Scientific Data

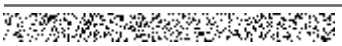
https://doi.org/10.1038/s41597-026-06710-9

Article in Press

A comprehensive IMU dataset for evaluating sensor layouts in human activity and intensity recognition

Received: 6 October 2025

Accepted: 24 January 2026



Cite this article as: Feng, M., Zhang, Q., Fang, H. A comprehensive IMU dataset for evaluating sensor layouts in human activity and intensity recognition. *Sci Data* (2026). https://doi.org/10.1038/s41597-026-06710-9

Mingfei Feng, Qiwei Zhang & Hongbin Fang

A comprehensive IMU dataset for evaluating sensor layouts in human activity and intensity recognition

Authors

Mingfei Feng[1,2], Qiwei Zhang[1,2], & Hongbin Fang[1,2]

Affiliations

1. College of Intelligent Robotics and Advanced Manufacturing, State Key Laboratory of Brain Function and Disorders, Fudan University, Shanghai 200433, China.
2. Yiwu Research Institute, Fudan University, Yiwu, Zhejiang 322000, China.

corresponding author(s): Hongbin Fang (fanghongbin@fudan.edu.cn)

Abstract

Human activity recognition (HAR) with wearable sensors is widely applied in health monitoring, fitness tracking, and smart environments, but the choice of sensor configuration remains a critical factor for balancing recognition performance with usability and comfort. Existing datasets often lack the full-body coverage required to systematically evaluate sensor placement strategies. We present a comprehensive dataset of 12 daily activities performed by 30 participants, recorded using 17 inertial measurement units (IMUs) distributed across the entire body. Each IMU provides tri-axial acceleration and angular velocity signals at 60 Hz, aligned within a standardized global coordinate system. The dataset further includes detailed anthropometric metadata, structured annotations of activity and effort level, and processing scripts to support feature extraction, segmentation, and baseline model training. Benchmark experiments with both machine learning and deep learning models demonstrate the usability of the dataset across multiple temporal windows and sensor subsets. This resource enables systematic evaluation of sensor layout strategies and supports the development of practical, generalizable HAR systems.

Background & Summary

Human Activity Recognition (HAR) has emerged a vibrant research field with wide-ranging applications in medical monitoring[1–3], fitness tracking[4–6], and intelligent living environments[7–9]. While vision-based methods have shown impressive performance, HAR systems based on wearable sensors remain more robust for long-term monitoring, privacy-preserving scenarios, and settings involving unconstrained movement. Among these, inertial measurement units (IMUs) are particularly attractive due to their lightweight design, portability, and ability to capture accurate motion signals[10]. A central challenge in IMU-based HAR, however, lies in the choice and configuration of sensor placement across the body.

Over the past decade, several public datasets have advanced the study of HAR using IMUs. The Skoda dataset[11], one of the earliest contributions, employed 19 accelerometers on the arms to classify 10 industrial gestures, explicitly exploring dynamic sensor selection but remaining restricted to upper-body tasks. The PAMAP2[12,13] dataset recorded 18 physical activities from 9 participants using 3 IMUs on the wrist, chest, and ankle—providing coverage of both torso and limbs, though at a relatively sparse density. The Opportunity[14] dataset extended this paradigm with a more complex multimodal setup of 7 IMUs, 12 accelerometers, and 4 localization sensors across the back, arms, and legs to support context-aware recognition in smart home environments. However, its heterogeneous sensor layout across body regions, combined with pronounced data imbalance, limit its generalizability. In contrast, widely used benchmarks such as WISDM[15] and HCI-HAR[16] adopt a single-device paradigm, focusing on using smartphone-based sensing from the pocket or waist. While convenient, these configurations provide

only a narrow view of whole-body dynamics. Similarly, the Capture-24 dataset[17] offered ecologically valid recordings under free-living conditions with more than 200 fine-grained activity labels, but relies solely on a single wrist-worn accelerometer, thereby restricting analyses of multi-sensor layout and trade-offs.

Despite their impact, existing datasets exhibit notable limitations in sensor coverage and layout. Some datasets emphasize specific regions, such as arms or legs, while others attempt full-body monitoring but suffer from uneven sensor distribution. Consequently, the effects of sensor placement on recognition performance remains insufficiently characterized [18–21]. To address these gaps, we present a new IMU-based dataset collected from 30 participants performing 12 daily activities spanning a range of intensities. Each subject was equipped with 17 IMUs uniformly distributed IMUs, enabling comprehensive coverage of full-body motion. Alongside activity labels, we provide detailed annotations of physical intensity, supporting detailed analysis of human movement. Crucially, the dataset is designed to facilitate systematic evaluation of sensor subsets—such as upper- or lower-body only, or wrist-based sensing alone—allowing researchers to quantify the trade-offs between recognition accuracy, sensor density, and deployment practicality. By expanding both the scale and uniformity of sensor coverage, this dataset establishes a new foundation for exploring efficient sensor layouts in real-world HAR applications.

Methods

Participants. Participants were recruited via campus advertisements. Thirty healthy adult participants volunteered to participate in this study. None reported neurological or musculoskeletal disorders that could affect motion performance. Participant characteristics are summarized in Table 1. All procedures were approved by the Ethics Committee of Fudan University, China (Application No. FE21124, approval date: 16 August, 2021), and informed consent for both their participation in the study and the open-access publication of their anonymized data was obtained from each participant prior to data collection. To protect privacy, all personally identifiable information was removed, and participants were assigned random ID codes (e.g., P01, P02).

Data collection. To minimize motion artifacts from clothing, participants wore athletic shirts and shorts, as well as athletic footwear to prevent injuries during physical activities. Prior to sensor placement, anthropometric measurements—including age, sex, height, weight, handedness, and body segment lengths—were recorded from each participant.

A total of 17 IMUs from the Perception Neuron Studio motion capture system (Noitom Technology Ltd., Beijing, China) were affixed to standardized anatomical landmarks, following the manufacturer's guidelines. Sensor locations were as follows: (1) posterior head; (2&3) shoulders (upper scapula), (4) mid-upper back (medial scapular area), (5&6) dorsal upper arms (midpoint between shoulder and elbow), (7&8) dorsal forearms (2/3 distal from elbow to wrist), (9&10) center of dorsal wrists, (11) lower back (L5 vertebra), (12&13) lateral

Table 1. Participant demographics

Characteristics	Mean±SD
Participants (n)	30
Sex (M/F)	20/10
Age (years)	24.6±3.1
Height (cm)	172.1±8.7
Weight (kg)	63.9±10.2
BMI (kg/m²)	21.5±2.4

Dominant hand	29 right / 1 left

thighs (just below the hips), (14&15) anterior shanks (just below the knees), and (16&17) center of dorsal feet. Sensors were secured with manufacturer-provided straps to minimize interference from muscle stretching or vibration (Figure 1(a)). Temporal synchronization across the 17 IMU nodes is ensured natively by the hardware architecture. The system utilizes a centralized receiver that aggregates and synchronizes data packets from all nodes.

