Where to Sense Before What to Classify: A Full-Body Tracking Pilot Study on Real-Time HAR

Ruben Schlonsak [1,2], Marco Gabrecht [1,2], Jierui Li [1,3], and Denys J.C. Matthies [1,2]

[1] Technical University of Applied Sciences Lübeck, Lübeck, Germany
ruben.schlonsak@th-luebeck.de, denys.matthies@th-luebeck.de
[2] Fraunhofer IMTE, Lübeck, Germany
[3] East China University of Science and Technology, Shanghai, China

Abstract. Human Activity Recognition (HAR) research has increasingly focused on complex deep learning models, although real-time wearable deployments remain constrained by inference latency and by *where* sensing occurs on the body. In this controlled, subject-dependent pilot study with three participants, we use a full-body motion capture suit with 17 IMUs to systematically compare 1D-CNN and LSTM classifiers across six daily activities, five window sizes, individual coordinate axes, and twelve anatomical sites (nine bilateral, three midline). Our analysis yields three main observations. First, LSTMs achieve higher classification accuracy than CNNs (M=0.95 vs. M=0.86, aggregated across all five window sizes), whereas CNNs perform inference 3.6× faster, highlighting an accuracy–latency trade-off relevant to interactive HAR. Second, temporal window size affects the two architectures differently (CNN: r=+0.22; LSTM: r=−0.45, over 8–128 frames), indicating that window length should be optimized separately for each model family. Third, sensor placement substantially influences recognition performance: the bilateral foot location reaches 96% accuracy with both architectures, while the bilateral toe location achieves 97% with the CNN, exceeding the CNN full-body baseline at W=32 by four percentage points. A leave-one-subject-out analysis reveals the expected reduction in absolute accuracy under cross-subject evaluation; however, the best-performing distal sensor location remains within three percentage points of full-body input in every fold, and the CNN's subject-dependent advantage of the toe over full-body input does not transfer to held-out participants. Although these findings require validation with larger participant cohorts and deployable low-cost wearable sensors, they suggest that sensor placement and temporal windowing are critical design parameters for real-time HAR. We discuss implications for body-worn HAR system design and the limitations of transferring results obtained from motion-capture suits to practical wearable deployments.

Keywords: Human Activity Recognition · Wearable Sensors · Motion Capture · Sensor Placement · Real-Time Classification.

1 Introduction

Human Activity Recognition (HAR) has evolved from handcrafted feature engineering toward deep learning, enabling substantial improvements across a wide range of benchmark datasets. However, translating these advances into practical wearable systems remains challenging because real-time deployments are constrained not only by recognition accuracy, but also by prediction time, computational resources, energy consumption, and sensor placement. Recent surveys therefore argue that continued progress depends not only on increasingly sophisticated model architectures, but also on understanding *how and where* human movement should be sensed [20]. While CNNs, LSTMs, and more recently Transformer-based models dominate current HAR research, their growing complexity can impede deployment on resource-constrained wearable hardware [11,5]. At the same time, body location substantially influences recognition performance [10,30], while practical edge deployments additionally require careful selection of temporal window size and consideration of prediction time [9,14]. Although these factors have received considerable attention, they are commonly investigated in isolation. Architecture comparisons typically assume a fixed sensor configuration and temporal window, whereas sensor-placement studies often evaluate only one model or a limited set of preprocessing choices. Consequently, there is limited guidance for deployable HAR systems, in which model architecture, sensing location, and temporal context jointly determine recognition performance and runtime behavior. A full-body motion capture suit provides a controlled setting for investigating these interactions because all body segments are recorded simultaneously, allowing candidate locations to be compared on identical movement sequences.

In this paper, we investigate these design dimensions through a controlled three-participant pilot study using a full-body motion tracking suit equipped with 17 inertial measurement units (IMUs). We compare a 1D convolutional neural network (1D-CNN) and a long short-term memory network (LSTM) across six daily activities. We systematically vary temporal window size, coordinate representation, and body location to examine their effects on classification accuracy, host-side prediction time, and cross-subject robustness.

Our contributions are:

- We present a controlled pilot study comparing 1D-CNN and LSTM classifiers across six activities, five temporal window sizes, multiple coordinate representations, and twelve anatomical sites recorded from identical movement sequences.
- We quantify how architecture choice, temporal window size, and body location jointly influence classification accuracy and prediction time, revealing architecture-specific trade-offs relevant to real-time wearable HAR.
- We provide preliminary evidence that distal lower-limb representations preserve much of the discriminative information available from full-body input for the investigated locomotion and posture activities. We further assess the robustness of this observation using leave-one-subject-out evaluation and

derive implications for body-site selection, temporal windowing, and model choice.

2 Related Work

2.1 Deep Learning in HAR

Early HAR systems relied on handcrafted features combined with conventional machine learning algorithms such as Support Vector Machines (SVMs) and Random Forests [2,27,31]. Although effective in controlled settings, these approaches required substantial feature engineering and often generalized poorly across users and activities. The adoption of deep learning shifted HAR toward end-to-end learning from raw sensor streams, reducing manual feature design while improving robustness and scalability [15]. Zeng et al. first demonstrated convolutional neural networks (CNNs) for accelerometer-based HAR [38], while Ordóñez and Roggen introduced DeepConvLSTM, combining convolutional feature extraction with recurrent temporal modeling [24].

Subsequent work explored increasingly sophisticated network architectures. CNN-based approaches were extended to multichannel wearable signals [35,36] and image-like sensor representations [16], while hybrid CNN–RNN models leveraged recurrent layers to capture longer temporal dependencies [12,37]. More recently, attention mechanisms and Transformer architectures have further improved recognition performance by modeling long-range temporal relationships and adaptive sensor weighting [21,13,17,11,5]. Although these architectures frequently achieve excellent benchmark performance, they generally increase computational complexity, memory requirements, and inference cost, making deployment on resource-constrained wearable devices more challenging.

Despite these architectural advances, most studies primarily optimize recognition accuracy, whereas inference latency is reported less consistently. Since latency directly determines the responsiveness of interactive wearable systems, evaluating both recognition performance and computational efficiency is essential when assessing models intended for real-time deployment. We therefore compare CNN and LSTM architectures with respect to both accuracy and prediction time.

