-stage process improves model robustness and generalizability by addressing the limited availability of labeled gait data. Further, Kluge et al. (2024) conducted a validation study comparing gait detection algorithms developed for wrist-worn sensors against real-world reference data, demon-

strating that these algorithms can achieve high performance in detecting gait sequences in real-life settings. Collectively, these methodologies highlight the growing sophistication in gait detection, leveraging deep learning, self-supervised models, and anomaly detection to enhance the accuracy and adaptability of wrist-worn sensor applications.

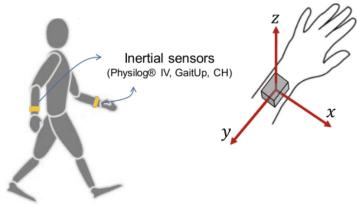


Figure 1: Wrist-worn tri-axial accelerometer (Soltani et al., 2020)

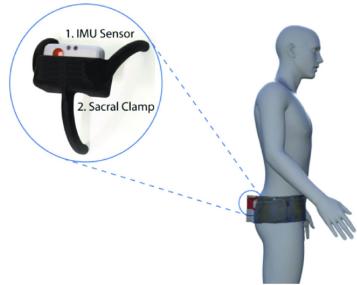
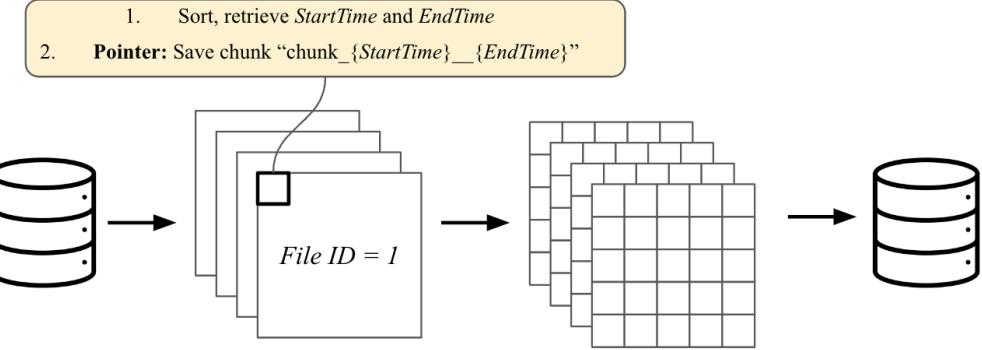


Figure 2: Lower back-worn tri-axial accelerometer (Vayalapra et al., 2022)

3.2 Gait Analysis

The reviewed literature on gait analysis using wrist-worn accelerometers shows a variety of advanced methodologies aimed at improving accuracy and applicability in real-world settings. Soltani et al. (2020) proposed a personalized approach to gait speed estimation, incorporating individual gait characteristics through online learning to enhance performance using wrist sensors. Their model fused biomechanically derived features with GNSS data, achieving high accuracy in both walking and running conditions. Femiano et al. (2022) validated two open-source algorithms—windowed peak detection and autocorrelation—specifically for cardiovascular participants, achieving high accuracy in controlled environments with errors below 10% for key movement categories. Pilkar et al. (2022) explored multiple wrist-based step-counting algorithms, including machine learning approaches, finding that movement frequency detection and machine learning models outperformed traditional algorithms

Generating Pointers and Chunks



Processing Chunks in Parallel

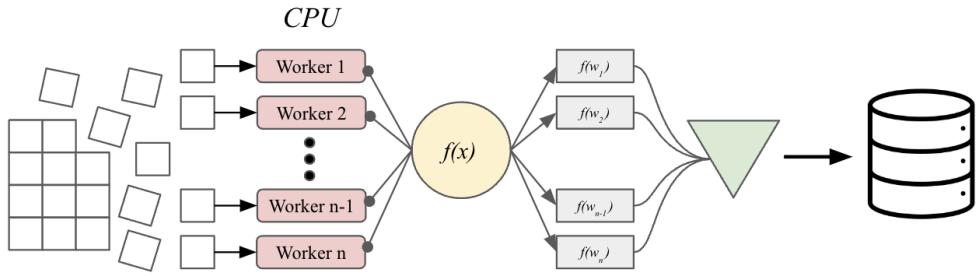


Figure 3: Demonstration of the Chunk Pointer Method and Processing

such as peak detection, particularly during varied walking speeds and patterns. Finally, Small et al. (2023) developed a machine learning-based step detection algorithm validated using the UK Biobank data. The self-supervised model showed high precision, demonstrating the feasibility of accurate step detection in large-scale, free-living environments. These studies collectively illustrate the shift towards personalized and adaptive algorithms to enhance the accuracy of gait analysis using wearable wrist devices.

4 Data Overview

4.1 Data Collection

Our dataset consists of sensor data collected from 28 participants, aged 21 to 80 years old, each equipped with two sensor devices placed on the wrist and the lower back (lumbar region). A McRoberts Dynaport device was worn on the lower back with a waist strap, and an Axivity AX6 6-Axis Logging Accelerometer was worn on the wrist as a watch. The lumbar sensor provides three-dimensional data from an accelerometer (in g), a gyroscope (in degrees per second), and a magnetometer (unknown units), while the wrist

sensor, in contrast, records three-dimensional accelerometer and gyroscope data. All sensor data streams are timestamped, allowing precise alignment and temporal analysis.

4.2 Missing Data

To further understand data distribution and detect patterns of missing or anomalous data, we generated heatmaps illustrating the percentage of zero values across different sensor axes and modalities for each participant. Accounting for zero-value percentages is essential, as zero measurements in the gyroscope and magnetometer data indicate that, although data recording is active, the sensor is likely not being worn.

Based on those figures, we observe that some participants have worse data quality than others and that not all participants have complete wrist data across all sensor types; in some cases, only accelerometer data is available.

The recordings sampled at a predominant frequency of 100 Hz. However, a subset of files exhibits lower sampling rates (35 Hz or 50 Hz) due to sampling errors. This discrepancy may affect the temporal precision of gait analysis in these cases, thus those recordings will not be considered when

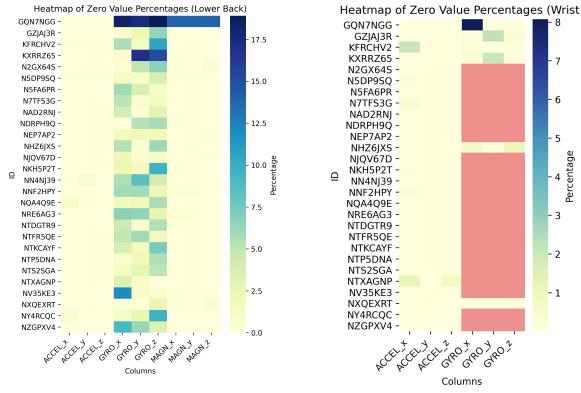


Figure 4: Zero values percentage for lumbar data

Figure 5: Zero values percentage for wrist data. Red represents missing data.

training.

4.3 Demographic Characteristics of the Study Cohort

The study's cohort consists of a diverse group, with a gender distribution of 5 male and 18 female participants. The age ranges from 21 to 80 years, with a mean age of 57.7 years. The participants' body weight varies from 50 kg to 113 kg, with an average weight of 75.1 kg, while the height of the participants ranges from 152 cm to 185 cm, with a mean height of 165.5 cm. The distribution of these characteristics ensures a varied representation of older participants, allowing for the exploration of spatio-temporal gait patterns across different body types and demographics.

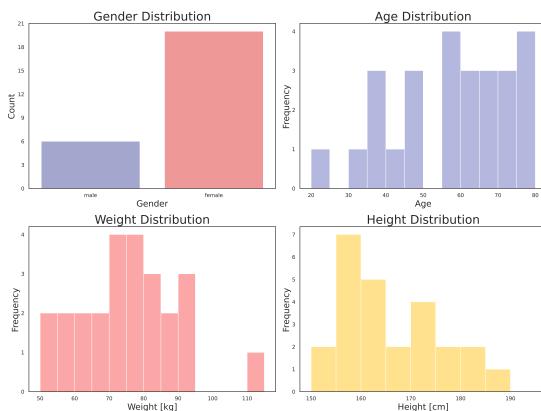


Figure 6: Acceleration signal from accelerometer in lower back

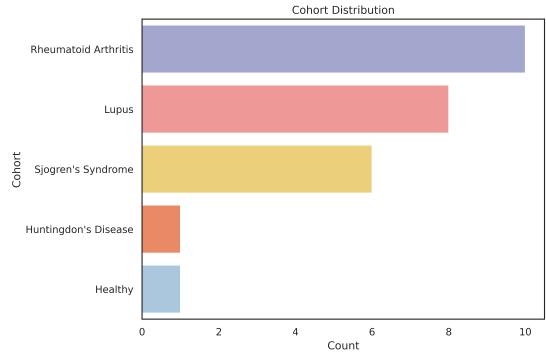


Figure 7: Acceleration signal from accelerometer in lower back

4.4 Wrist Recordings

Statistic	Value
Total Duration	379 days, 15.7 hours
Average Duration per File	13 days
Total Number of Rows	3.034 billion
Average Rows per File	108.36 million
Recording Frequencies:	
100 Hz	25 files
50 Hz	2 files
35 Hz	1 file

4.5 Lower Back Data And Wrist Data

Statistic	Value
Total Duration (Combined)	2 years, 29 days
Total Number of Rows (Combined)	6.07 billion rows

4.6 Time Overlap Analysis

To determine the feasibility of using participant data, an initial “Time Overlap Analysis” was conducted. This involved:

1. Visualizing and verifying time overlaps between *omx* and wrist recordings.
2. Categorizing participants into three groups:
 - Those with perfectly aligned data, immediately usable.
 - Those requiring further processing using the Mobilize-D algorithm.
 - Those with insufficient alignment, excluded from the project.