Before recording, each participant performed two calibration poses, A-pose and T-pose (Figure 1(a)), following the system's user guide. In the A-pose, participants stood with palms resting against the thighs and feet parallel. In the T-pose, participants abducted their arms to 90˚ with palms facing downward. Data were transmitted wirelessly to a remote desktop at a sampling frequency of 60 Hz (Figure 1(b)).

Participants then performed 12 daily activities: lying, sitting, standing, slow walking, moderate walking, brisk walking, ascending stairs, descending stairs, cycling, running, jumping, and rowing, as depicted in Figure 1(c). Details descriptions of each activity are provided in Table 2. To reduce fatigue and ensure consistent performance, activities were divided into shorter trials based on their metabolic equivalent of tasks (MET)[22,23]. Activities of varying intensity were interleaved to form seven protocols (Figure 1(d) and Table 3), thereby balancing workload across sessions and minimizing error due to fatigue.

Figure 1 goes here

Benchmarks. To demonstrate the usability of the dataset for HAR, we conducted benchmarking experiments using both traditional machine learning classifiers and modern deep learning models. Models were evaluated under varying temporal window sizes and sensor configurations, and performance was assessed on two tasks: activity classification and intensity classification.

Tasks. The activity classification task required recognition of 12 activity categories, illustrated in Figure1(c). Transient or irrelevant movements were excluded from labeling. The intensity classification task aimed to distinguish activity intensity level, defined according to MET[22,23]. Activities were grouped into four categories:

(i) Sedentary effort (< 1.5 METs): lying and sitting;
(ii) Light effort (1.5-3.0 METs): standing and slow walking;
(iii) Moderate effort (3.0-6.0 METs): moderate walking, brisk walking, descending stairs, cycling, and rowing;
(iv) Vigorous effort (> 6.0 METs): ascending stairs, running, and jumping.

All benchmark experiments used only the raw inertial signals (accelerometer and gyroscope) as model inputs. Quaternion data, available from the motion capture system, was deliberately excluded to ensure comparability with typical wearable-sensor datasets.

Models. We evaluated four representative models widely used in HAR research:

(i) Random forest (RF): A balanced random forest classifier with 100 trees was implemented using the default hyperparameters provided by scikit-learn[24], which have been shown to be robust for similar tasks.
(ii) Support vector machine (SVM): A radial basis function (RBF) kernel was used, with default values for the regularization parameter, kernel coefficient, and other hyperparameters.
(iii) Convolutional neural network (CNN): We adopted a 1D variant of ResNet-18[25], in which 2D convolutional blocks were replaced with 1D convolutions to capture temporal dependencies. The architecture consists of four residual stages with {2,2,2,2} basic blocks, followed by global average pooling and a fully connected output layer.

(iv) Recurrent neural network (RNN): A bidirectional Long Short-Term Memory (LSTM) network with two layers and 128 hidden units per direction was implemented. The output from the final time step was passed through a dropout layer (p=0.5) and a fully connected classifier.

Table 2. Detailed descriptions of 12 daily activities

Activity	Activity	Description
1	Lying	Lying quietly and still, allowing only slight movement, such as changing the lying posture
2	Sitting	Sitting in a chair in any posture the participant finds comfortable, allowing adjustments to sitting postures
3	Standing	Standing still or talking while standing still, possibly gesticulating
4	Slow walking	Walking at a low speed of 0.9 m/s on an indoor treadmill
5	Moderate walking	Walking at a moderate speed of 1.2 m/s on an indoor treadmill
6	Brisk walking	Walking at a fast speed of 1.5 m/s on an inside treadmill
7	Ascending stairs	Performed at moderate intensity indoors, requiring crossing two floors when ascending
8	Descending stairs	Performed at moderate intensity indoors, requiring crossing two floors when descending
9	Cycling	Performed indoors on a stationary bicycle at a slow to moderate pace, as if the participant were commuting to work or cycling for leisure (rather than as a sport activity)
10	Running	Jogging on an indoor treadmill, with speed adjusted according to the participant's personal preference
11	Jumping	The basic jump action performed by jumping with both feet simultaneously, possibly gesticulating
12	Rowing	Performed indoor on a rowing machine at slow to moderate pace, as if the participant were engaged in leisurely rowing (rather than as a sport activity)
0	Other	Transient or irrelevant activities

Table 3. Data collection protocols

Protocol	Action	MET	Iteration
Protocol 1	Standing for 1 minute Slow walking for 2 minutes Running for 2 minutes Resting for 2 minutes	1.8 MET 2.5 MET 7.0-8.0 MET –	5 times
Protocol 2	Standing for 1 minute Moderate walking for 2 minutes Brisk walking for 2 minutes Resting for 1 minute	1.8 MET 3.3 MET 3.8 MET –	5 times
Protocol 3	Sitting for 1 minute Jumping for 1 minute Resting for 1 minute	1.8 MET 8.0-10.0 MET –	10 times
Protocol 4	Lying for 10 minutes	1.0 MET	1 time
Protocol 5	Rowing for 1 minute Resting for 1 minute	7.0 MET –	10 times
Protocol 6	Cycling for 3 minutes Resting for 1 minute	4.0 MET –	4 times

	Ascending stairs for 30 seconds	8.0 MET	
Protocol 7	Descending stairs for 30 seconds	3.0 MET	20 times
	Resting for 1 minute	–	

Input representations. For RF and SVM classifiers, we extracted handcrafted features from the tri-axial accelerometer and gyroscope signals of each IMU channel. A total of 18 features, commonly adopted in prior HAR literature[16,17,26], were considered. These include 14 time-domain features—mean, variance, maximum, minimum, range, skewness, energy, entropy, interquartile range (IQR), median absolute deviation (MAD), root mean square (RMS), signal magnitude area (SMA), zero-crossing rate (ZCR), and mean-crossing rate (MCR)—and 4 frequency-domain features: spectral centroid, spectral variance, spectral entropy, and dominant frequency.

To systematically examine the effects of feature engineering and dimensionality on classification performance, we designed four feature subsets with increasing complexity:

(i) Set of 4 features: mean, variance, spectral centroid, spectral variance (24 features per IMU per window).

(ii) Set of 7 features: Set of 4 features + maximum, minimum, dominant frequency (42 features per IMU per window).

(iii) Set of 9 features: Set of 7 features + ZCR, MCR (54 features per IMU per window).

(iv) Set of 18 features: Full set of 18 features (108 features per IMU per window).

In addition, we benchmarked RF and SVM classifiers directly on raw inertial signals to evaluate the benefit of feature extraction. For the deep learning models (ResNet1D and bidirectional LSTM), raw accelerometer and gyroscope signals were used exclusively, without handcrafted features, in line with standard practice.