2.2 Sensor Placement

Sensor placement fundamentally determines which body dynamics can be observed and therefore strongly influences recognition performance. Previous studies consistently report that lower-body placements outperform many upper-body locations for locomotion-related activities because they directly capture gait dynamics [1,25]. Davoudi et al. compared five common wearable locations and found that a single hip-mounted accelerometer achieved the highest balanced accuracy, while additional sensors yielded only modest improvements [10]. Similar conclusions have been reported for waist, hip, thigh, and ankle placements across a variety of activity recognition tasks [7,26,29], as well as for fall detection,

where waist- and thigh-mounted sensors consistently outperform arm-mounted alternatives [32].

More specialized applications further emphasize the importance of distal lower-limb sensing. Ankle-mounted IMUs provide highly discriminative gait information [22,3], while recent work demonstrates that foot-mounted sensors can reconstruct lower-limb motion and characterize walking terrain with high fidelity [34]. At the same time, multimodal sensor fusion has been shown to improve recognition performance by combining complementary information from multiple sensing modalities [4,6,23].

A common limitation of sensor placement studies is that each candidate location typically requires a separate device and often an independent recording session. Consequently, different placements are evaluated on similar rather than identical movement instances, introducing additional variability. A full-body motion tracking suit overcomes this limitation by recording all body segments simultaneously, allowing different sensor locations to be compared using exactly the same movements. This controlled setting enables a more direct assessment of how body location alone affects recognition performance.

2.3 Real-Time Constraints

Real-time HAR systems must balance recognition accuracy against computational efficiency. For interactive wearable applications such as activity coaching, rehabilitation, or safety monitoring, overall system latency is often required to remain well below one second to provide timely feedback [18]. This latency depends on the complete processing pipeline, including sensor sampling, temporal windowing, preprocessing, and model inference [8].

Temporal window size represents one of the primary design parameters. Larger windows provide additional temporal context and can improve recognition accuracy but also increase response latency because more sensor samples must be collected before inference can begin [39,29,33]. Czekaj et al. demonstrated that reducing the window length substantially lowers latency while often maintaining competitive recognition performance [9]. However, window size introduces only one component of overall latency.

It is therefore useful to distinguish between two separate sources of delay. The first is the *accumulation delay*, which depends solely on sampling frequency and window length. The second is the *prediction time* required by the trained model, which depends on the network architecture and implementation. These two latency components are frequently discussed together despite representing different design trade-offs. In this work, we report them separately to better characterize the practical implications of model selection and window configuration.

Recent TinyML research has demonstrated that neural networks can be deployed on highly resource-constrained microcontrollers through quantization, pruning, and lightweight architectures [40,19,14]. Nevertheless, much of this work focuses on reducing model complexity after the model architecture has been selected. Comparatively less attention has been paid to whether appropriate sensor

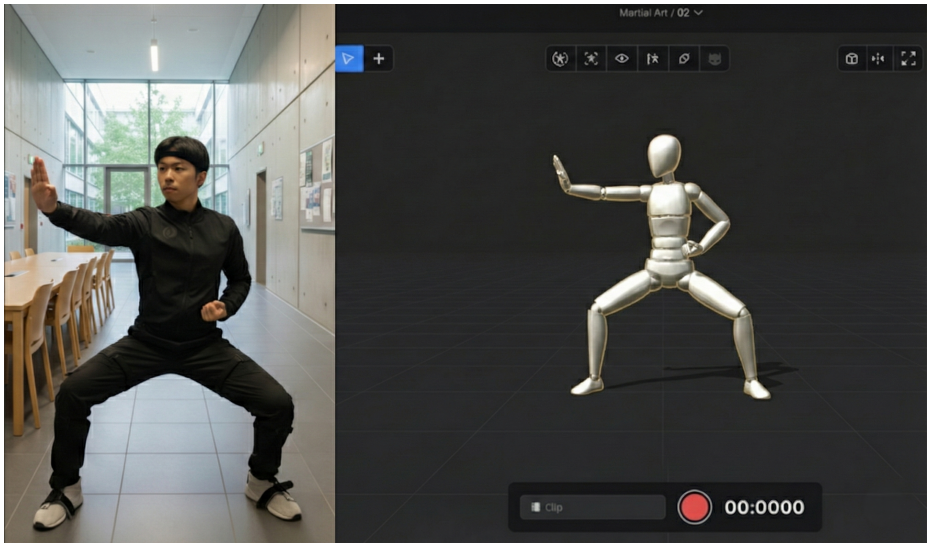


Fig. 1. The Rokoko Smartsuit Pro II used for data collection (left) and the corresponding avatar visualization in Rokoko Studio (right).

placement and architecture-specific window design can achieve substantial improvements before model compression becomes necessary. Our study investigates these design dimensions jointly under controlled experimental conditions.

3 Method

3.1 Apparatus and Data Collection

We recorded full-body motion using the Rokoko Smartsuit Pro II [28], shown in Fig. 1. The suit contains 17 inertial measurement units (IMUs), each combining an accelerometer, a gyroscope, and a magnetometer. Motion data are transmitted wirelessly to Rokoko Studio, which applies a proprietary sensor-fusion pipeline and estimates the global 3D positions of 21 body segments at 30 Hz. These segments comprise the head, neck, and hips, together with the left and right shoulder, arm, forearm, hand, thigh, shin, foot, toe, and toe tip.

Our models operate on the fused segment positions produced by Rokoko Studio rather than directly on raw accelerometer, gyroscope, or magnetometer signals. This representation places all segments in a common coordinate system and allows body locations to be compared on the same movement instances. However, sensor fusion and drift correction are delegated to the suit software. The resulting measurements should therefore be interpreted as suit-derived kinematic data rather than as direct measurements from independently deployed, low-cost IMUs.

Recordings were exported from Rokoko Studio as Biovision Hierarchy (BVH) files and converted to CSV using the `bvh-converter` library. Each exported table contained a timestamp, 63 position channels corresponding to 21 segments and three coordinate axes, and one auxiliary hip-height channel. Rokoko Studio reports positions in centimeters in a gravity-aligned global frame, in which the vertical y-axis points upwards. Figure 2 summarizes the resulting processing pipeline.

Three participants took part in the study. All participants were male, aged 22–23 years, and had body heights between 170 and 174 cm. The suit was calibrated individually for each participant before recording. The small and demographically homogeneous participant group reflects the pilot character of the study and limits the extent to which the results can be generalized to a broader population.

Each participant performed six activities indoors: jogging, jumping, lying, sitting, standing, and walking. The activity set includes periodic locomotion activities, namely jogging and walking; an impulsive movement, namely jumping; and three predominantly static postures, namely lying, sitting, and standing. For each participant and activity, approximately 20 s were recorded for training. An additional recording of approximately 9 s was collected separately for testing. Consequently, training and test windows did not originate from the same recording, although the same participants occurred in both sets.