This analysis also identified data redundancy due to the extensive volume of pre-processed and labeled files, particularly in the mapped data. By quantifying temporal alignment among wrist, back, and classification data (*combined_classification_df.csv*), we

streamlined the processing pipeline, reducing redundant operations and improving efficiency.

Although pre-labeled data existed in classification files within participant folders, we prioritized accuracy by reprocessing raw files and merging them with classification data. Figures 8 and 9 highlight contrasting cases:

- **Figure 8:** Perfect alignment between the classification file (`combined_classification_df.csv`), AX6 wrist recording (`combined_ax6_df.csv`), and mapped OMX recordings, making this participant suitable for model training.
- **Figure 9:** Misalignment between AX6 wrist recordings and both classification and OMX data, excluding this participant due to inconsistent time alignment.

This structured approach ensured precise labeling, minimized redundancies, and optimized data utilization for model training.

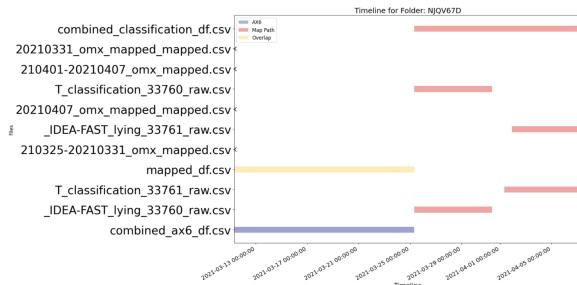


Figure 8: Timeline for Folder N2GX64S: aligned data sources.

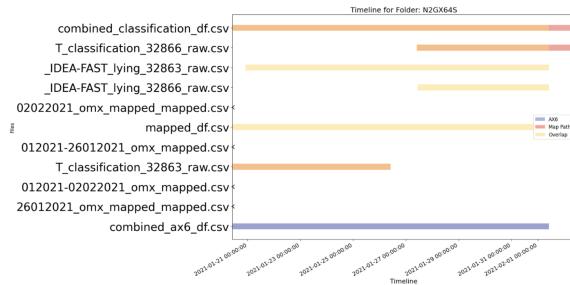


Figure 9: Timeline for Folder NJQV67D: misaligned data sources.

After conducting the time overlap analysis, we identified six participants with over 97% overlap between wrist and lower back data, who could be labeled using the provided classification files immediately. Additionally, labeling was possible for

eight participants using the Mobilize-D classification algorithm, despite varying levels of overlap between wrist and lower back recordings.

The remaining participants were excluded from the classification portion due to one or more of the following reasons: lack of metadata required for Mobilize-D processing, insufficient overlap between wrist and lower back data, or the absence of classification files. In total, 14 participants were included in the study for the classification process.

4.7 Signal Visualization

To better understand typical gait patterns, we visualized accelerometer data from a randomly selected participant during both walking and non-walking instances, captured simultaneously from the lower back and wrist.

The accelerometer signals from the lower back and wrist demonstrate distinct characteristics for walking and non-walking activities. As shown in Figure 10 and Figure 11, the signals during walking exhibit noticeable differences between the wrist and lower back. The wrist signals (Figure 10) display higher variability and sharper peaks, reflecting the arm's dynamic motion, while the lower back signals (Figure 11) are smoother and more periodic, capturing the body's core movements. Similarly, during non-walking activities, the wrist signals (Figure 12) remain highly irregular, influenced by incidental arm movements, whereas the lower back signals (Figure 13) show low amplitude and stability. These differences underscore the impact of sensor placement on activity classification and gait characterization.

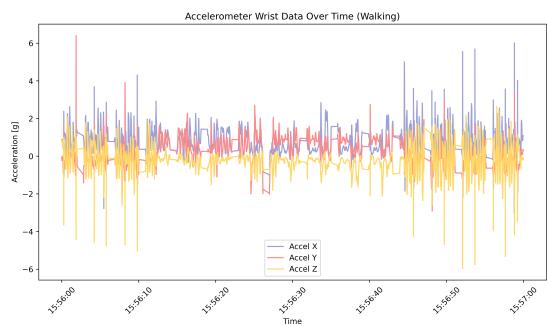


Figure 10: Acceleration signal from accelerometer in lower back

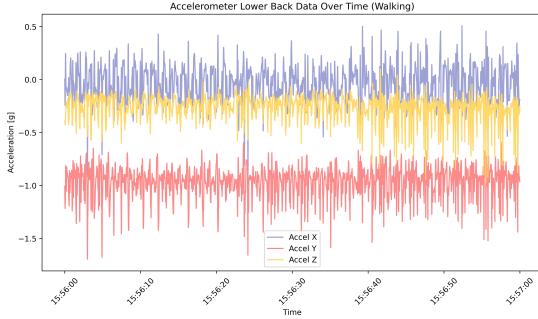


Figure 11: Acceleration signal from accelerometer in lower back

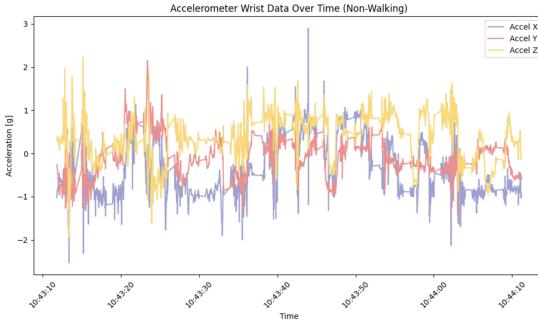


Figure 12: Acceleration signal from accelerometer in lower back

5 Data Preprocessing

5.1 File Processing and Data Handling

The dataset used in this study contained extensive time series data from 28 participants, including both wrist (.ax6) and lower back (.omx) sensor recordings. Each participant’s data consisted of:

- **Wrist recordings:** Converted from .ax6 to .CSV.
- **Lower back recordings:** Converted from .omx to .CSV.

Given the sheer size of these files, attempting to load them into memory all at once was infeasible and would crash the computing environment.

Before proceeding with complex manipulations, we developed an **automated big file processing pipeline** (Figure 3) designed to handle large datasets in a fast and efficient manner. This system was specifically optimized to work seamlessly with parallel processing, enabling high-speed data manipulation and information retrieval across the overall data chunks. Use cases included extracting essential metadata (e.g., number of rows, timestamp ranges, sensor frequencies, participant attributes such as gender and age) and

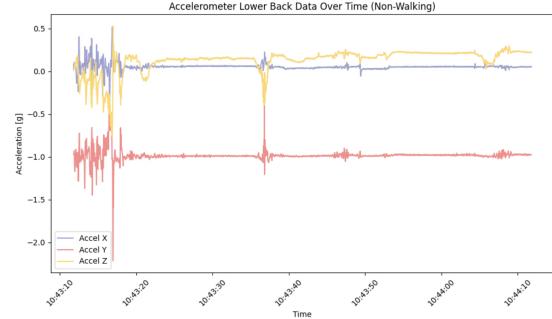


Figure 13: Acceleration signal from accelerometer in lower back

performing time overlap analyses among wrist, lower back, and pre-labeled classification files to identify redundant recordings. Its design ensured smooth integration with **our chunking strategy**, leveraging parallel processing to achieve scalable and efficient handling of the dataset.

5.1.1 Chunking and Parallel Processing Strategy

To handle the large datasets effectively, we adopted a chunking strategy combined with parallel processing. Instead of depending on libraries like Dask or PyArrow, we introduced a custom ”pointer” approach (refer to Algorithm 1):

Chunked Input: Using the pandas chunk parameter, we split the data into manageable segments. Each segment was assigned a name reflecting its timestamp range.

Direct Access via Timestamps: Rather than maintaining a monolithic file, smaller chunks allowed direct file access by timestamp references, streamlining data retrieval.

Better Runtime Estimation and Debugging: Pandas’ native chunking eased debugging, supported runtime estimation (via `tqdm`), and improved error handling—capabilities that were not as conveniently accessible with Dask or PyArrow.

To understand the structure and scale of the data before processing, we developed an automated file information analyzer. This script extracted basic file-level details such as the number of rows, timestamp ranges, sensor frequency, and participant metadata like gender and age. It also performed a time overlap analysis to identify overlaps between wrist, lower back, and pre-labeled classi-

Algorithm 1: File Generation and Processing with Specific Access

1. Generating Files:

```

file_path, chunk_size: begin
    while not EOF(file_path) do
        Read and process each chunk chunk  $\leftarrow$  Read next
        chunk of size chunk_size
        sorted_chunk  $\leftarrow$  Sort chunk by timestamp
        Get time range for the chunk min_time  $\leftarrow$  minimum
        timestamp in sorted_chunk
        max_time  $\leftarrow$  maximum timestamp in sorted_chunk
        Save the chunk with time-based naming Save
        sorted_chunk as:
        chunk_{min_time}-{max_time}.feather
    
```

2. Processing Files:

```

chunk_dir, process_function, max_parallel: begin
    paths  $\leftarrow$  Retrieve all Feather file paths in chunk_dir
    Worker(chunk_path), paths, max_workers = max_parallel
    chunk_path: begin
        Process a single chunk chunk  $\leftarrow$  Read Feather file at
        chunk_path
        Apply process_function to chunk
    
```

3. Access Specific Data by Time Range:

```

chunk_dir, time_range: begin
    Identify chunks within the specified time range
    matching_chunks  $\leftarrow$  Filter Feather file paths in chunk_dir
    where:
        min_time  $\leq$  time_range.start  $\wedge$  max_time  $\geq$ 
        time_range.end
    Load and process only the matching chunks foreach
    chunk_path in matching_chunks do
        chunk  $\leftarrow$  Read Feather file at chunk_path
        Return or Process chunk as required
    
```

fication documents. The overlap analysis helped us understand the extent of redundant data within the dataset due to repeated recordings.