Evaluation Protocol. To ensure robust and subject-independent evaluation, we adopted a subject-wise 5-fold cross-validation. In each fold, data from five participants were held out for testing, while the remaining 20 are used for training. Sliding-window segmentation with 50% overlap was applied, ensuring that all data from a given participant were strictly assigned to either training or testing set, thereby eliminating the risk of data leakage[27].

To examine the effect of temporal granularity on classification performance, we evaluated model performance under eight window sizes: 0.5, 1, 1.5, 2, 2.5, 5, 7.5, and 10 seconds. These windows span short to long temporal contexts commonly employed in HAR tasks. Windows containing multiple activities were excluded. Each model was trained independently for each window size, and performance was reported as the average accuracy across the five folds. For RF and SVM trained on raw data, only the four longest windows (2.5s to 10s) were considered due to computational constraints.

Sensor Configuration Analysis. To quantify the impact of sensor placement on classification accuracy, we evaluated six reduced configurations (Figure 1(f)): an upper-body setup (10 IMUs), a lower-body setup (7 IMUs), an L5-thigh-shank setup (3 IMUs located at L5 vertebra, left thigh, and left shank), and three single-IMU setups located at the left thigh, left foot, and left wrist.

These reduced configurations were designed to simulate practical scenarios with constrained sensor deployment. For each configuration, the best-performing model identified in the full-body experiments was retrained using only the corresponding subset of sensors. The resulting performance provides baseline references for sensor selection studies and offers insight into the sufficiency of partial-body input for real-world HAR applications.

Implementation Details. All experiments were conducted in Python (Python Software Foundation, https://www.python.org/). The RF and SVM classifiers were implemented using the *scikit-learn* library[24], while the DL models (ResNet1D and LSTM) were implemented in PyTorch[28] and trained on a single Nvidia GeForce RTX 2080 Ti GPU. For the DL models, optimization was performed using the Adam algorithm with a learning rate of 0.001 using cross-entropy loss. Training was subject to early stopping

with a patience of 10 epochs to prevent overfitting. All input signals were Z-score normalized using the mean and standard deviation of the training set, ensuring feature scaling within the range 0–1 prior to model training.

Data Records

Each IMU provides multi-channel signals, including tri-axial acceleration from an accelerometer (Acc) and tri-axial angular velocity from the gyroscope (Gyr). In addition, the motion capture system computes the sensor's orientation in four quaternions (Quat). All signals are expressed relative to a world coordinate system established during calibration. In this system, the *Z*-axis points forward, aligned with the participant's facing direction during calibration (from back to front), the *Y*-axis points vertically upward, and the *X*-axis completes the right-handed coordinate system, pointing laterally to the participant's left. The origin is defined as the horizontal projection of the sensor positioned at the L5 vertebra during calibration.

The raw readings are stored as CSV files in the 'data' folder, available at Figshare[29]. Each participant's trial data file is named as 'P#.csv', where '#' denotes the participant identifier. The total size of the dataset is approximately 21.2 GB. Anthropometric information for all participants is provided in a separate CSV file 'anthropometric information.csv'. The structure of the trial data files is summarized in Table 4, which lists the columns definitions for accelerometer, gyroscope, and quaternion channels.

Table 4. Data columns description. 'BodyPart*' represents 17 body parts equipped with IMUs, including: LowerBack, RightThigh, RightShank, RightFoot, LeftThigh, LeftShank, LeftFoot, UpperBack, Head, RightShoulder, RightUpperArm, RightForeArm, RightWrist, LeftShoulder, LeftUpperArm, LeftForeArm, and LeftWrist

Columns	Format	Units	Descriptions
Activity	Int	N/A	Activity ID
Acc_X_BodyPart*	Float	m/s^2	Acceleration along the *X*-axis
Acc_Y_BodyPart*	Float	m/s^2	Acceleration along the *Y*-axis
Acc_Z_BodyPart*	Float	m/s^2	Acceleration along the *Z*-axis
Gyr_X_BodyPart*	Float	rad/s	Angular velocity about the *X*-axis
Gyr_Y_BodyPart*	Float	rad/s	Angular velocity about the *Y*-axis
Gyr_Z_BodyPart*	Float	rad/s	Angular velocity about the *Z*-axis
Quat_X_BodyPart*	Float	N/A	*X*-component of orientation quaternion
Quat_Y_BodyPart*	Float	N/A	*Y*-component of orientation quaternion
Quat_Z_BodyPart*	Float	N/A	*Z*-component of orientation quaternion
Quat_W_BodyPart*	Float	N/A	Scalar component of orientation quaternion

Data Overview

Figure 2 illustrates acceleration and angular velocity data from 6 sensors (head, upper back, left wrist, lower back, left thigh, and left foot) captured within a 3-second window from participant P01 during a moderate walking trial, serving as an example of raw data. Table 5 provides the average duration and standard deviation of each activity across participants. Table 6 lists the number of resulting samples for each activity when using the window sizes ranging from 0.5 to 10 seconds (with 50% overlap).

Figure 2 goes here

Technical Validation

Benchmark Results and Analysis. We validate the usability of the dataset for HAR through extensive benchmarking experiments spanning classification tasks, model families, temporal window sizes, and sensor configurations.

Table 5. Activity durations

Activity_ID	Activities	Duration(s)	Activity_ID	Activities	Duration(s)
1	Lying	564.3±285.5	7	Ascending stairs	473.4±157.6
2	Sitting	719.0±416.6	8	Descending stairs	440.8±152.5
3	Standing	652.9±210.6	9	Cycling	610.7±130.6
4	Slow walking	612.1±153.0	10	Running	609.6±114.8
5	Moderate walking	610.0±176.3	11	Jumping	571.2±160.2
6	Brisk walking	633.1±95.2	12	Rowing	566.6±193.8
			0	Others	332.7±194.3

Table 6. Number of samples for each activity and window size

Activity_ID	Activities	0.5s	1s	1.5s	2s	2.5s	5s	7.5s	10s
1	Lying	67647	33783	22501	16853	13475	6699	4441	3312
2	Sitting	86011	42864	28482	21290	16975	8353	5484	4043
3	Standing	77748	38560	25510	18983	15065	7273	4701	3409
4	Slow walking	73362	36632	24393	18267	14594	7244	4798	3576
5	Moderate walking	73101	36503	24297	18201	14535	7218	4774	3557
6	Brisk walking	75872	37885	25225	18892	15098	7498	4967	3700
7	Ascending stairs	54589	26154	16685	11961	9103	3418	1581	698
8	Descending stairs	50541	24060	15229	10834	8164	2888	1241	521
9	Cycling	73198	36554	24343	18234	14578	7248	4804	3579
10	Running	72996	36410	24220	18121	14469	7153	4712	3498
11	Jumping	68126	33853	22426	16710	13285	6443	4153	3024
12	Rowing	67909	33913	22580	16919	13515	6717	4451	3323

Activity Classification. Figure 3 illustrates the performance of RF, SVM, ResNet1D and LSTM on the 12-class activity recognition task across window sizes ranging from 0.5 to 10 seconds. Among traditional machine learning classifiers, the SVM with 18 handcrafted features achieves the highest accuracy, exceeding 95.8% across all window sizes and peaking at 97.1% with a 7.5-second window. The RF classifier demonstrates competitive performance, reaching 95.8% with 18 features at 2 seconds, while also achieving 95.8% with only 9 features on longer windows (10 seconds).