Each activity recording was stored in a separate file and assigned its corresponding activity label. The exported streams contained no gaps or missing sensor frames. We therefore applied no additional signal filtering or missing-value imputation before coordinate transformation and segmentation.

3.2 Preprocessing and Segmentation

Rokoko Studio exports segment positions in a global coordinate system. Consequently, otherwise similar activities performed at different locations produce different absolute trajectories. We reduced this dependence on global position by applying first-order differencing to the two horizontal coordinate axes using `pandas.DataFrame.diff()`. The resulting values describe frame-to-frame displacement rather than absolute horizontal position. Differencing was applied independently within each activity recording, so no displacement was ever computed across a recording boundary; the single `NaN` produced at the first frame of each recording was dropped before segmentation.

Differencing provides translation invariance with respect to the starting position, but it does not make the signals invariant to rotations of the global coordinate system. Direction- and heading-dependent effects may therefore remain in the horizontal channels. The gravity-aligned vertical coordinate was additionally evaluated in its unmodified form because absolute segment height contains information that distinguishes postures such as standing, sitting, and lying. The resulting three-axis representation therefore combines relative horizontal displacement with absolute vertical position. Because absolute vertical height (in cm) and frame-to-frame horizontal displacement (in cm) differ in value

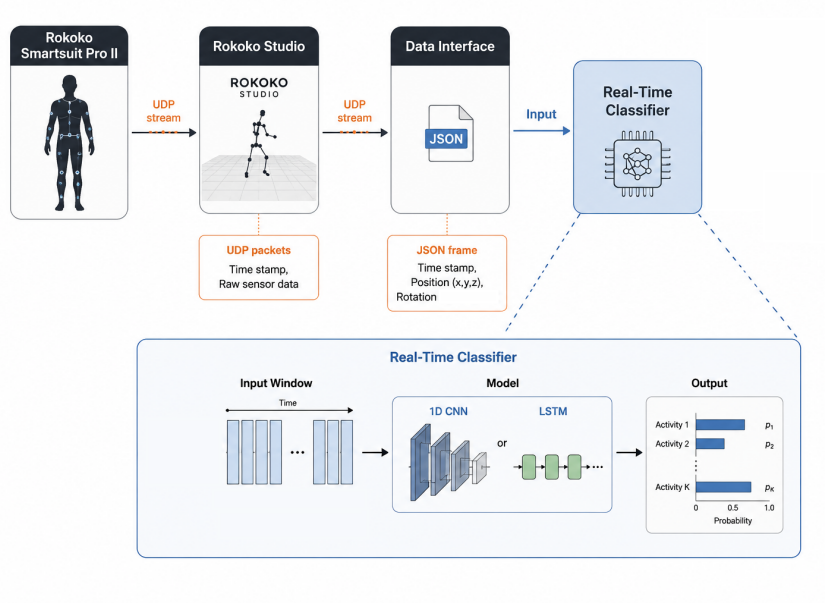


Fig. 2. Processing pipeline from motion recording with the Rokoko suit to window-based activity classification using a 1D-CNN or LSTM.

range and variance, all input channels were z-score normalized per channel using the training-set mean and standard deviation before being passed to the models.

For the full-body experiments, the model received the selected coordinate channels from all tracked body segments. The axis ablations retained only the channels associated with the evaluated coordinate axis. For the body-site experiments, channels were selected according to anatomical site. The 21 tracked segments correspond to twelve anatomical site types: three midline sites, namely head, neck, and hips, and nine bilateral sites, namely shoulder, arm, forearm, hand, thigh, shin, foot, toe, and toe tip. For a bilateral site, the channels from the left and right body sides were used jointly, thereby representing a candidate placement as a bilateral sensor configuration.

Activity streams were divided into fixed-length sliding windows with a stride of two samples. We evaluated five window sizes,

$$W \in \{8, 16, 32, 64, 128\} \text{ frames},$$

corresponding to approximately 0.27, 0.53, 1.07, 2.13, and 4.27 s at 30 Hz. Windows were generated independently within each activity recording and therefore never crossed activity or recording boundaries. Unless stated otherwise, the reference configuration used W=32, corresponding to approximately one second of motion data.

Because recording duration was approximately fixed, larger windows produced fewer training examples. At the evaluated window sizes, each participant's per-activity training set contained approximately 318, 315, 305, 290, and 258 windows, and the corresponding per-activity test set contained approximately 134, 130, 121, 106, and 74 windows. Aggregated over the three participants and six activities, the subject-dependent test split therefore comprised approximately 2,412, 2,340, 2,181, 1,908, 1,332 windows across the five window sizes; the W=32 reference configuration thus contained 2,181 test windows. Consecutive windows are also strongly overlapping, particularly at larger window sizes. The generated windows should therefore not be interpreted as statistically independent observations. Repeated training runs characterize optimization variability, whereas participant-level generalization is examined separately.

3.3 Models

We compared two conventional architectures for sensor-based HAR, a one-dimensional convolutional neural network (1D-CNN) and a long short-term memory network (LSTM). The 1D-CNN consists of two temporal convolutional layers with 16 filters, kernel size 3, and ReLU activations, followed by max pooling, a dense layer with 100 ReLU units, dropout (0.5), and a six-class softmax output. It contains 26,978 trainable parameters.

The LSTM uses a recurrent layer with 64 units, followed by dropout (0.5), a dense layer with 64 ReLU units, and a six-class softmax output. It contains 37,574 trainable parameters.

Both models were implemented in TensorFlow and trained for 50 epochs with a batch size of 64 using categorical cross-entropy and Adam with its default learning rate of 10^{-3}. We intentionally used established architectures without extensive task-specific optimization to examine how the two model families respond to changes in window size, coordinate representation, and body location.

3.4 Evaluation Protocol

The evaluation was organized around four hypotheses:

- **H1:** The 1D-CNN provides a more favorable inference-time profile than the LSTM for real-time activity classification.
- **H2:** Increasing the temporal window size improves classification accuracy.
- **H3:** Selected body sites retain sufficient activity information to approach the recognition performance obtained from full-body input.
- **H4:** For at least one of the two architectures, an axis-specific coordinate representation can match or exceed the accuracy obtained when using all three coordinate axes.

The primary evaluation followed a subject-dependent protocol. Training and test windows originated from separate recordings, but all three participants were

represented in both sets. The resulting measurements therefore characterize generalization to a separate recording session involving previously observed participants. They should not be interpreted as evidence of generalization to unseen users. We examine the latter separately using leave-one-subject-out evaluation in Sect. 4.5.