6 Methodology

The larger goal of this project is to label wrist-worn sensor data, identify periods of walking and extract gait characteristics such as: step count and cadence.

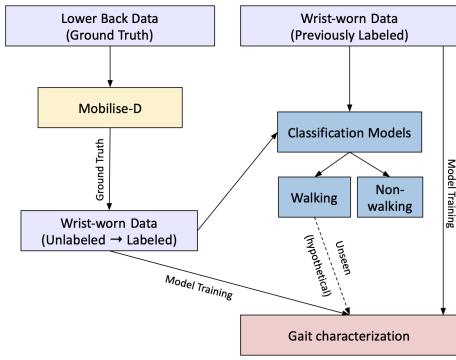


Figure 14: Methodology Workflow

6.1 Leveraging Lumbar Data

A significant portion of the wrist data provided to us was unlabeled which posed a challenge for ac-

tivity classification and gait characterization since at first only 6 out of 28 participants' data were usable.

To address this, we leveraged state-of-the-art methods for walking bout detection and gait characterization using lumbar (lower back) IMU data. These tasks are generally simpler to solve on lumbar data due to the reduced degrees of freedom in movement at the lower back. Consequently, this data benefits from highly accurate algorithms, making it an effective intermediary solution. By synchronizing the wrist data with the lower back data using timestamps, we were able to derive labels for the wrist data.

As part of this process, we deployed algorithmic pipelines from the Mobilise-D (Kirk et al., 2024; Micó-Amigo et al., 2023) package on the labeled lower back data. The way the pipeline works is by directly using the output from one task as the input for the next, following a specific schema, which is shown in Figure 15. The algorithms of the pipeline utilizes both accelerometer data and participant metadata to facilitate accurate analysis. This pipeline approach allowed us to validate the most effective algorithms on the labeled lower back data for each task in the sequence, ensuring their suitability for our dataset, before running them on the unlabeled data.

Before utilizing the pipelines, we preprocessed the data to meet the specifications outlined by Mobilise-D. This included not only converting the acceleration signal from gravitational units (g) to SI units (m/s^2), but also performing several additional tasks, primarily related to data formatting. These preprocessing steps were essential to ensure the data's compatibility with the pipeline requirements. To further optimize efficiency, we processed the preprocessed data in parallel, running the pipelines on multiple data chunks simultaneously, which significantly reduced the overall processing time.

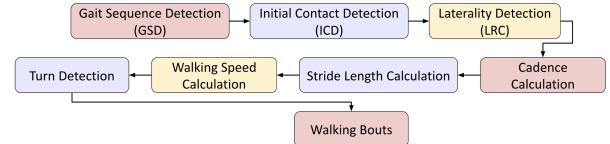


Figure 15: Mobilise-D Pipeline

Table 1 displays the performance metrics for the different configurations tested. Due to the structure of the pipeline, we did not need to test every

possible combination of algorithms for each step. Instead, we optimized the configuration by selecting the best algorithm for the first step, followed by the best algorithm for the second step, and so on.

	Cfg1	Cfg2	Cfg3	Cfg4	Cfg5	Cfg6
GSD	Iluz	Adaptive Ionescu	Ionescu	Ionescu	Ionescu	Ionescu
ICD	Shin	Shin	Shin	Ionescu	HKLee	Shin
LRC	Ullrich	Ullrich	Ullrich	Ullrich	Ullrich	McCamley
Acc.	0.966	0.941	0.975	0.976	0.976	0.975
F1	0.000	0.527	0.719	0.707	0.707	0.719
Prec.	0.000	0.367	0.603	0.624	0.624	0.603
Recall	0.000	0.926	0.891	0.815	0.815	0.891

Table 1: Performance of different configurations

Based on the results shown in Table 1, the optimal configuration, considering all metrics, is Cfg3. However, the results also reveal a recurring issue with false positives across all configurations.

After determining the optimal pipeline, we applied it to generate files containing walking bout detection start and end timestamps, along with gait parameters for each time interval. By mapping the timestamps, we were able to generate labels for 8 additional previously unlabeled participants. However, we were unable to label data for all participants due to missing demographic information for some or poor alignment between the wrist and lumbar data, which made labeling for these participants impractical.

6.2 Activity Classification

We developed a custom model to compare its performance with a pre-trained model, HarNet, which is known to be successful for activity classification tasks. This dual approach aimed to evaluate the effectiveness of our novel method while gaining insights into the elements that contribute significantly to classifying activities such as walking and non-walking. While the specifics of HarNet will be addressed in later sections, this section focuses on our novel approach and its foundational development.

6.2.1 Initial Approach and Preprocessing for Machine Learning

Our initial exploration began with visualizing the raw accelerometer data to understand its distribution and separability between walking and non-walking activities. One key observation was the presence of a direct current (DC) component in

the sensor signals, primarily due to the sensors being positioned at different heights on the body. This discrepancy necessitated zero-centering the signals to eliminate biases and align the data more effectively.

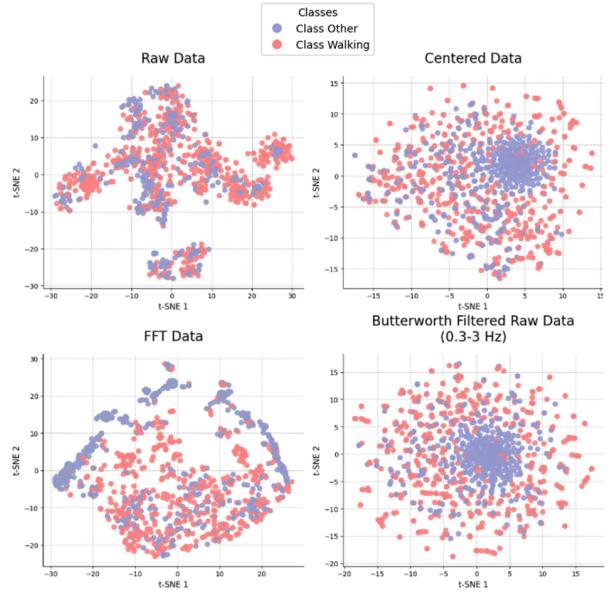


Figure 16: t-SNE Visualization of Accelerometer Data with Various Normalization Representations Highlighting Activity Separability

To zero-center the data, we subtracted the mean value from each axis individually. Post zero-centering, the separation between walking and non-walking activities became more evident, as confirmed through visualization. Recognizing the importance of preserving the variance relationship across axes, we proceeded to normalize the data by dividing each axis by its respective standard deviation. This step ensured the integrity of variance while maintaining consistency across all dimensions.

Subsequent analysis using T-SNE (Figure 17) showed significant improvement in the separability of the data after these preprocessing steps. The clusters for walking and non-walking activities became more distinct, highlighting the importance of zero-centering and variance normalization in preparing the data for activity classification.

6.2.2 Data Splitting and Windowing

The nature of time-series accelerometer data, particularly in scientific applications, necessitates careful consideration of the windowing strategy. To preserve the temporal continuity and capture meaningful patterns, we adopted a sliding window

approach with a half-stride overlap of the original time window. This method ensures comprehensive coverage of the data, capturing transitions between activities more effectively while maintaining computational efficiency.

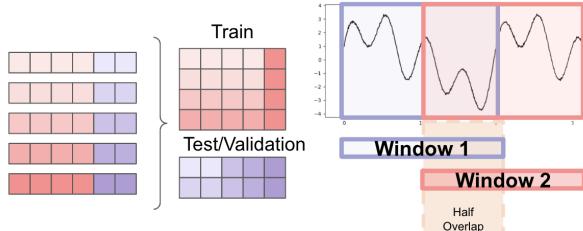


Figure 17: Data Splitting and Time Windowing

A critical challenge in handling such data is the risk of data leakage, which can compromise the validity of the classification results. To address this issue, we ensured that the training, validation, and test sets were split prior to windowing, thereby avoiding any overlap of windows across these datasets. Specifically, the dataset was divided into 80% training and 20% testing and validation sets, with the latter further split into a 60:40 ratio for validation and testing, respectively. This stringent separation protocol was designed to ensure robust model evaluation, avoiding artificially inflated performance metrics due to data leakage.

6.2.3 Novel Approach: Wavelet Model

One of the driving motivations for the Novel Approach was the need to utilize the data to its maximum capacity. Due to the large size of the HarNet model and GPU constraints, memory allocation for the entirety of our dataset was not feasible, often resulting in kernel crashes. This necessitated the creation of a model optimized for quick training, leveraging the substantial CPU power available. The approach enabled efficient handling of the large dataset provided by the CAPTURE24 paper, **which included approximately 27 hours of data from 151 additional participants**. By adopting this strategy, we effectively utilized our computational resources while ensuring the model's scalability and effectiveness for activity classification.

After data representations to detect and understand the nature of the frequencies that are helping our classification, we converted raw signal data into FFT, using only the positive frequencies to perform feature selection. To identify the most

significant frequencies, we utilized a strong L1 regularization technique integrated within a Sequential Neural Network model (Figure 19). This allowed us to effectively isolate and emphasize the frequencies contributing the most to classification accuracy.