Figure 3 goes here

Both models show sensitivity to feature richness: SVM benefits significantly from richer feature representations, with accuracy dropping by 2.8±0.5% when reduced to only 4 features. RF is less sensitive, showing a 0.9±0.2% reduction when features are reduced from 18 to 4. In contrast, when trained on raw signals without feature engineering, both SVM and RF models degrade significantly—particularly SVM, which falls below 70% accuracy on longer windows—highlighting the necessity of handcrafted features for traditional classifiers. DL models, by contrast, are generally more robust to raw input. ResNet1D perform competitively, particularly on shorter windows, reaching 95.1% accuracy with raw data on a 2-second window. LSTM models trail slightly, plateauing around 94.5%.

Figure 4 depicts normalized confusion matrices for the best-performing traditional classifier (SVM with 18 features) and DL model (ResNet1D, raw input), both evaluated at the 2-second window. A common source of confusion occurs between static postures, particularly the distinction between sitting and standing. This is expected because the static nature of the signals can be highly similar across these classes. Another notable confusion arises in fine-grained categories, such as walking activities at

different speeds—through varying in intensity, they exhibit similar patterns. Overall, SVM demonstrates stronger performance, as reflected by the higher diagonal values. However, SVM also shows a distinct misclassification pattern: running mislabeled as jumping. This confusion occurs due to the similar impulsive vertical acceleration peaks and lower-limb dynamics shared by these two high-intensity activities. It suggests that SVM with handcrafted features emphasizes localized signal characteristics and is less effective at capturing the global movement differences. In contrast, the ResNet1D captures global temporal patterns and is better able to separate such distinct activities, resulting in minimal confusion between jumping and running. In addition, note that both models encounter challenges with rowing-sitting/lying confusion pair. Unlike the confusion between running and jumping that stems from dynamic properties, the confusion between rowing and sitting/lying arises from quasi-static phases. The catch and recovery phases of rowing involve a seated posture with a brief reduction in speed, potentially leading classifiers to misinterpret them as static sitting. Additionally, participants may exhibit brief pauses during the recovery phase of the stroke. In a 2s window, these phases are statistically difficult to distinguish from static postures. This limitation is effectively mitigated by increasing the window size. When the window is extended to 10s, the SVM results in only a single instance of rowing being misclassified as sitting. Under the same 10s window, ResNet1D misclassifies 28 rowing samples (0.8% of all rowing samples) as sitting and 6 samples (0.2% of all rowing samples) as lying.

Figure 4 goes here

Intensity Classification. Figure 5 summarizes model performance on the four-level intensity classification task across window sizes ranging from 0.5 to 10 seconds. Here, the SVM with 18 features again achieves the highest overall accuracy, reaching 97.9% with a 5-second window. RF models with 18 features also performs competitively, achieving 97% at 1.5 seconds, with only marginal gains for longer windows. Deep learning models exhibit similar trends, with ResNet1D peaking at 96.2% and LSTM trailing slightly.

Figure 5 goes here

Across all models, intensity classification accuracy is consistently higher than activity recognition accuracy, except SVM using raw data at 2.5s. Moreover, the gap between handcrafted-feature models and raw-input DL models is narrower. This is likely because, for intensity classification task, grouping activities into broader intensity levels reduces inter-class confusion and lowers task difficulty.

Effects of Window Sizes. We further investigated the effects of temporal granularity on classification performance. In both tasks, performance improves with increasing window size, but gains saturated beyond moderate durations across all tasks, models, and input types. For activity classification, accuracy rises noticeably between 0.5 s and 2 s, beyond which improvements became marginal. Intensity classification follows a similar trend, with modest gains between 0.5 s and 5 s, and negligible or even negative changes beyond 5 s. These findings suggest that longer windows capture richer temporal context, windows of 2~5 seconds provides an effective balance between recognition accuracy and latency. This balance is especially beneficial for real-time and wearable applications, where responsiveness is critical.

Effects of Sensor Configuration. To evaluate the trade-off between recognition accuracy and sensor deployment, we evaluated six reduced sensor configurations: an upper-body setup (10 IMUs), a lower-body setup (7 IMUs), a L5-thigh-shank setup (3 IMUs: L5 vertebra, left thigh, left shank), and three single-IMU setups (left thigh, left foot, left wrist). Figure 6 and Figure 7 respectively show classification performance for activities and intensities across these configurations.

Figure 6 goes here

Figure 7 goes here

Overall, the lower-body configuration demonstrates exceptional performance, even surpassing the full-body setup in certain cases. Particularly in intensity classification, for both SVM and ResNet1D, accuracy achieved using only lower-body inputs outperforms the full-body configuration in nearly all cases (with only three cases in SVM showing slight decreases in accuracy $\leq 0.1\%$). The slight drop in accuracy of the full-body setup can be attributed to two factors: redundancy in certain activities and the curse of dimensionality. On one hand, the activity protocol is dominated by lower-body driven locomotion (e.g., running, cycling), in which upper-body movements are often rhythmically coupled with lower-body motion, providing limited additional discriminative information. On the other hand, expanding the sensor set from 7 to 17 significantly increases the feature space dimension, which may introduce noise or lead to overfitting, thereby degrading generalization performance compared to more compact, informative sensor subsets. The L5-thigh-shank configuration also achieves strong results, exceeding 90% accuracy on both tasks, indicating its potential as a lightweight yet effective sensor setup. By contrast, the upper-body configuration shows notable performance drops: 7.4±2.1% lower accuracy in activity recognition and 6.9±1.6% lower in intensity recognition compared to full-body configuration. The observed accuracy drop can be explained by the fact that upper-body motion contributes less to distinguishing the predominantly locomotor activities in this dataset. Single-IMU setups exhibit the steepest declines in performance, with the wrist-only condition performing worst. However, the thigh-only and the foot-only setups yield moderate accuracies, particularly when paired with deep learning models. Notably, ResNet1D mostly outperforms traditional classifiers in single-sensor conditions, highlighting the capability of deep architectures to extract richer temporal features from limited inputs.

The above results demonstrate that full-body sensor coverage is not always necessary for robust HAR. Carefully selected subsets of sensors, can give rise to competitive performance while substantially reducing hardware requirements, setup complexity, and energy cost. This finding has practical implications for the design of lightweight, efficient HAR systems suitable for real-world deployment.