Because neural network optimization is stochastic, each configuration was trained and evaluated ten times with independent random seeds. This procedure was applied to the model and window-size comparison as well as to the coordinate-axis and body-site ablations. We report the mean (M) and standard deviation (SD) of classification accuracy across the ten runs. These statistics describe variability across repeated optimization runs and do not represent variability across a broader participant population. Confusion matrices are reported for the reference window size of W=32.

In addition to classification accuracy, we measured the prediction time per window. Timing measurements were performed on a consumer laptop with an Intel Core i7-1165G7 processor and 16 GB of RAM, running Windows 11, Python 3.10, and TensorFlow 2.12. Inference was executed on the CPU without GPU acceleration. Prediction time was measured as the wall-clock duration of a single-window inference call using a batch size of one. We excluded 100 initial warm-up calls and then averaged the measurements over the test set. Timing was repeated for each of the ten trained instances of a configuration.

The reported prediction times include overhead introduced by Python and the TensorFlow inference call. They therefore support a relative comparison between the two architectures on the evaluated platform, but they do not estimate the latency achievable after hardware-specific optimization or deployment on a microcontroller.

3.5 Live Classification Pipeline

We additionally implemented a live classifier to verify that the processing pipeline could operate on streaming motion data. The program loads a trained model, connects to Rokoko Studio, and receives motion frames as UDP packets at 30 Hz. Each packet contains JSON-encoded timestamp and position values. Incoming frames are parsed and buffered until the selected temporal window has been filled, after which the resulting window is passed to the model for inference.

For an initially empty buffer, the earliest possible prediction requires the accumulation of W frames. The corresponding context-accumulation time is

$$t_{\text{window}} = \frac{W}{f_s},$$

where f_s=30 Hz. The earliest first output is therefore additionally affected by preprocessing and model inference time. This accumulation delay is conceptually distinct from the model prediction time reported in the experiments. It is also distinct from the interval between successive predictions once a streaming buffer has been populated.

Table 1. Aggregate model performance across the five window-size conditions. Each model–window-size configuration was evaluated over 10 runs. M denotes the mean across window-size conditions, while SD is the mean within-condition standard deviation and therefore reflects run-to-run optimization variability rather than variation across window sizes.

Model	Accuracy		Prediction time [s]	
	M	SD	M	SD
1D-CNN	0.86	0.0064	**0.140**	0.0002
LSTM	**0.95**	**0.0005**	0.498	0.0046

The live implementation was used as a functional proof of concept and produced activity predictions for each of the six classes during informal tests. However, we did not instrument the complete path from sensor acquisition to prediction output. The reported timing results therefore characterize only model inference and should not be interpreted as measurements of end-to-end system latency.

3.6 Ethics and Consent

All participants were adults and provided written informed consent before data collection. The participant shown in Fig. 1 additionally consented to publication of the photograph. Under the applicable local institutional regulations, this non-invasive pilot study did not require formal ethics approval.

4 Results

Given the pilot character of the study and the sample of three participants, we treat the following analyses as descriptive. Unless stated otherwise, means (M) and standard deviations (SD) summarize ten repeated training runs of the same configuration. The reported standard deviations therefore quantify variability introduced by stochastic model optimization; they do not represent uncertainty across participants.

4.1 CNN vs. LSTM: Accuracy and Prediction Time

Table 1 summarizes model performance across all five window sizes. The LSTM achieved higher mean classification accuracy than the 1D-CNN (M=0.95 vs. M=0.86). Its reported run-level variability was also lower (SD=0.0005 vs. SD=0.0064), indicating that repeated optimization runs produced more similar results for the LSTM under the evaluated configurations.

The accuracy difference was accompanied by a substantial difference in prediction time. The 1D-CNN required an average of 0.140 s per window, compared with 0.498 s for the LSTM. The resulting ratio, $0.498/0.140 \approx 3.6$, indicates that

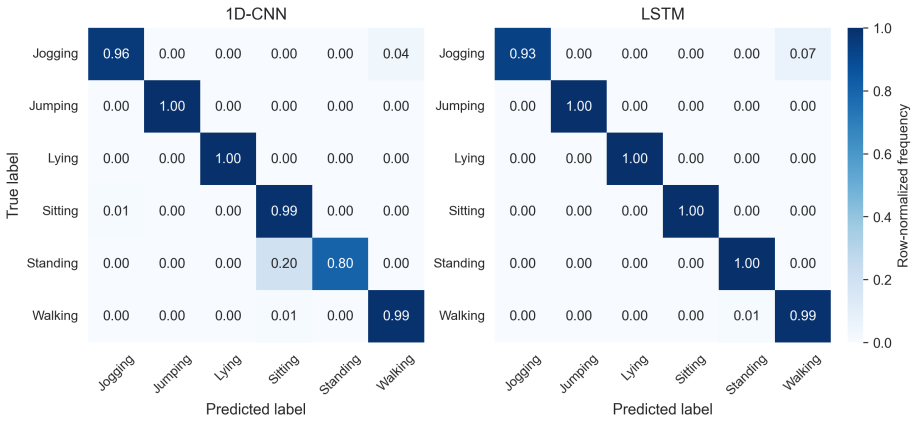


Fig. 3. Row-normalized confusion matrices for a single 1D-CNN run (left) and a single LSTM run (right) at W=32 with full-body input. The matrices illustrate the class-level error structure; absolute accuracies are reported in Tables 2 and 4. A prominent error involves standing versus sitting for the CNN and jogging versus walking for both models.

the CNN executed approximately 3.6× faster on the evaluated host platform. As described in Sect. 3, these measurements characterize relative TensorFlow inference time on the test laptop rather than end-to-end latency or performance on embedded hardware.

The values in Table 1 are aggregated across window sizes and should not be conflated with the reference results at W=32 used in the subsequent ablations. At that reference configuration, the full-body accuracies were 0.93 for the 1D-CNN and 0.98 for the LSTM, as reported in Table 4.

Figure 3 illustrates the class-specific error patterns for one run at the reference configuration. The most frequent confusion for the CNN was standing classified as sitting. Both models also confused some jogging windows with walking, while lying and jumping were classified without error in the shown run. As the matrices represent one run rather than an aggregation across repetitions, they illustrate the error structure but not its run-to-run variability.