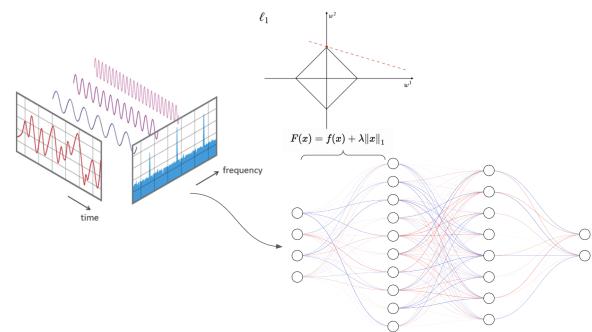


Figure 18: Illustration of FFT Model Training For Understanding Frequency Contributions

This analysis informed the subsequent architecture of our custom model, ensuring that it was tailored to leverage these key frequency features. Additionally, it highlighted which frequency information would be vital for future analysis, leading to the use of a Butterworth filter and enhancing the model's complexity.

6.2.4 Expanding Feature Space: Statistical and Frequency-Domain Analysis

Building upon the promising results obtained using FFT-transformed data, we sought to further enrich our feature space by extracting a wide array of statistical, distributional, and time-frequency-based descriptors from the raw and preprocessed signals. Initially, we employed only the raw FFT magnitudes across the full frequency spectrum to assess whether sufficient classification accuracy could be achieved using this straightforward approach. While this method demonstrated promising separability, we hypothesized that incorporating additional features capturing subtle temporal and spatial characteristics could further enhance model performance.

By exploring numerous statistical measures, autocorrelation structures, wavelet decompositions, and higher-order derivatives such as jerk and snap, we created a more comprehensive feature set to better distinguish between walking and non-walking activities.

Wavelet Transform and Motion Dynamics: To analyze motion dynamics, we derived *jerk* and

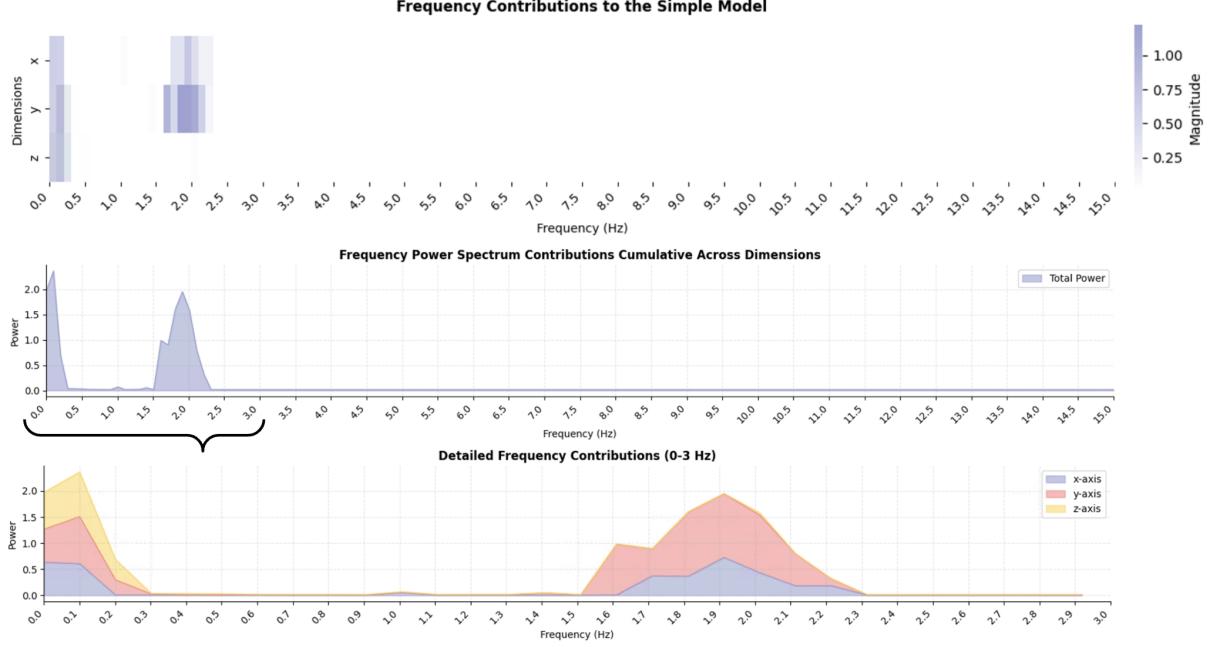


Figure 19: Frequency Contributions to the Simple Model. The top panel shows the distribution of magnitudes across dimensions, while the middle panel displays the cumulative power spectrum contributions. The bottom panel highlights the detailed frequency contributions within the range of 0-3 Hz, emphasizing the x-, y-, and z-axis contributions.

snap from accelerometer data, representing the first and second derivatives of acceleration, respectively. These derivatives highlight transient patterns and abrupt transitions in activities such as walking and running.

Jerk and Snap: Acceleration data, $\mathbf{a}(t) = [a_x(t), a_y(t), a_z(t)]$, represents velocity changes over time across three spatial axes. The *jerk*, $\mathbf{j}(t)$, and its magnitude are defined as:

$$\mathbf{j}(t) = \frac{d\mathbf{a}(t)}{dt}, \quad |\mathbf{j}(t)| = \sqrt{\sum_{i=x,y,z} \left(\frac{da_i(t)}{dt} \right)^2}.$$

Similarly, the *snap*, $\mathbf{s}(t)$, captures the rate of change of jerk:

$$\mathbf{s}(t) = \frac{d^2\mathbf{a}(t)}{dt^2}, \quad |\mathbf{s}(t)| = \sqrt{\sum_{i=x,y,z} \left(\frac{d^2a_i(t)}{dt^2} \right)^2}.$$

These metrics provide time-domain insights but may miss finer patterns.

Wavelet Transform: To enhance the analysis, we applied the *Wavelet Transform (WT)*, which decomposes signals into localized time-frequency components. Unlike the Fourier Transform, which averages frequency over the entire signal, WT captures variations across different scales and times.

Mathematically, WT computes coefficients, $W(s, \tau)$, by convolving the signal, $f(t)$, with a scaled and shifted wavelet, ψ :

$$W(s, \tau) = \int f(t)\psi^* \left(\frac{t-\tau}{s} \right) dt,$$

where s controls scale (frequency), and τ controls time localization.

This decomposition reveals multi-resolution patterns critical for identifying distinct activities.

The visualization below shows the decomposition of accelerometer data into approximation and detail coefficients using the Daubechies-4 wavelet (db4).

Feature Extraction: The feature extraction process involved calculating a diverse set of descriptors, including:

- **Basic Statistical Features:** Means, standard deviations, root mean squares (RMS), maxima, minima, and ranges for both acceleration, jerk, and snap signals, as well as their magnitudes. These values characterize the central tendency and variability of the signals.
- **Frequency-Domain Features:** Computation of FFT-based measures, including spectral centroid, bandwidth, skewness, kurtosis, and

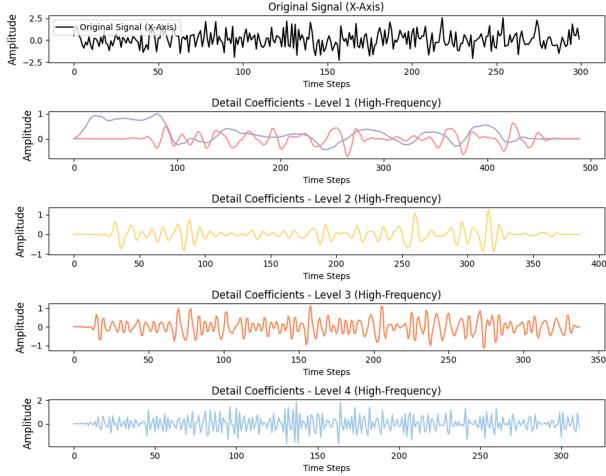


Figure 20: Wavelet decomposition of accelerometer data. The approximation coefficients (low-frequency) capture smooth trends, while detail coefficients (high-frequency) reveal abrupt changes and noise.

energy concentration in different frequency bands (e.g., 0 – 3 Hz). These features were derived not only from raw acceleration but also from jerk and snap signals to capture multi-scale dynamics.

- **Cross-Correlation and Autocorrelation Features:** Quantifying temporal and inter-axis relationships allowed us to detect synchronized patterns or lagged dependencies within and across the x , y , and z axes.
- **Distributional and Zero-Crossing Metrics:** Percentiles, interquartile ranges, zero and mean crossing rates, and entropy measures provide insights into the statistical distribution and complexity of the signal.
- **Wavelet Decomposition:** By applying a discrete wavelet transform (e.g., using the ‘db4’ wavelet), we captured both time-localized frequency content and transient features that can complement the FFT-based analysis.

Initial Approach: FFT in Three Axes The initial approach employed a straightforward representation of the data by concatenating the raw FFT magnitudes across the three spatial axes (x , y , and z). This method leveraged frequency-domain representations to distinguish between walking and non-walking activities. While the FFT approach effectively captured dominant periodic patterns and provided reasonable separability, it lacked the

ability to account for finer temporal and spatial nuances within the signal. This limitation motivated the development of a more comprehensive feature extraction methodology.

Expanding Feature Space: Statistical and Frequency-Domain Analysis Building upon the initial FFT-based approach, the new method sought to enhance the feature space by extracting a diverse array of statistical and frequency-domain descriptors from various signal representations. In addition to raw acceleration signals, we computed features from derived signals such as FFT, jerk, magnitude, and snap. This expansion aimed to capture subtler patterns and temporal dynamics that were overlooked in the initial FFT-only approach.

Rationale for Expanded Feature Extraction

The initial approach, utilizing raw FFT magnitudes concatenated across three axes, demonstrated that frequency-domain representations could differentiate between walking and non-walking activities. However, its simplicity limited the ability to capture finer temporal and spatial patterns that could improve classification performance. By incorporating features such as jerk, snap, and wavelet-based descriptors, we aimed to:

- Capture transient and abrupt changes in motion, reflected in jerk and snap derivatives.
- Exploit both global (FFT) and local (wavelet) frequency characteristics.
- Enrich the statistical representation of signals with features such as variance, skewness, and kurtosis across multiple domains (raw, FFT, jerk, magnitude, and snap).
- Enhance the robustness of the model to noise and variability by integrating multi-scale and multi-domain information.