Usage Notes

Data Storage and Processing. All participant data are stored directly within the dataset directory. Anthropometric metadata is stored in a single file, *anthropometric_information.csv*. To support reproducible pre-processing, we provide scripts alongside the dataset. The script *extract_features.py* processes raw signals to compute all 18 handcrafted features across eight window sizes (0.5s~10s), generating participant-specific feature files (*features_P#.csv*) stored in a separate feature directory. For traditional machine learning, *train_base_traditional_models.py* trains SVM and RF classifiers across all feature subsets, window sizes, and sensor configurations. For deep learning experiments, *train_base_nn.py* first segments raw data into eight window sizes and stores training samples (*P#_X.npy*), activity labels (*P#_Y_act.npy*), and intensity labels (*P#_Y_int.npy*) in subdirectories organized by window sizes within the dataset directory. It then trains ResNet1D and LSTM models across all combinations of window sizes and sensor configurations.

Limitations. While the dataset provides a rich set of full-body inertial measurements for a diverse set of daily activities, certain limitations should be noted. All recordings were recorded in a controlled indoor environment, and all participants were young healthy adults (age range: 18-32 years) without mobility impairments. As a result, generalizability to outdoor environments, older populations, children, or individuals with disabilities remains untested. These factors should be considered when applying the

dataset to real-world scenarios. Consequently, applying models trained on this data to populations with markedly different activity patterns (e.g., elderly or clinical groups) would require domain adaptation techniques like transfer learning. To address this limitation and enhance generalizability, future versions of this dataset plan to incorporate participants across a broader age range. Furthermore, this dataset focuses on exercises and locomotion. Fine-grained Activities of Daily Living (ADLs) involving complex hand-object interactions, such as cooking or typing, are not included. Consequently, this dataset is more suitable for research into gaits and whole-body exercises rather than fine-grained gesture analysis.

Research directions. This dataset is intended as a flexible benchmark for a wide range of HAR research problems. It focuses on basic postures, locomotion, and exercise activities. Beyond standard activity recognition, the inclusion of structured intensity levels enables the study of effort-based classification tasks, which have been less commonly explored. The comprehensive full-body coverage further supports investigations into sensor subsets optimization, enabling researchers to evaluate the trade-off between recognition accuracy and deployment cost. Such studies could leverage advanced methods including reinforcement learning, attention-based feature selection, or multi-task learning. More broadly, the dataset may contribute to advancing research in transfer learning, personalized HAR models, and domain adaptation across sensor layouts and populations.

Data Availability

The dataset described in this study is available on Figshare[29] (DOI: 10.6084/m9.figshare.30234940) under a non-commercial license. The repository contains CSV files for each participant's trial and a CSV file providing anthropometric information.

Code Availability

All project code including data segmentation, feature extraction, and model training process is released under a non-commercial license on the project's repository at https://github.com/FudanBSRL/Comprehensive-IMU-Dataset.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (Grant No. 12532002) and the Shanghai Pilot Program for Basic Research - Fudan University, China (Grant No. 21TQ1400100-22TQ009).

Author contributions

M.F. and H.F. designed the experimental protocol; M.F. performed data collection, annotation, technical validation, and drafted the manuscript. Q.Z. contributed to discussion of results and helped draft the manuscript; H.F. conceived of the study, supervised the study at all stages, commented on the approach and results, and critically revised the manuscript. All authors gave final approval for publication.

Competing interests

The authors declare no competing interests.

References

1. Wang, Y., Cang, S. & Yu, H. A survey on wearable sensor modality centred human activity recognition in health care. *Expert Syst. Appl.* **137**, 167–190, https://doi.org/10.1016/j.eswa.2019.04.057 (2019).

2. Bisio, I., Garibotto, C., Lavagetto, F. & Sciarrone, A. When eHealth Meets IoT: A Smart Wireless System for Post-Stroke Home Rehabilitation. *IEEE Wirel. Commun.* **26**, 24–29, https://doi.org/10.1109/MWC.001.1900125 (2019).
3. Alhazmi, A. K. *et al.* Intelligent Millimeter-Wave System for Human Activity Monitoring for Telemedicine. *Sensors* **24**, 268, https://doi.org/10.3390/s24010268 (2024).
4. Bock, M., Kuehne, H., Van Laerhoven, K. & Moeller, M. WEAR: An Outdoor Sports Dataset for Wearable and Egocentric Activity Recognition. **8**, 175, https://doi.org/10.1145/3699776 (2023).
5. Liang, L. *et al.* WMS: Wearables-Based Multisensor System for In-Home Fitness Guidance. *IEEE Internet Things J.* **10**, 17424–17435, https://doi.org/10.1109/JIOT.2023.3274831 (2023).
6. Czekaj, Ł. *et al.* Real-Time Sensor-Based Human Activity Recognition for eFitness and eHealth Platforms. *Sensors* **24**, 3891, https://doi.org/10.3390/s24123891 (2024).
7. Hiremath, S. K. & Plötz, T. The Lifespan of Human Activity Recognition Systems for Smart Homes. *Sensors* **23**, 7729, https://doi.org/10.3390/s23187729 (2023).
8. Li, Y., Yang, G., Su, Z., Li, S. & Wang, Y. Human activity recognition based on multienvironment sensor data. *Inf. Fusion* **91**, 47–63, https://doi.org/10.1016/j.inffus.2022.10.015 (2023).
9. Qureshi, T. S., Shahid, M. H., Farhan, A. A. & Alamri, S. A systematic literature review on human activity recognition using smart devices: advances, challenges, and future directions. *Artif. Intell. Rev.* **58**, 276, https://doi.org/10.1007/s10462-025-11275-x (2025).
10. Gomaa, W. & Khamis, M. A. A perspective on human activity recognition from inertial motion data. *Neural Comput. Appl.* **35**, 20463–20568, https://doi.org/10.1007/s00521-023-08863-9 (2023).
11. Zappi, P. *et al.* Activity recognition from on-body sensors: Accuracy-power trade-off by dynamic sensor selection. in *Verdone, R. (eds) Wireless Sensor Networks* vol. 4913 17–33, https://doi.org/10.1007/978-3-540-77690-1_2 (Springer, Berlin, 2008).
12. Reiss, A. & Stricker, D. Introducing a new benchmarked dataset for activity monitoring. in *2012 16th international symposium on wearable computers* 108–109, https://doi.org/10.1109/ISWC.2012.13 (IEEE, Newcastle, UK, 2012).
13. Reiss, A. & Stricker, D. Creating and benchmarking a new dataset for physical activity monitoring. in *PETRA '12: Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments* https://doi.org/10.1145/2413097.2413148 (Association for Computing Machinery, New York, NY, USA, 2012).
14. Roggen, D. *et al.* Collecting complex activity datasets in highly rich networked sensor environments. in *2010 Seventh International Conference on Networked Sensing Systems (INSS)* 233–240, https://doi.org/10.1109/INSS.2010.5573462 (IEEE, Kassel, Germany, 2010).
15. Kwapisz, J. R., Weiss, G. M. & Moore, S. A. Activity recognition using cell phone accelerometers. *ACM SIGKDD Explor. Newsl.* **12**, 74–82, https://doi.org/10.1145/1964897.1964918 (2011).
16. Anguita1, D., Ghio, A., Oneto, L., Parra, X. & Reyes-Ortiz, J. L. A public domain dataset for human activity recognition using smartphones. in *ESANN 2013 Proceedings, 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (Bruges, Belgium, 2013).
17. Chan, S. *et al.* CAPTURE-24: A large dataset of wrist-worn activity tracker data collected in the wild for human activity recognition. *Sci. Data* **11**, 1135, https://doi.org/10.1038/s41597-024-03960-3 (2024).
18. Davoudi, A. *et al.* The effect of sensor placement and number on physical activity recognition and energy expenditure estimation in older adults: Validation study. *JMIR mHealth uHealth* **9**, e23681, https://doi.org/10.2196/23681 (2021).