With respect to **H1**, the results support the hypothesized prediction-time advantage of the 1D-CNN. However, they do not support a broader interpretation that the CNN is categorically better suited to the task, because its lower prediction time is accompanied by lower recognition accuracy. The choice between the two architectures therefore depends on the relative importance of classification performance and model-computation latency.

4.2 Window Size: Architecture-Specific Effects

Window size affected the two model families differently. Table 2 reports the per-window-size means over the ten runs for both architectures. We report de-

Table 2. Per-window-size accuracy and prediction time at full-body input (M over 10 runs).

	1D-CNN		LSTM	
W	Accuracy	Pred. time [s]	Accuracy	Pred. time [s]
8	0.84	0.108	0.96	0.342
16	0.81	0.125	0.92	0.421
32	0.93	0.141	0.98	0.509
64	0.85	0.156	0.95	0.578
128	0.87	0.172	0.92	0.642

scriptive Pearson correlations between window length and mean accuracy over $W \in \{8, 16, 32, 64, 128\}$.

For the 1D-CNN, window size showed a weak positive association with accuracy (r=+0.22). In contrast, LSTM accuracy showed a negative association with window size (r=−0.45). Neither model exhibited a monotonic accuracy trajectory across the evaluated window sizes. Larger windows therefore did not consistently improve recognition performance.

Both trajectories were shaped by individual conditions rather than by a smooth trend. For the 1D-CNN, the W=32 condition lies 6 to 12 percentage points above every other condition, substantially exceeding the run-to-run variability of that configuration (SD=0.0041; Table 4). Excluding this condition, the remaining CNN values increase monotonically from W=16 onwards, and the correlation rises to r=+0.84. The weak overall coefficient therefore results primarily from the high accuracy observed at the intermediate window size. The LSTM likewise attained its maximum at W=32. Because the correlations are based on only five conditions and are sensitive to individual values, we base our interpretation primarily on the observed non-monotonicity rather than on the magnitude of r. It should also be noted that W=32, which is used for all subsequent ablations, produced the highest accuracy for both architectures.

Prediction time increased monotonically with window size for both architectures, with a strong association for the 1D-CNN (r=0.93) and the LSTM (r=0.91). The recurrent model incurred the larger absolute prediction-time cost across all conditions, while both architectures required more computation for longer input sequences.

Window length additionally imposes an accumulation delay independently of model inference. At the 30 Hz sampling rate, the evaluated windows correspond to approximately 0.27 s at W=8, 0.53 s at W=16, 1.07 s at W=32, 2.13 s at W=64, and 4.27 s at W=128. These durations specify the minimum amount of temporal context that must be collected for an initially empty window before a first prediction can be produced.

Overall, **H2** is not supported as a general hypothesis. The CNN exhibited only a weak positive relationship between window length and accuracy, whereas the corresponding relationship was negative for the LSTM. The results instead

indicate that window size must be selected separately for each architecture and evaluated jointly with its accumulation and prediction-time costs.

4.3 Body-Site Ablation

To evaluate **H3**, we trained both models at W=32 using one anatomical site type at a time. Because the full-body suit recorded all segments simultaneously, each site was evaluated on the same underlying movement instances. For paired anatomical locations, we used the bilateral representation defined in Sect. 3.2. The resulting conditions therefore represent nine paired site types and three midline sites; they do not characterize strictly one-sided sensing.

Figure 4 summarizes the per-site results, and Table 3 reports the corresponding means and standard deviations. Averaged across the twelve evaluated sites, the LSTM reached an accuracy of 0.92, compared with 0.89 for the 1D-CNN. This ordering is consistent with the aggregate full-body comparison, although the magnitude of the difference varied substantially by site.

The strongest results were obtained from distal lower-limb locations. For the 1D-CNN, the toe representation achieved the highest accuracy (M=0.97, SD=0.0003), followed by the foot (M=0.96, SD=0.0007). Relative to the CNN full-body reference of 0.93 at W=32, four of the twelve sites achieved higher accuracy. The toe exceeded the full-body reference by four percentage points.

For the LSTM, the foot produced the highest site-specific accuracy (M=0.96, SD=0.0008), followed by the toe, thigh, and shoulder at 0.95. These values remained close to, but did not exceed, the LSTM full-body reference of 0.98. The foot condition was two percentage points below the full-body input.

The hand was the weakest site for both architectures, reaching 0.74 with the CNN and 0.85 with the LSTM. The next-lowest CNN results were obtained from the forearm (M=0.81) and shin (M=0.87), whereas the next-lowest LSTM results were obtained from the shin (M=0.86) and hips (M=0.87).

The range across body sites differed between architectures. CNN accuracy varied by 23 percentage points, from 0.74 at the hand to 0.97 at the toe. LSTM accuracy varied by 11 points, from 0.85 at the hand to 0.96 at the foot. Within this activity set, the LSTM was therefore less sensitive to the selected input location, whereas CNN performance depended more strongly on body site. The low standard deviations for the toe and foot conditions ($SD \leq 0.0008$ for both models) further indicate that their high accuracies were reproduced consistently across the ten optimization runs.

These subject-dependent results support **H3**: selected body sites approached the recognition performance of full-body input, and the best site exceeded the full-body reference for the CNN. However, the evaluated inputs are suit-derived segment positions from a bilateral representation. The result should therefore be interpreted as evidence from a controlled body-site ablation, not as validation that an independently mounted, low-cost IMU at the same location would achieve identical performance.

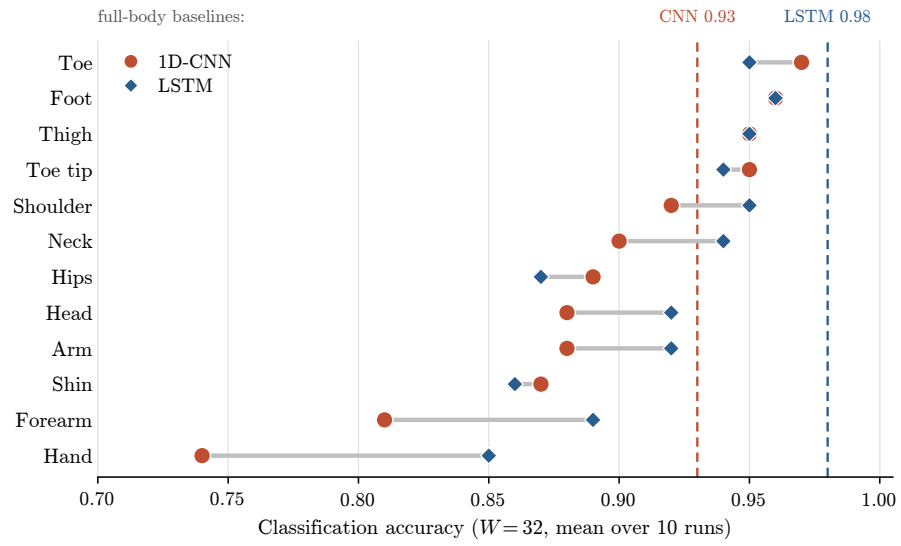


Fig. 4. Per-site classification accuracy at W=32, averaged over 10 runs. Sites are sorted by the CNN mean, and the dashed line marks the CNN full-body reference of 0.93. Four sites exceed this reference. The LSTM full-body reference is 0.98 and is not drawn. Exact means and standard deviations are reported in Table 3.