The expectation was that combining this broader spectrum of features would provide richer insights and significantly boost classification performance, addressing the limitations observed in the initial FFT-only approach.

Feature Selection and Analysis: Neural Network with L1 Regularization Utilizing the combination of these enriched features, we extracted a total of 310 features spanning various domains, including raw signals, FFT-transformed

data, jerk, magnitude, snap, and wavelet representations. To identify which features contributed the most to classification accuracy, we applied the previously designed neural network model augmented with L1 regularization. This approach enabled us to penalize less informative features, effectively filtering the feature set and isolating those with the highest relevance for distinguishing between walking and non-walking activities.

To visualize the importance of these features, we generated the following visualizations:

- **A Sorted Heatmap of Feature Importance**, using a magma color scale, illustrating the relative weights assigned to each feature by the neural network model.
- **A Cumulative Feature Importance Plot**, highlighting the cumulative contribution of features in descending order of importance. This plot illustrates how many features are necessary to capture a significant portion of the model’s performance.
- **A Feature Contribution Analysis Figure**, categorizing features by data type (raw, FFT, jerk, etc.), dimension (x , y , z , or magnitude), and extraction method (e.g., statistical, wavelet, etc.). This visualization highlights which signal representations and methodologies were most impactful in driving classification performance.

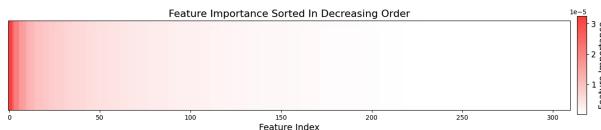


Figure 21: Sorted Heatmap of Feature Importance. Colors represent feature weights, with darker shades indicating higher importance .

Feature Importance Analysis and Final Feature Pool: The heatmap (Figure 21) and cumulative feature importance plot (Figure 23) reveal that a small subset of features accounted for a significant portion of the model’s performance. By systematically adding and removing features, we observed notable changes in the AUC and loss metrics (Figure 24).

The most efficient configuration, balancing accuracy with minimal feature usage, was achieved through the integration of wavelet decomposition

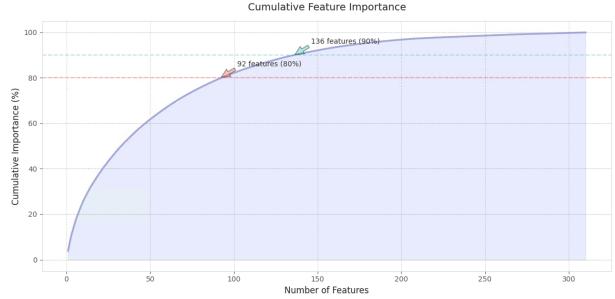


Figure 22: Cumulative Feature Importance Plot. The x-axis represents the number of features, and the y-axis represents cumulative importance.

with statistical measures, including percentiles, entropy, log energy, peak-to-peak values, and basic statistics across various coefficient levels. Notably, removing spectral features alone had a minimal impact on AUC (-0.00059), whereas excluding wavelet features resulted in a more significant decrease (-0.00515). The combined usage of wavelet and spectral features showed a balanced performance (AUC: 0.88937). A deeper dive into FFT-only configurations (AUC: 0.86487) highlighted potential limitations when frequency-domain features were isolated, as the model misclassified standing activities more frequently. This underscores the need for a multi-faceted approach combining wavelet, raw, and jerk-derived features for optimal classification.

The best and most stable method for training extracted tabular data was determined to be XGBoost. This decision was based on its ease of use, computational efficiency, and robustness across various feature configurations and time windows. Figure 25 illustrates the performance of XGBoost as a function of max_depth for different prediction time windows (10, 5, and 30 seconds). The plot compares the average score, computed as the mean of precision, recall, and F1-score, across various configurations.

From this analysis, it was evident that a max_depth of 6 offered an optimal balance between model complexity and performance stability, as highlighted by the vertical dashed line in Figure 25. This configuration minimizes the runtime and prediction time without significantly compromising accuracy, making it a robust choice for tabular data classification tasks.

6.2.5 HarNet

In this section, we delve into the pre-trained HarNet models, which were introduced by Yuan et al.

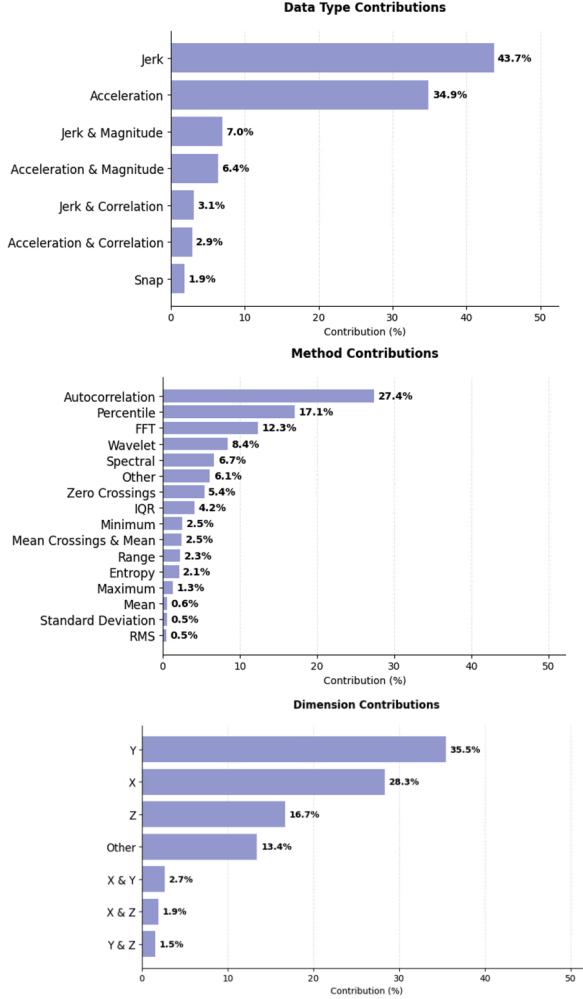


Figure 23: Feature Contribution Analysis by Data Type, Dimension, and Extraction Method. This figure shows the relative impact of each category on classification performance.

(2024). These models were developed using the extensive UK Biobank accelerometer dataset, encompassing over 700,000 person-days of continuous, free-living human motion data. This dataset is unparalleled in scale and diversity, featuring a broad range of natural human activities captured over seven days from more than 100,000 participants. The HarNet models leverage this data to fully realize the potential of deep learning for human activity recognition, overcoming the traditional limitations of small, lab-constrained datasets.

The training pipeline for the HarNet models incorporated multi-task self-supervised learning to pre-train a deep neural network, which was subsequently evaluated on eight diverse activity recognition benchmarks using transfer learning. This approach ensured that the models could generalize effectively across external datasets, devices, and

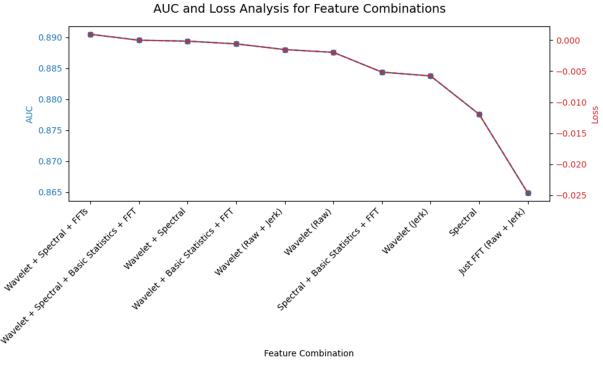


Figure 24: Feature Contribution Analysis by Controlled Feature Removal.

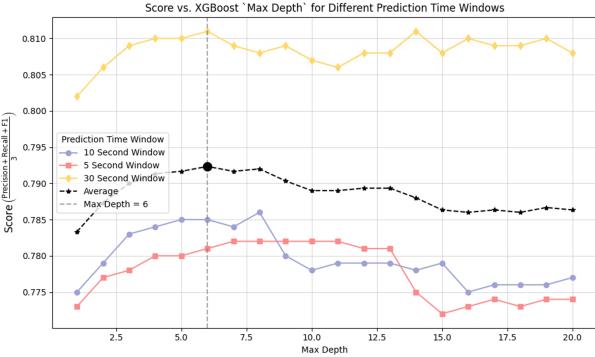


Figure 25: Score vs. XGBoost *max_depth* for Different Prediction Time Windows.

populations. The pre-trained models—HarNet5, HarNet10, and HarNet30—are tailored to classify activities in 5-second, 10-second, and 30-second windows of wrist-worn accelerometer data, respectively. These models consistently outperformed conventional baselines, achieving up to 130.9% relative improvement in F1 scores on smaller datasets.

The HarNet5, HarNet10, and HarNet30 models from the Oxford Wearables repository are designed to classify 5-second, 10-second, and 30-second long data samples. Each window of data represents acceleration across three channels: Accel-X, Accel-Y, and Accel-Z.

To process the already split data a two-step Butterworth filter was used: a low-pass filter to remove high-frequency noise, and a high-pass filter to eliminate gravitational components. The filtered signals were then normalized per axis using standard scaling, ensuring zero mean and unit variance, to account for differences in sensor sensitivity. Next, the data was clipped to lie within a range of -3g to +3g, removing extreme outliers and ensuring consistency with data used to pre-train the HarNet models.