19. Dang, X., Li, W., Zou, J., Cong, B. & Guan, Y. Assessing the impact of body location on the accuracy of detecting daily activities with accelerometer data. *iScience* **27**, 108626, https://doi.org/10.1016/j.isci.2023.108626 (2024).
20. Urukalo, D., Nates, F. M. & Blazevic, P. Sensor placement determination for a wearable device in dual-arm manipulation tasks. *Eng. Appl. Artif. Intell.* **137**, 109217, https://doi.org/10.1016/j.engappai.2024.109217 (2024).
21. Tsukamoto, A., Yoshida, N., Yonezawa, T., Mase, K. & Enokibori, Y. Where Are the Best Positions of IMU Sensors for HAR ? - Approach by a Garment Device with Fine-Grained Grid IMUs -. in *UbiComp/ISWC '23 Adjunct: Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing* 445–450, https://doi.org/10.1145/3594739.3610736 (Association for Computing Machinery, New York, NY, USA, 2023).
22. Ainsworth, B. E. *et al.* Compendium of physical activities: An update of activity codes and MET intensities. *Med. Sci. Sports Exerc.* **32**, https://doi.org/10.1097/00005768-200009001-00009 (2000).
23. Ainsworth, B. E. *et al.* 2011 compendium of physical activities: A second update of codes and MET values. *Med. Sci. Sports Exerc.* **43**, 1575–1581, https://doi.org/10.1249/MSS.0b013e31821ece12 (2011).
24. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830, (2011).
25. He, K., Zhang, X., Ren, S. & Jian, S. Deep Residual Learning for Image Recognition. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778, https://doi.org/10.1002/chin.200650130 (2016).
26. Chen, Z., Zhang, L., Cao, Z. & Guo, J. Distilling the Knowledge from Handcrafted Features for Human Activity Recognition. *IEEE Trans. Ind. Informatics* **14**, 4334–4342, https://doi.org/10.1109/TII.2018.2789925 (2018).
27. Saeb, S., Lonini, L., Jayaraman, A., Mohr, D. C. & Kording, K. P. The need to approximate the use-case in clinical machine learning. *Gigascience* **6**, 1–9, https://doi.org/10.1093/gigascience/gix019 (2017).
28. Paszke, A. *et al.* PyTorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **32**, (2019).
29. Feng, M., Zhang, Q. & Fang, H. A Comprehensive IMU Dataset for Evaluating Sensor Layouts in Human Activity and Intensity Recognition. *Figshare* https://doi.org/10.6084/m9.figshare.30234940 (2025).

Figure legends

Fig. 1 Overview of the dataset creation and benchmark: (a) Placement of 17 IMU sensors on the participant's body, including the A-pose and T-pose used for calibration. (b) Data transmission setup, including wireless connection of sensors to the receiver and data streaming to a laptop. (c) The twelve activities included in the dataset. (d) Intensity levels of these activities are defined based on METs and grouped into four levels: sedentary, light, moderate, and vigorous. (e) Signal processing pipeline. Raw accelerometer and gyroscope signals are segmented into overlapping windows (50% overlap) of various lengths (0.5s to 10s). Handcrafted features are extracted from each window and combined into four feature sets of increasing dimensions. These feature vectors or raw signals are then fed into different models: random forest, support vector machine, ResNet1D, and LSTM. All models are evaluated using subject-wise 5-fold cross-validation. (f) Sensor configurations, including the full-body setup and six reduced configurations: upper-body, lower-body, L5-thigh-shank, left-thigh, left-foot, and left-wrist.

Fig. 2 Tri-axial (a) acceleration and (b) angular velocity data from 6 sensors captured within a 3-second window from participant P01 during a moderate walking trial.

Fig. 3 Accuracy across different window sizes for activity classification. (a) RF using handcrafted features (4, 7, 9, 18) and raw data. (b) SVM using handcrafted features (4, 7, 9, 18) and raw data (note that SVM using raw input exhibits substantially lower accuracy and is therefore shown in an embedded inset for better visibility). (c) Deep learning models (LSTM and ResNet1D) trained on raw input.

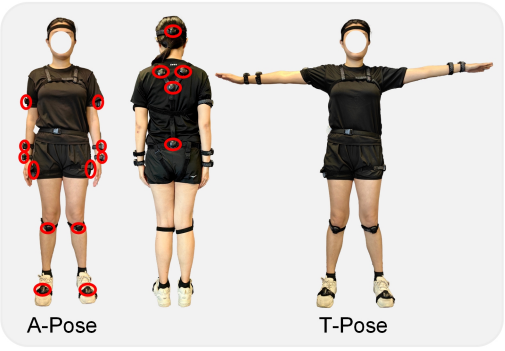
Fig. 4 Normalized confusion matrices (in %) for (a) SVM with 18 features and (b) ResNet1D with raw input, both using a 2-second window in the activity classification task.

Fig. 5 Accuracy across different window sizes for intensity classification. (a) RF using handcrafted features (4, 7, 9, 18) and raw data. (b) SVM using handcrafted features (4, 7, 9, 18) and raw data (note that SVM using raw input exhibits substantially lower accuracy and is therefore shown in an embedded inset for better visibility). (c) Deep learning models (LSTM and ResNet1D) trained on raw input.

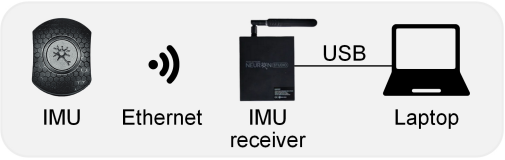
Fig. 6 Accuracy for activity classification under different sensor configurations. (a) SVM trained with 18 features using seven sensor configurations: full-body (17 IMUs), upper-body (10 IMUs), lower-body (7 IMUs), L5-thigh-shank (3 IMUs), left-thigh (one IMU), left-foot (one IMU) and left-wrist (one IMU). (b) ResNet1D trained on raw input using the same seven sensor configurations.