4.4 Coordinate Representation

We next evaluated each coordinate axis separately at W=32 using full-body input. As described in Sect. 3.2, the horizontal coordinates were represented through frame-to-frame displacement, whereas the gravity-aligned vertical coordinate retained absolute segment height. The axis ablation therefore evaluates both the information carried by each spatial dimension and the associated coordinate representation.

Table 4 reports the resulting accuracies. The vertical y-axis was the strongest individual axis for both architectures. With y-only input, the CNN achieved M=0.85 (SD=0.0026), whereas the LSTM achieved M=0.98 (SD=0.0002). At the reported two-decimal precision, the latter matched the LSTM's all-axes result at the same window size; the two conditions are equal as reported rather than demonstrably identical.

The horizontal axes were less informative when used in isolation. For the CNN, x-only and z-only input resulted in accuracies of 0.78 and 0.75, respectively. For the LSTM, the corresponding values were 0.81 and 0.83. Thus, the relative ordering of the two horizontal axes differed between architectures, but both remained below the vertical-axis condition.

The benefit of using only the vertical coordinate was architecture-dependent. For the CNN, combining all three axes produced an accuracy of 0.93, eight

Table 3. Per-site classification accuracy at W=32 (M and SD over 10 runs), sorted by CNN mean. The highest mean for each model is shown in bold.

Site	1D-CNN		LSTM	
	M	SD	M	SD
Toe	**0.97**	0.0003	0.95	0.0006
Foot	0.96	0.0007	**0.96**	0.0008
Thigh	0.95	0.0011	0.95	0.0005
Toe tip	0.95	0.0009	0.94	0.0008
Shoulder	0.92	0.0025	0.95	0.0007
Neck	0.90	0.0028	0.94	0.0009
Hips	0.89	0.0042	0.87	0.0019
Head	0.88	0.0031	0.92	0.0012
Arm	0.88	0.0033	0.92	0.0011
Shin	0.87	0.0057	0.86	0.0006
Forearm	0.81	0.0024	0.89	0.0015
Hand	0.74	0.0019	0.85	0.0021

Table 4. Coordinate-axis ablation at W=32 with full-body input (M and SD over 10 runs). The highest single-axis mean for each model is shown in bold.

Input axes	1D-CNN		LSTM	
	M	SD	M	SD
x only (horizontal)	0.78	0.0038	0.81	0.0008
y only (vertical)	**0.85**	0.0026	**0.98**	0.0002
z only (horizontal)	0.75	0.0042	0.83	0.0007
All three axes (full-body)	0.93	0.0041	0.98	0.0009

percentage points above the y-only condition. In contrast, the LSTM obtained 0.98 from the vertical axis alone, matching its all-axes result.

Because **H4** is formulated as an existence statement ("for at least one architecture"), it is supported: for the LSTM, single-axis y-input matched the all-axes result. However, the effect is not general, since for the CNN retaining all three axes produced the substantially stronger result. The ablation nevertheless identifies the gravity-aligned vertical coordinate as the most informative individual axis for the evaluated activity set.

4.5 Cross-Subject Generalization

The preceding experiments used a subject-dependent split. To examine whether the main placement observation persisted for unseen participants, we conducted an additional leave-one-subject-out (LOSO) evaluation at W=32. In each of the three folds, the models were trained on two participants and tested on the held-out third participant, using that participant's separate test recording. We

Table 5. Leave-one-subject-out accuracy at W=32, reported as the mean (M) and standard deviation (SD) over 10 runs per fold. S1–S3 denote the held-out participant.

		S1		S2		S3	
Model	Input	M	SD	M	SD	M	SD
1D-CNN	Full-body	0.81	0.0041	0.76	0.0055	0.84	0.0036
1D-CNN	Toe	0.79	0.0048	0.73	0.0061	0.82	0.0044
LSTM	Full-body	0.89	0.0009	0.85	0.0014	0.92	0.0007
LSTM	Foot	0.87	0.0012	0.83	0.0016	0.90	0.0010

evaluated full-body input and the best site identified for each architecture in the subject-dependent analysis: the toe for the CNN and the foot for the LSTM. Each fold and input configuration was repeated over 10 training runs.

As shown in Table 5, absolute accuracy decreased under cross-subject evaluation. For full-body input, the CNN achieved a mean LOSO accuracy of 0.80, compared with its subject-dependent reference of 0.93. The LSTM achieved 0.89 under LOSO, compared with 0.98 in the subject-dependent evaluation. The two full-body models therefore showed reductions of 13 and 9 percentage points, respectively, although the two protocols differ in their participant composition and should be compared descriptively.

The selected distal sites remained close to their corresponding full-body inputs in all three folds. The CNN toe condition was two points below full-body input for S1, three points below for S2, and two points below for S3. The LSTM foot condition remained two points below full-body input in every fold. Averaged across folds, the selected sites reached 0.78 for the CNN and 0.87 for the LSTM, compared with 0.80 and 0.89 for their respective full-body conditions.

The LOSO analysis changes the interpretation of the CNN placement result. Under the subject-dependent protocol, the toe representation exceeded the CNN full-body reference by four percentage points. Under LOSO, full-body input outperformed the toe in every fold. Thus, the apparent superiority of the toe does not transfer to the held-out participants. What persists is the smaller observation that the selected distal site remained within three percentage points of full-body input.

These results provide preliminary cross-subject support for the near-parity component of **H3**, but not for the superiority of a distal site over full-body input. Moreover, the toe and foot conditions were selected based on the subject-dependent results rather than through site selection nested within each LOSO fold. The LOSO experiment should therefore be interpreted as a targeted robustness check for two preselected sites, not as an unbiased cross-subject comparison of all body sites.