When fine-tuning the HarNet models on our dataset, we explored a range of configurations to assess their impact on performance. All experiments incorporated early stopping with a patience of 5 epochs and utilized a batch size of 512 to ensure efficient training. To ensure reproducibility and comparability, each experiment was run three times, with each run having an assigned random seed. The fine-tuning configurations included unfreezing the classification layer, unfreezing the classification and last 2 feature extraction layers, and fully fine-tuning the entire model. These approaches allowed us to systematically evaluate the effectiveness of different fine-tuning strategies on model performance, to ultimately determine the best performing model.

6.3 Gait Characterization

6.3.1 Previous Findings and Data

Assessment of gait can be useful for characterizing different aspects of overall health. Not only does increasing step count benefit cardiovascular and musculoskeletal health, but the manner in which a person walks can also provide valuable information on their well-being. After running models to label walking bouts in our lower back and wrist sensor data, we can then use these labels to study walking periods.

From previous literature, we know that usual gait speed for walking is between 1 and 1.6 meters per second (Soltani et al., 2020) and usual walking cadence is between 110 and 125 steps per minute (Murtagh et al., 2021). Then we can conclude that in usual walking bouts, humans walk between 0.545 and 0.768 meters per step. According to Murtagh, a “slow” walking speed is measured between 0.53 and 1.04 meters per second. From these metrics, we can calculate regular and slow cadence.

Walking Type	Cadence (steps/min)
Usual Walking	110–125
Slow Walking	58–114

Because of our elderly population as described in the data section, we assume that our participants possibly walk slower than average. We can expect our participants to have a cadence between 58 and 114 steps per minute for slow to regular walking bouts.

Using the previously labeled data from the lower back sensor, we can study the cadence of

participants. Figure 26 shows cadence during walking bouts for four examples of anonymous participants. We see that our study follows what the literature tells us about typical cadence for slow walking, with an average cadence of 73.45 steps per minute.

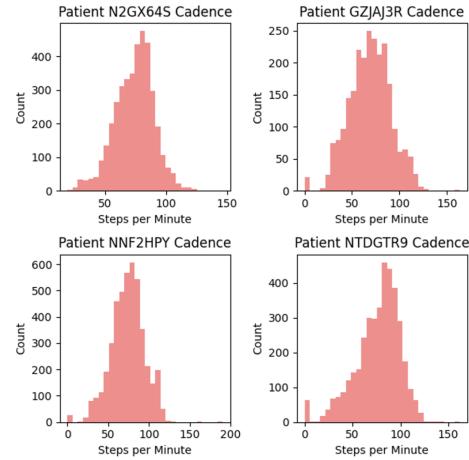


Figure 26: Cadence during periods labeled as “walking” for anonymous participants.

It is important to note that these walking bouts are not necessarily prolonged periods of walking, but could come from periods when the participant is doing a mixture of walking, standing and shuffling, though we only use cadence during times labeled as walking. Therefore, true cadence could differ, as the participants might be walking just a few steps at a time.

6.3.2 Cadence Algorithms

For the purpose of our study, we decided to focus our gait characterization on walking cadence. Once we had the labeled wrist-worn walking bout data, we were able to use it to study gait patterns measured via the wrist.

Find Peaks Algorithm According to Small (Small et al., 2023), the find peaks algorithm from the SciPy python package can be used for step detection. As an initial attempt, we used wrist-worn accelerometer data from labeled walking bouts as input to the algorithm. In addition to the find peaks function, we pre-processed the data by first filtering each of the three directional signals using a low-pass Butterworth filter. We then calculate the magnitude of the filtered signals, and standardize the signals. Finally, we run the find peaks algorithm with a distance parameter of half a second,

a dynamic threshold based on the average peak height, and a small prominence to capture small movements, inspired by Small’s work.

Algorithm 2: Estimate Strides

Inputs:
 accel_data: Acceleration data (x, y, z)
 cadence: Known cadence (strides/min)
 duration: Recording duration (s)
 height: Height of individual (m)
 sampling_rate: Sampling rate (Hz)

Outputs:
 Estimated cadence
 True cadence

Filter acceleration data: Low pass Butterworth filter for x, y, z
Compute acceleration magnitude: Calculate magnitude of filtered signals
Standardize Data: Standardize with mean/std
Set dynamic threshold:
 $\text{threshold} \leftarrow \text{mean}(\text{standardized magnitude})$
Detect peaks: peaks $\leftarrow \text{find_peaks}(\text{magnitude}, \text{distance} = 50\text{Hz}, \text{height} = \text{threshold}, \text{prominence} = 0.1)$
Calculate step durations and cadence:
 $\text{step_durations} \leftarrow \text{time between peaks}$
 $\text{stride_frequency} \leftarrow \frac{1}{\text{mean}(\text{step_durations})}$
 $\text{est_cadence} \leftarrow \frac{\text{len}(\text{peaks})}{\text{duration}/60}$
return {est_cadence, cadence}

Running this algorithm successfully captures the peaks in the walking bouts, as displayed below.

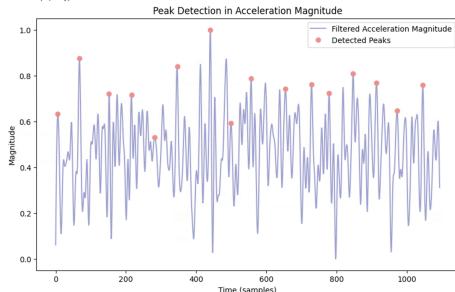


Figure 27: Example Peak Output from Estimate Strides algorithm

However, when we use the output of find peaks directly as a proxy for step count to calculate cadence, the accuracy is extremely low. When we average the results across the 6 tested participants, we achieve the results in Table 2 when comparing true versus estimated cadence.

Mean Difference	MAE	MAPE	R ²
-2.1	18.6	25.4%	0

Table 2: Results for Find Peaks Cadence estimation

Essentially, the algorithm mainly predicted cadence within a specific band as seen in Fig 28, and did not seem to capture the low or high cadence very well, most likely due to short walking bouts where there are not many steps detected.

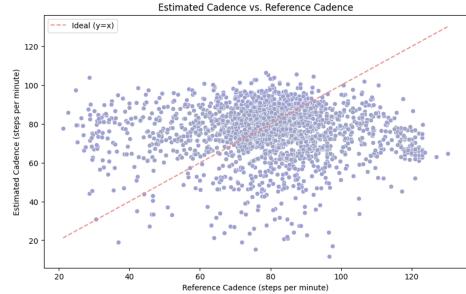


Figure 28: Predicted vs Actual Cadence

6.3.3 XGBoost with Find Peaks

Though the find peaks algorithm directly did not yield accurate estimations for cadence, previous literature led us to believe that the algorithm could still be valuable. Therefore, we opted to use the output from the find peaks algorithm as a feature in a machine learning model, along with demographic features that are correlated with cadence.

Methodology For each labeled walking bout, we run the find peaks algorithm as described above, returning "peaks" as an estimate for steps from the wrist-worn accelerometer signal. We use the number of peaks found, along with walking bout duration, subject age, subject weight, and subject height, as features in an XGBoost model to predict step count, which we can then use to calculate cadence.

$$\text{Cadence} = \frac{\text{Step Count}}{\text{Walking Duration (seconds)}} \times 60$$

The model is trained and validated using previously measured data, where Johnson Johnson provides step counts for periods of time, along with the labeled wrist-worn data from Mobilise-D. We have 15 total participants in the training data. Once we have trained and tested our model, we can then use the model to predict step counts for the data we have labeled as walking or non-walking. The signal from walking bouts from wrist-worn data can be passed to the find peaks algorithm to get the number of peaks. We also have duration for these walking bouts, as well as participant demographics. Thus, we are able to predict step count for all of our labeled walking bouts using our trained model.

To train the model, we use an 80-20 training and testing data split on the measured data, and choose the hyper-parameters in (Table 3) using GridSearchCV.

Parameter	Value
Number of Trees	50
Learning Rate	0.1
Maximum Tree Depth	4
Data Subsampling per Tree	80%
Feature Sampling per Tree	80%
L2 Regularization	2.0
Random State	42

Table 3: XGBoost model parameters from GridSearchCV

We run the model for walking bouts greater than 3 seconds to capture all walking bouts with more than one step. Then we zoom in and run the model with bouts 3-10 seconds, 10-30 seconds, and greater than 30 seconds.

7 Results

7.1 Activity Classification

7.1.1 Novel Wavelet Model

The performance of the custom wavelet model across different prediction windows (5, 10, and 30 seconds) is summarized in Figure 30. The model’s metrics, including Overall F1-Score, Precision, Sensitivity, and Specificity, show consistent improvement as the prediction window increases.

- **Overall F1-Score** steadily improves, reaching a peak of 0.826 for the 30-second window.
- **Precision (Walking)** demonstrates an increase from 0.695 to 0.734, indicating better accuracy in identifying walking instances.
- **Sensitivity (Walking)** remains high across all windows, with a maximum value of 0.883 for the 30-second window, reflecting the model’s robustness in detecting true positives.
- **Specificity (Walking)** improves to 0.784 for the largest window, showcasing reduced false-positive rates.

This analysis highlights the model’s ability to leverage longer prediction windows for enhanced performance across key metrics.