Fig. 7 Accuracy for intensity classification under different sensor configurations. (a) SVM trained with 18 features using seven sensor configurations: full-body (17 IMUs), upper-body (10 IMUs), lower-body (7 IMUs), L5-thigh-shank (3 IMUs), left-thigh (one IMU), left-foot (one IMU) and left-wrist (one IMU). (b) ResNet1D trained on raw input using the same seven sensor configurations.

Data Acquisition

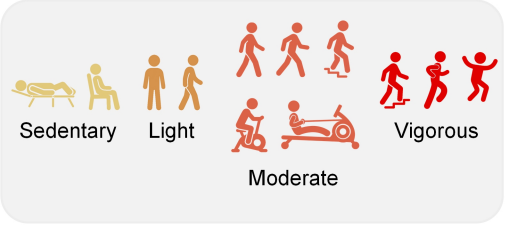


(a) Sensor placement



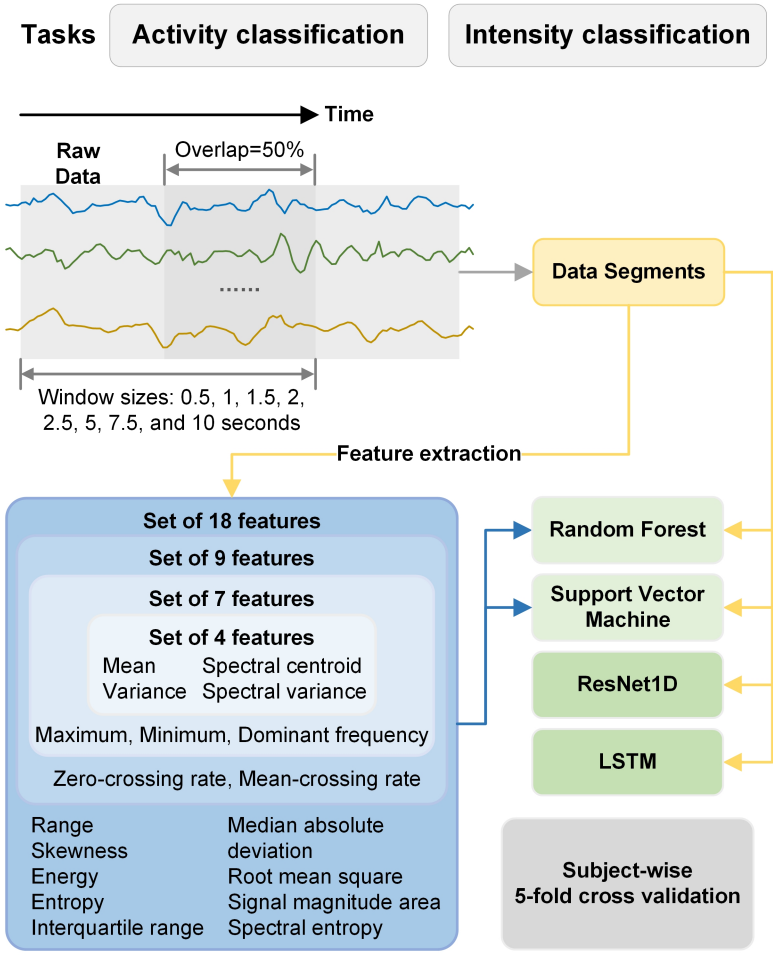