5 Discussion

The results indicate that sensing location and temporal windowing materially influence HAR performance alongside model architecture. In the subject-dependent evaluation, selected distal lower-limb representations approached or exceeded full-body input, whereas the effect of window size differed between the two architectures. These findings suggest that sensing and preprocessing choices should be considered jointly with model selection. However, the results remain descriptive because they are based on three participants, six coarse-grained activities, and suit-derived positional data.

5.1 Interpretation of the Findings

Distal lower-limb representations captured the studied activities well. Foot and toe representations performed best for the investigated locomotion and posture activities, likely because they capture gait dynamics, impacts, and stance-related motion. Upper-limb movement was less consistently related to the activity labels, which may explain the weaker hand and forearm results. This agrees with previous findings on lower-limb sensing [1,10], but the suit-derived segment positions are not equivalent to measurements from standalone IMUs.

Window size interacted with model architecture. Larger windows did not improve accuracy consistently, and the observed trends differed between the CNN and LSTM. This may reflect differences in temporal modeling as well as the decreasing amount of independent training information at longer, strongly overlapping windows. Window size should therefore be optimized separately for each architecture.

The vertical coordinate carried substantial task information. The gravity-aligned vertical axis was the strongest individual coordinate for both models, likely because it captures vertical motion and posture height. Removing absolute vertical information may therefore discard useful task-specific features, while horizontal differencing alone does not provide heading invariance.

Accuracy and prediction time favored different architectures. The LSTM achieved higher accuracy, whereas the CNN required less host-side prediction time. However, system responsiveness also depends on the time required to accumulate a complete input window. Architecture and window size should therefore be selected jointly against a shared latency budget rather than optimized independently.

5.2 Limitations and Threats to Validity

The principal limitation is the small and homogeneous sample of three young male participants. The subject-dependent evaluation therefore measures generalization to separate recordings from known participants rather than to a broader

population. Although the LOSO analysis evaluates unseen participants, its three folds are insufficient for a reliable population-level placement ranking. Moreover, only the best subject-dependent site for each architecture was evaluated, making the analysis a targeted robustness check rather than an unbiased comparison of all sites.

The six coarse-grained locomotion and posture activities likely favor lower-limb sensing and the gravity-aligned vertical coordinate. Tasks involving gestures, object interaction, or asymmetric motion may produce different results. In addition, Rokoko provides fused segment positions rather than raw measurements from standalone IMUs. Its proprietary fusion and kinematic model may affect signal separability, and the bilateral site conditions do not represent strictly one-sided sensing.

Finally, the strongly overlapping windows are not statistically independent, and the reported standard deviations characterize optimization variability rather than population-level uncertainty. The window-size correlations are based on only five conditions and should be interpreted descriptively. Prediction time was measured using CPU-only TensorFlow inference on a consumer laptop, including framework overhead. End-to-end latency, memory consumption, and energy use were not measured, while the live pipeline provides only a functional demonstration of streaming inference.

5.3 Design Implications

Select sensing locations from the target activity dynamics. For the locomotion and posture activities evaluated here, distal lower-limb representations retained much of the information available from full-body input. Foot- or toe-mounted sensing is therefore a reasonable starting point for similar applications. This recommendation should not be transferred directly to gesture, manipulation, or upper-body tasks, for which placement must be evaluated separately.

Tune architecture and window size jointly. The two architectures responded differently to increasing temporal context. A window length adopted from previous work may therefore be inappropriate for a different model or dataset. Model selection should account for both prediction time and the accumulation delay W/f_s. Reducing model inference time provides limited practical benefit when a long window already determines the dominant response delay.

Preserve task-relevant coordinate information. The results support retaining the gravity-aligned vertical coordinate when posture height and vertical dynamics are informative. Horizontal coordinates can be transformed into relative displacement to reduce dependence on the recording location, but heading variation requires an additional normalization strategy. Coordinate preprocessing should consequently be documented and evaluated rather than treated as an implementation detail.

Report deployment conditions explicitly. HAR evaluations intended to inform real-time systems should report the sampling rate, window size, stride, window overlap, hardware platform, inference batch size, and prediction-time procedure alongside recognition accuracy. Training and test recordings should also be separated before window generation to prevent overlapping windows from crossing dataset splits. These details are necessary to distinguish improvements in recognition performance from changes in evaluation protocol or computational setting.

6 Future Work

Broader cross-subject evaluation. The most immediate next step is validation with a larger and more diverse participant sample, including variation in age, sex, body dimensions, gait patterns, and mobility. Longer recordings and less constrained activity execution would allow the placement ranking to be assessed under more realistic within- and cross-subject variation. Site selection should be performed within each training fold to avoid selecting placements using information from held-out participants.

Validation with deployable sensors. The suit-derived findings should be tested using standalone foot-mounted IMUs, instrumented insoles, or other low-cost wearable devices. Such an evaluation should compare unilateral and bilateral configurations and measure the effects of sensor noise, attachment position, footwear, sampling rate, and magnetic disturbance. Deployment experiments should additionally report end-to-end latency, memory use, and energy consumption on the target hardware.

Adaptive segmentation and privacy. Event-triggered segmentation based on gait events or activity transitions may reduce the accumulation delay imposed by fixed windows. Confidence-based early classification may further reduce computation for clearly distinguishable activities. At the same time, future work should examine whether reduced sensing configurations limit the identity information contained in full-body kinematics while preserving activity recognition utility.

7 Conclusion

This controlled pilot study compared 1D-CNN and LSTM classifiers across temporal window sizes, coordinate representations, and body sites for six daily activities. Aggregated across window sizes, the LSTM achieved higher accuracy, whereas the CNN executed approximately 3.6× faster on the evaluated host platform. Window size showed architecture-specific rather than uniformly beneficial effects, indicating that temporal context should be tuned jointly with model choice. In the subject-dependent evaluation, the bilateral toe representation reached 97% accuracy with the CNN and the bilateral foot reached 96% with both architectures, exceeding the CNN full-body reference at $W=32$ while

remaining below the LSTM full-body reference. Under LOSO evaluation, the selected distal representation remained within three percentage points of full-body input in every fold, although the toe's subject-dependent advantage over full-body input did not transfer to held-out participants.
These findings motivate an embodiment-first design process in which the sensing location, coordinate representation, and temporal context are evaluated before model complexity is increased. They do not yet establish an optimal wearable placement because the study used three participants and fused positions from a full-body motion tracking suit. Validation with broader populations, raw wearable sensors, and instrumented edge deployments is required before the observed design patterns can be transferred to practical HAR systems.