7.1.2 HarNet

We evaluated the performance of the HarNet5, HarNet10, and HarNet30 models for activity classification across various configurations, aiming to

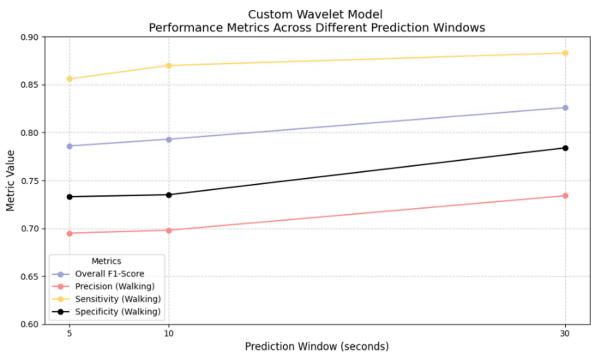


Figure 29: Performance Metrics Across Different Prediction Windows for the Custom Wavelet Model.

identify the optimal model and settings for our labeled wrist-worn accelerometer dataset. The results, summarized in Table 4, compare the performance of the pre-trained HarNet models with and without fine-tuning. Additionally, we examined the effects of different temporal window sizes and dataset sizes on model performance.

Our initial analysis began with data from six participants, consisting of labeled wrist-sensor data. By leveraging the Mobilise-D algorithm to label data from lower-back sensors, we were able to include an additional eight participants. Finally, the CAPTURE-24 dataset provided data from 151 more participants, significantly enhancing the diversity and volume of the dataset.

The results in the Table 4 demonstrate the F1 walking scores for the HarNet5, HarNet10, and HarNet30 models across different fine-tuning settings and datasets. We chose to evaluate the models using F1 scores since this provides a good balance between recall and precision of the walking detection.

1. **Comparison of HarNet Models:** Among the three HarNet models, HarNet30 consistently achieves the highest F1 walking scores across most configurations. This superior performance is likely attributed to its ability to process longer temporal windows (30 seconds), offering richer contextual information for walking classification. In contrast, the performance difference between HarNet5 and HarNet10 is less pronounced, with each model outperforming the other in specific configurations, indicating that their effectiveness may depend on the particular dataset or fine-tuning settings.
2. **Effect of Fine-tuning:** The impact of the

F1 Walking Scores					
Model	Data	Finetune Last Layer	Finetune Last 2 Layers	Fully Finetune	
HarNet5	Base	0.6077	0.5949	0.5945	
	Mob-D	0.6919*	0.6939	0.6656	
	Mob-D and CAPTURE	0.7586	0.7688	0.7742 , 0.6778 ‡, 0.6790 †	
HarNet10	Base	0.5836	0.6029	0.6103	
	Mob-D	0.6748	0.6900	0.6905	
	Mob-D and CAPTURE	0.7532	0.7586	0.7664	
HarNet30	Base	0.6366	0.6944	0.7023	
	Mob-D	0.7012	0.7544	0.7134	
	Mob-D and CAPTURE	0.7683	0.7839	0.7777	

Table 4: F1 Walking scores for different datasets for HarNet models with different levels of fine-tuning. † No filter applied. ‡ No standardization applied.

fine-tuning configuration appears to be less pronounced and is strongly influenced by the size of the dataset, which is intuitive. Smaller datasets may lack sufficient data to effectively update all parameters, potentially leading to overfitting or suboptimal adaptation. However, the results indicate that fine-tuning more than one layer generally improves the F1 walking score. This trend suggests that enabling the model to update a greater number of parameters allows it to adapt more effectively to the data, but achieving significant performance gains with deeper fine-tuning requires substantially larger and more diverse datasets.

3. **Impact of Datasets:** The inclusion of data from both the “Mobilise-D” and “CAPTURE-24” datasets consistently enhances the F1 walking score compared to using only the labeled wrist-worn data. This improvement highlights the value of incorporating diverse and larger datasets, as they provide additional variability and context, enabling the models to better generalize and adapt to complex walking classification tasks.

The best performing pre-trained model based on walking F1-score was HarNet30, finetuned on the last 2 feature extraction and classification layers using all 3 datasets. This model yielded the following metrics:

- **Overall F1-Score:** 0.7983
- **Precision (Walking):** 0.6849

- **Sensitivity (Walking):** 0.9164
- **Specificity (Walking):** 0.7161
- **F1 (Walking):** 0.7839

Finally, for the HarNet5 with the Mobilise-D and CAPTURE-24 dataset, we conducted additional experiments by removing filtering and standardization from the preprocessing pipeline. These results revealed that these preprocessing steps have a substantial impact on model performance, highlighting their importance in achieving optimal results.

Overall, the findings emphasize the critical role of both fine-tuning strategies and dataset diversity in enhancing model performance. Among the models, HarNet30 consistently demonstrated superior results. The best F1 scores for each HarNet model are highlighted in bold in the table.

7.1.3 Final Model Comparison

The comparative performance of Harnet and Wavelet models across key metrics (i.e. sensitivity, specificity, precision, and F1-score) was evaluated at different time intervals ($t = 5$, $t = 10$, $t = 30$). The average of those metrics across all time intervals was also used to compare models. As shown in Figure 30, the Wavelet model consistently achieves higher precision and specificity, particularly at $t = 10$ and $t = 30$. The Harnet model demonstrates slightly stronger sensitivity across the evaluated intervals, while both models show similar trends in F1-score. The results suggest that the Wavelet model provides more balanced performance across most metrics. On the

other hand, the Harnet model may be advantageous in applications prioritizing sensitivity while Wavelet performance is better for specificity and precision.

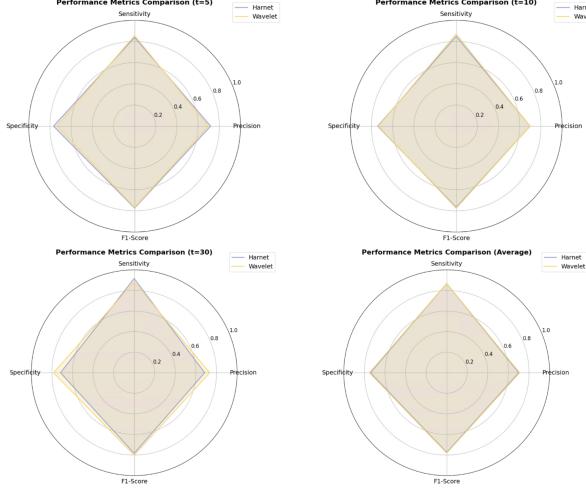


Figure 30: Performance metrics comparison of Harnet and Wavelet models across varying time intervals ($t = 5, 10, 30$) and their average performance.

7.2 Gait Characterization

XGBoost with Find Peaks The results for the XGBoost models using find peaks, described in the methodology section, are as follows in Table 5, Figure 31, and Figure 32.

Walking Duration	R ²	MAPE	RMSE (steps)
> 3 seconds	0.97	19.0%	6.6
3–10 seconds	0.86	16.6%	2.0
10–30 seconds	0.90	15.6%	5.4
> 30 seconds	0.95	12.0%	18

Table 5: Results from XGBoost model using Find Peaks for different walking bout durations

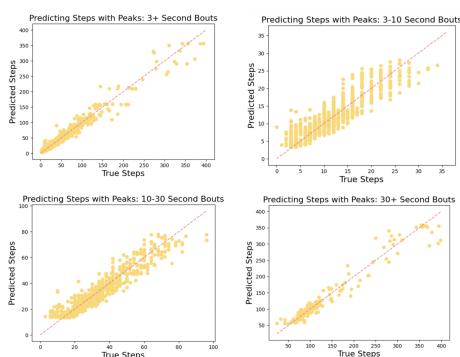


Figure 31: Predicted vs Actual Steps

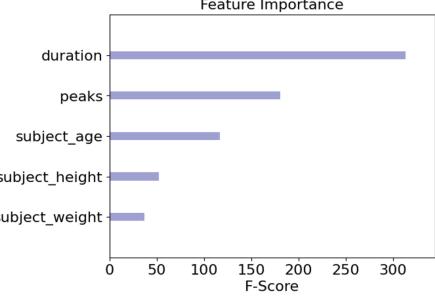


Figure 32: Feature Importance for 3+ second model

Evaluation The model using all walking bouts greater than 3 seconds gives us the best results. We can see from the results plot in Fig 31 as well as the results from the model using 3–10 second walking bouts that the model does begin to struggle with very short walking bouts. This is expected, as signal data from shorter walking bouts are not as consistent in terms of peaks, since these could come from bouts with only a few steps, or with a combination of walking and shuffling. Especially with wrist-worn devices, very short walking bouts could be misidentifying arm movements as steps, also leading to lower accuracy in our model.

Further, the model using all walking bouts greater than 3 seconds evidently has more data than the other models, which contributes to the better performance of the model. There is more variation in the data, adding complexity that aids in making predictions on unseen data.

Each model has the same order for feature importance based on F-score, with the walking bout duration and detected peaks being the most important features, followed by participant demographics (age, height, weight), as seen in Fig 32. The model results are impressive using only duration and peaks, however including the demographics improves the performance of each model.

With the high accuracy of the model using all walking bouts greater than 3 seconds, we can confidently use the predicted steps and walking bout duration to estimate cadence from a walking bout using wrist-worn data.

8 Discussion

8.1 Summary

Walking bout classification is generally a difficult task for wrist-worn devices because of the noise introduced by extraneous wrist movements. However, to counter the cost of manually labeling billions of samples with an activity, the Novel

Wavelet and HarNet models aim to classify walking bouts for unlabeled wrist accelerometer samples. The best performing HarNet model was HarNet30 when it was fully finetuned on the entire available dataset, yielding a F1 walking score of 0.7839. The results suggest the models are able to capture walking bouts and reasonably label data that can be used for gait characterization. With further optimizations to training parameters, model performances can be improved.

With unseen wrist-worn accelerometer data, walking bout duration, height, weight, and age, we can estimate gait characteristics. Since our models to predict step count are highly accurate, we can use the predicted step count to calculate cadence, even for very short walking bouts. This is a success, because this shows that participants can use wrist-worn devices in a real-world setting to accurately monitor gait parameters, without having to wear a large lower-back sensor or measure walking bouts in a lab setting.