(b) Sensor connection diagram

(c) Activity categories

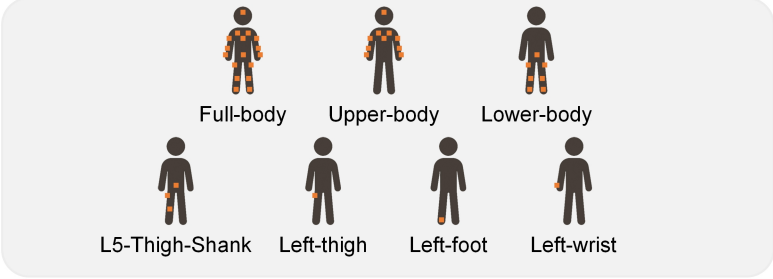


(d) Intensity levels

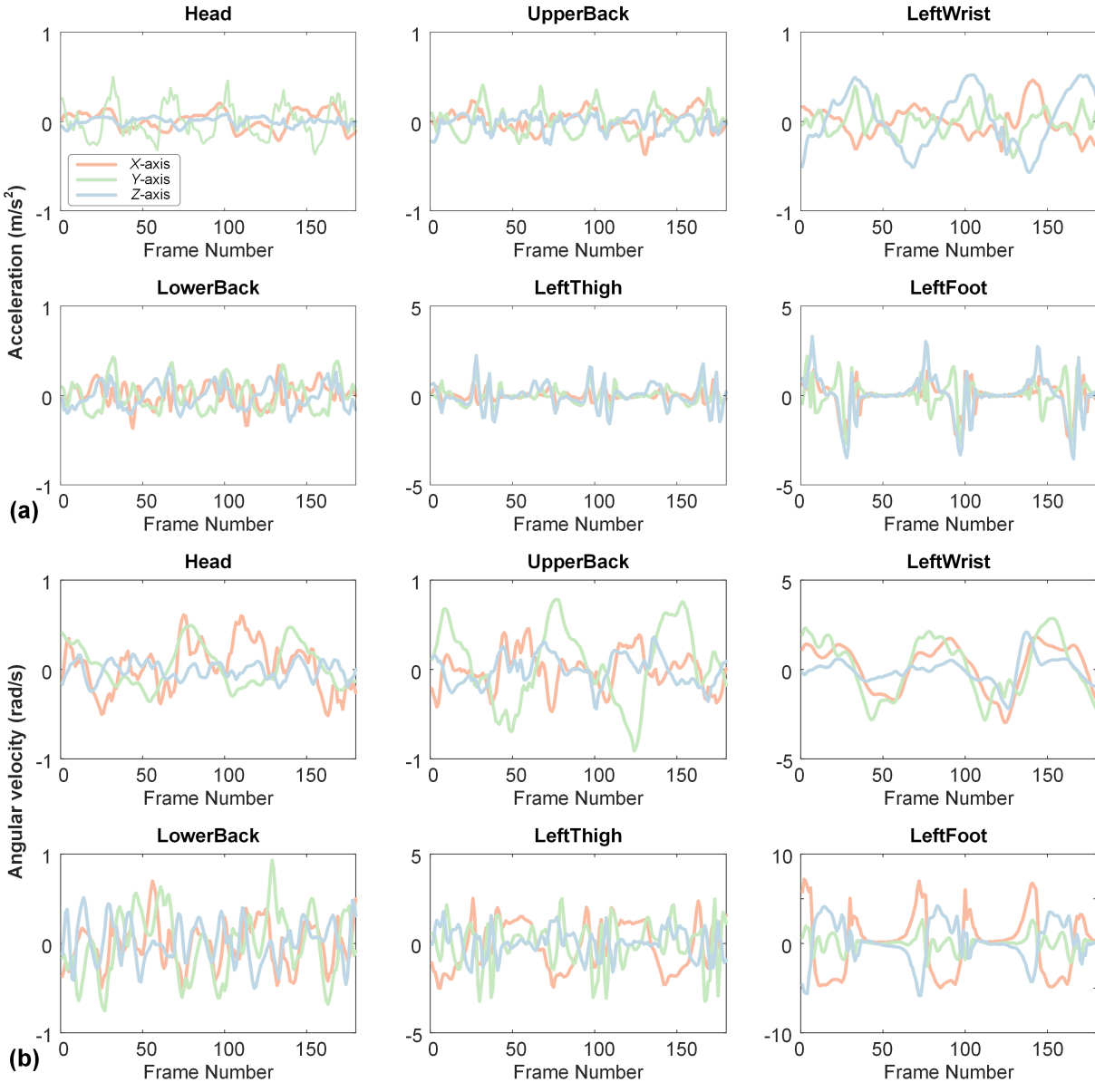
Benchmark

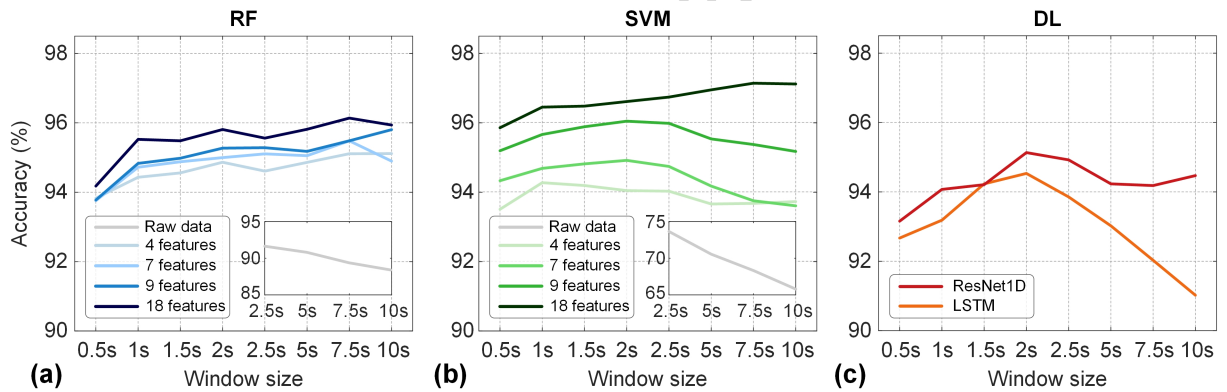


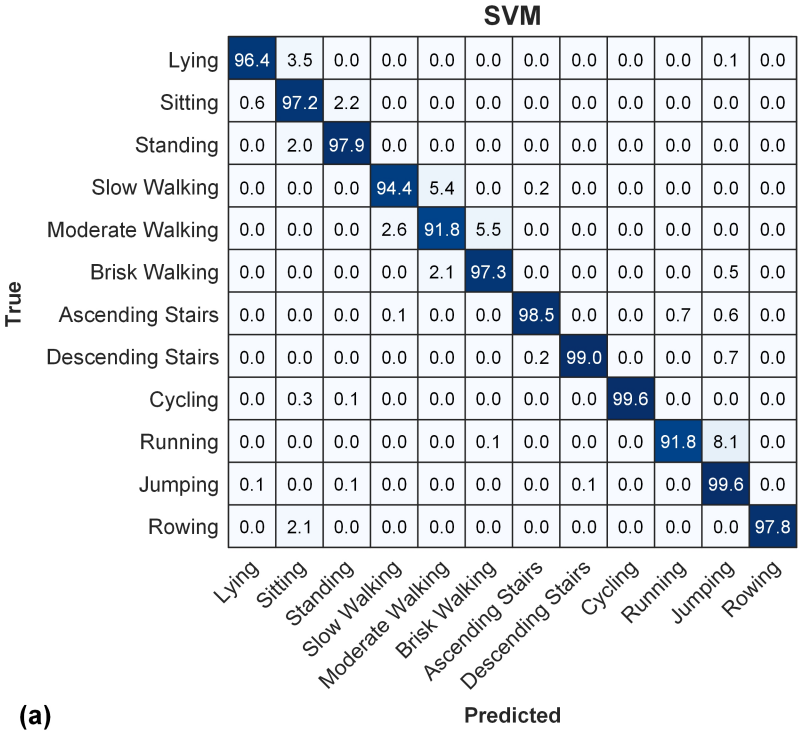
(e) Signal processing pipeline



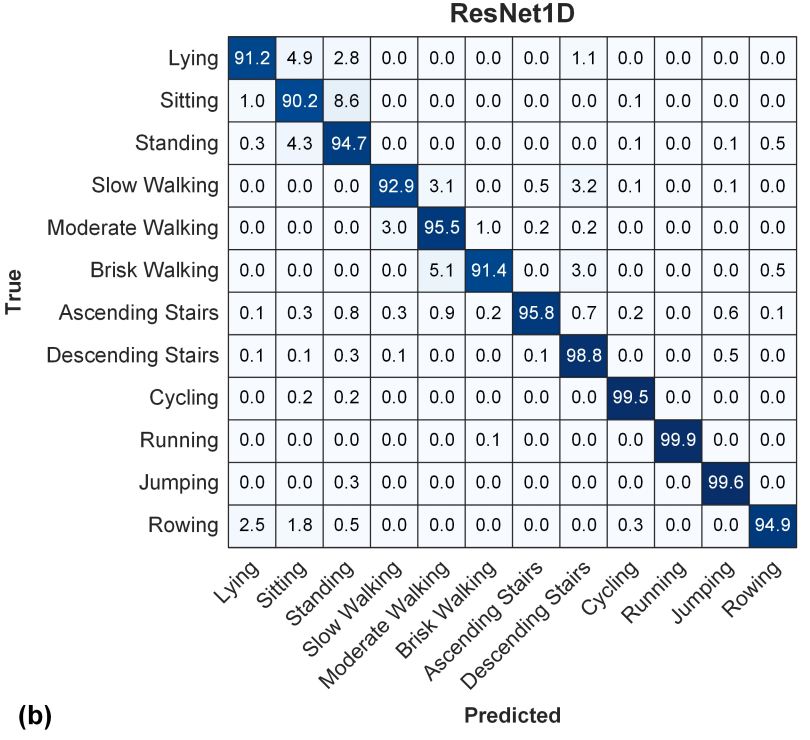
(f) Sensor configurations


Head
UpperBack
LeftWrist
LowerBack
LeftThigh
LeftFoot
X-axis
Y-axis
Z-axis
Acceleration (m/s²)
Frame Number
(a)
Angular velocity (rad/s)
(b)





(a)



(b)