8.2 Conclusion

In conclusion, we can successfully and accurately label walking bouts in real-world settings from wrist-worn accelerometer data using novel and pre-trained classification models. From this labeled data, we can also obtain precise estimations for step count and cadence during the associated walking bout. Ultimately, the objective of this pipeline is to aid in disease detection and monitoring for patients with neurological disorders, as tracking gait characteristics can provide insight into disease progression.

8.3 Future work

In the future, we would like to focus on gait parameters beyond cadence, such as speed or stride length. We would also like to leverage the model results to evaluate the participants' health risk, and understand how gait characteristics can be used as positive or negative indicators for each of our participants' conditions.

Further, because of the mechanics of wrist-worn sensors, this study was conducted only with white participants. In the future, we would like to expand the study to other races. We would also like to include more healthy participants and a wider, more balanced range of ages in the study to truly understand the variety of gait patterns across different groups.

8.4 Ethical Considerations

As a project that uses personal health data, there were many privacy precautions taken for the project. All data was anonymous, and only those pre-approved by Johnson & Johnson have access to this data.

9 Contributions

- Introduction Literature: Anouck, Lucia
- Data Overview: Aikaterini, Deha
- Data Preprocessing: Deha
- Mobilise-D: Aikaterini
- Classification Models: Anouck, Bhavana, Deha
- Gait Characterization: Lucia

References

- Abedinifar, M., B. Kline, S. R. van der Walt, and A. Wiskott. 2024. *KielMAT: Kiel Motion Analysis Toolbox - An open-source Python toolbox for analyzing neurological motion data from various recording modalities*. Journal of Open Source Software, 9(102), 6842. <https://doi.org/10.21105/joss.06842>
- Brand, Yonatan E., Dafna Schwartz, Eran Gazit, Aron S. Buchman, Ran Gilad-Bachrach, and Jeffrey M. Hausdorff. 2022. *Gait Detection from a Wrist-Worn Sensor Using Machine Learning Methods: A Daily Living Study in Older Adults and People with Parkinson's Disease*. Sensors, 22(18), 7094. MDPI. <https://doi.org/10.3390/s22187094>
- Brand, Yonatan E., Felix Kluge, Luca Palmerini, Anisoara Paraschiv-Ionescu, Clemens Becker, Andrea Cereatti, Walter Maetzler, Basil Sharrack, Beatriz Vereijken, Alison J. Yarnall, Aron S. Buchman, and Jeffrey M. Hausdorff. 2024. *Self-supervised Learning of Wrist-worn Daily Living Accelerometer Data Improves the Automated Detection of Gait in Older Adults*. Scientific Reports, 14(20854). <https://doi.org/10.1038/s41598-024-71491-3>
- Chan, Lloyd L. Y., Tiffany C. M. Choi, Stephen R. Lord, and Matthew A. Brodie. 2022. *Development and Large-scale Validation of the Watch Walk Wrist-worn Digital Gait Biomarkers*. Scientific Reports, 12(16211). <https://doi.org/10.1038/s41598-022-20327-z>
- Chan, Shing, Hang Yuan, Catherine Tong, et al. 2024. *CAPTURE-24: A Large Dataset of Wrist-Worn Activity Tracker Data Collected in the Wild for Human Activity Recognition*. *Scientific Data*, 11, 1135. <https://doi.org/10.1038/s41597-024-03960-3>.

- Demirel, Berken U. and Christian Holz. 2024. *An Unsupervised Approach for Periodic Source Detection in Time Series*. Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. arXiv preprint arXiv:2406.00566.
- Femiano, Riccardo, Charlotte Werner, Matthias Wilhelm, and Prisca Eser. 2022. *Validation of Open-Source Step-Counting Algorithms for Wrist-Worn Tri-Axial Accelerometers in Cardiovascular participants*. Gait Posture, 92, 206-211. <https://doi.org/10.1016/j.gaitpost.2021.11.035>
- Kirk, C., Küderle, A., Micó-Amigo, M.E., et al. 2024. *Mobilise-D insights to estimate real-world walking speed in multiple conditions with a wearable device*. Scientific Reports, 14, 1754. <https://doi.org/10.1038/s41598-024-51766-5>
- Kluge, Felix, Yonatan E. Brand, M. Encarna Micó-Amigo, Stefano Bertuletti, Ilaria D'Ascanio, Eran Gazit, Tecla Bonci, Cameron Kirk, Arne Küderle, Luca Palmerini, Anisoara Paraschiv-Ionescu, Francesca Salis, Abolfazl Soltani, Martin Ullrich, Lisa Alcock, Kamiar Aminian, Clemens Becker, Philip Brown, Joren Buekers, Anne-Elie Carsin, Marco Caruso, Brian Caulfield, Andrea Cereatti, Lorenzo Chiari, Carlos Echevarria, Bjoern Eskofier, Jordi Evers, Judith Garcia-Aymerich, Tilo Hache, Clint Hansen, Jeffrey M. Hausdorff, Hugo Hiden, Emily Hume, Alison Keogh, Sarah Koch, Walter Maetzler, Dimitrios Megaritis, Martijn Niessen, Or Perlman, Lars Schwickert, Kirsty Scott, Basil Sharrack, David Singleton, Beatrix Vereijken, Ioannis Vogiatzis, Alison Yarnall, Lynn Rochester, Claudia Mazzà, Silvia Del Din, and Arne Mueller. 2024. *Real-World Gait Detection Using a Wrist-Worn Inertial Sensor: Validation Study*. JMIR Formative Research, 8(1), e50035. <https://doi.org/10.2196/50035>
- Micó-Amigo, M., Bonci, T., Paraschiv-Ionescu, A., et al. 2023. *Assessing real-world gait with digital technology? Validation, insights and recommendations from the Mobilise-D consortium*. Journal of NeuroEngineering and Rehabilitation, 20, 78. <https://doi.org/10.1186/s12984-023-01198-5>
- Mikos, Val, Shih-Cheng Yen, Arthur Tay, Chun-Huat Heng, Chloe Lau Ha Chung, Sylvia Hui Xin Liew, Dawn May Leng Tan, and Wing Lok Au. 2018. *Regression Analysis of Gait Parameters and Mobility Measures in a Healthy Cohort for Subject-Specific Normative Values*. PLoS ONE, 13(6), e0199215. <https://doi.org/10.1371/journal.pone.0199215>
- Murtagh, Elaine M., Jacqueline L. Mair, Elroy Aguiar, Catrine Tudor-Locke, and Marie H. Murphy. 2021. *Outdoor Walking Speeds of Apparently Healthy Adults: A Systematic Review and Meta-analysis*. Sports Medicine, 51(1), 125-141. <https://doi.org/10.1007/s40279-020-01351-3>
- Paraschiv-Ionescu, A., B. Newman, A. A. Sangeux, and K. Aminian. 2019. *Locomotion and cadience detection using a single trunk-fixed accelerometer: validity for children with cerebral palsy in daily life-like conditions*. Journal of NeuroEngineering and Rehabilitation, 16(1), 24. <https://doi.org/10.1186/s12984-019-0494-z>
- Paraschiv-Ionescu, A., B. Newman, M. Hamill, A. A. Sangeux, and K. Aminian. 2020. *Real-world speed estimation using a single trunk IMU: methodological challenges for impaired gait patterns*. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. <https://doi.org/10.1109/EMBC44109.2020.9176281>
- Pilkar, Rakesh, Dawid Gerstel, Ethan Toole, Matt Biggs, Tyler Guthrie, Marta Karas, Christopher Moufawad el Achkar, Philippe Renevey, Abolfazl Soltani, Sarah Sloan, Joe Nguyen, Matthew R. Patterson, Damien Ferrario, Mathieu Lemay, Ali Neishabouri, and Christine Guo. 2022. *Performance Analyses of Step-Counting Algorithms Using Wrist Accelerometry*. Research Square, rs-2183645. <https://doi.org/10.21203/rs.3.rs-2183645/v2>
- Small, Scott R., Shing Chan, Rosemary Walmsley, Lennart von Fritsch, Aidan Acquah, Gert Mertes, Benjamin G. Feakins, Andrew Creagh, Adam Strange, Charles E. Matthews, David A. Clifton, Andrew J. Price, Sara Khalid, Derrick Bennett, and Aiden Doherty. 2023. *Development and Validation of a Machine Learning Wrist-Worn Step Detection Algorithm with Deployment in the UK Biobank*. medRxiv preprint, 2023.02.20.23285750. <https://doi.org/10.1101/2023.02.20.23285750>
- Soltani, Abolfazl, Hooman Dejnabadi, Martin Savary, and Kamiar Aminian. 2020. *Real-World Gait Speed Estimation Using Wrist Sensor: A Personalized Approach*. IEEE Journal of Biomedical and Health Informatics, 24(3), 658–667. <https://doi.org/10.1109/JBHI.2019.2914940>
- Vayalapra, Sushanth, Xueyang Wang, Arham Qureshi, Abhinav Vepa, Usama Rahman, Arnab Palit, Mark A. Williams, Richard King, and Mark T. Elliott. 2022. *Repeatability of Inertial Measurement Units for Measuring Pelvic Mobility in participants Undergoing Total Hip Arthroplasty*. Sensors, 23(1), 377. <https://doi.org/10.3390/s23010377>
- Yuan, Hang, Shing Chan, Andrew P. Creagh, Catherine Tong, Aidan Acquah, David A. Clifton, and Aiden Doherty. 2024. *Self-supervised Learning for Human Activity Recognition Using 700,000 Person-Days of Wearable Data*. npj Digital Medicine, 7(1). <https://doi.org/10.1038/s41746-024-01062-3>