Acknowledgments. This publication is a result of the research funded by the Federal Ministry for Economic Affairs and Climate Protection of Germany (KK5646401RH4) and by the Federal Ministry of Research, Technology and Space of Germany (6SV9597, 03DPC0711A).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Attal, F., Mohammed, S., Dedabrishvili, M., Chamroukhi, F., Oukhellou, L., Amirat, Y.: Physical human activity recognition using wearable sensors. Sensors **15**(12), 31314–31338 (2015). https://doi.org/10.3390/s151229858
2. Bao, L., Intille, S.S.: Activity recognition from user-annotated acceleration data. In: Ferscha, A., Mattern, F. (eds.) Pervasive Computing. pp. 1–17. Springer Berlin Heidelberg, Berlin, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24646-6_1
3. Beaufils, B., Chazal, F., Grelet, M., Michel, B.: Robust stride detector from ankle-mounted inertial sensors for pedestrian navigation and activity recognition with machine learning approaches. Sensors **19**(20) (2019). https://doi.org/10.3390/s19204491
4. Benos, L., Tsaopoulos, D., Tagarakis, A.C., Kateris, D., Bochtis, D.: Optimal sensor placement and multimodal fusion for human activity recognition in agricultural tasks. Applied Sciences **14**(18) (2024). https://doi.org/10.3390/app14188520
5. Cao, K., Wang, M.: Human behavior recognition based on sparse transformer with channel attention mechanism. Frontiers in Physiology **14**, 1239453 (2023). https://doi.org/10.3389/fphys.2023.1239453
6. Cao, M., Wan, J., Gu, X.: Clear: Multimodal human activity recognition via contrastive learning based feature extraction refinement. Sensors **25**(3) (2025). https://doi.org/10.3390/s25030896
7. Casale, P., Pujol, O., Radeva, P.: Human activity recognition from accelerometer data using a wearable device. In: Vitrià, J., Sanches, J.M., Hernández, M. (eds.) Pattern Recognition and Image Analysis. pp. 289–296. Springer Berlin Heidelberg, Berlin, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21257-4_36

8. Cero Dinarević, E., Baraković Husić, J., Baraković, S.: Step by step towards effective human activity recognition: A balance between energy consumption and latency in health and wellbeing applications. Sensors **19**(23) (2019). https://doi.org/10.3390/s19235206
9. Czekaj, Ł., Kowalewski, M., Domaszewicz, J., Kitłowski, R., Szwoch, M., Duch, W.: Real-time sensor-based human activity recognition for efitness and ehealth platforms. Sensors **24**(12), 3891 (2024). https://doi.org/10.3390/s24123891
10. Davoudi, A., Mardini, M.T., Nelson, D., Albinali, F., Ranka, S., Rashidi, P., Manini, T.M.: The effect of sensor placement and number on physical activity recognition and energy expenditure estimation in older adults: Validation study. JMIR Mhealth Uhealth **9**(5), e23681 (May 2021). https://doi.org/10.2196/23681
11. Dirgová Luptáková, I., Kubovčík, M., Pospíchal, J.: Wearable sensor-based human activity recognition with transformer model. Sensors **22**(5) (2022). https://doi.org/10.3390/s22051911
12. Dua, N., Singh, S.N., Semwal, V.B.: Multi-input cnn-gru based human activity recognition using wearable sensors. Computing **103**(7), 1461–1478 (Jul 2021). https://doi.org/10.1007/s00607-021-00928-8
13. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7132–7141 (2018). https://doi.org/10.1109/CVPR.2018.00745
14. Jaramillo, I.E., Jeong, J.G., Lopez, P.R., Lee, C.H., Kang, D.Y., Ha, T.J., Oh, J.H., Jung, H., Lee, J.H., Lee, W.H., Kim, T.S.: Real-time human activity recognition with imu and encoder sensors in wearable exoskeleton robot via deep learning networks. Sensors **22**(24) (2022). https://doi.org/10.3390/s22249690
15. Jeyakumar, J.V., Sarker, A., Garcia, L.A., Srivastava, M.: X-CHAR: A concept-based explainable complex human activity recognition model. Proc ACM Interact Mob Wearable Ubiquitous Technol **7**(1) (Mar 2023). https://doi.org/10.1145/3580804
16. Jiang, W., Yin, Z.: Human activity recognition using wearable sensors by deep convolutional neural networks. In: Proceedings of the 23rd ACM International Conference on Multimedia. pp. 1307–1310. MM '15, Association for Computing Machinery, New York, NY, USA (2015). https://doi.org/10.1145/2733373.2806333
17. Khan, Z.N., Ahmad, J.: Attention induced multi-head convolutional neural network for human activity recognition. Applied Soft Computing **110**, 107671 (2021). https://doi.org/10.1016/j.asoc.2021.107671
18. Lane, N.D., Georgiev, P.: Can deep learning revolutionize mobile sensing? In: Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications. pp. 117–122. HotMobile '15, Association for Computing Machinery, New York, NY, USA (2015). https://doi.org/10.1145/2699343.2699349
19. Lin, J., Zhu, L., Chen, W.M., Wang, W.C., Han, S.: Tiny machine learning: Progress and futures [feature]. IEEE Circuits and Systems Magazine **23**(3), 8–34 (2023). https://doi.org/10.1109/mcas.2023.3302182
20. Liu, H., Gamboa, H., Schultz, T.: Sensor-based human activity and behavior research: Where advanced sensing and recognition technologies meet. Sensors **23**(1) (2023). https://doi.org/10.3390/s23010125
21. Mahmud, S., Tonmoy, M.T.H., Bhaumik, K.K., Rahman, A.K.M.M., Amin, M.A., Shoyaib, M., Khan, M.A.H., Ali, A.A.: Human activity recognition from wearable sensor data using self-attention. In: ECAI 2020: 24th European Conference on Artificial Intelligence. Frontiers in Artificial Intelligence and Applications, vol. 325, pp. 1332–1339. IOS Press (2020). https://doi.org/10.3233/FAIA200236

22. Mannini, A., Rosenberger, M., Haskell, W.L., Sabatini, A.M., Intille, S.S.: Activity recognition in youth using single accelerometer placed at wrist or ankle. Medicine & Science in Sports & Exercise **49**(4), 801–812 (2017). https://doi.org/10.1249/MSS.0000000000001144
23. Oh, Y., Kim, S.: Multi-modal lifelog data fusion for improved human activity recognition: A hybrid approach. Information Fusion **110**, 102464 (2024). https://doi.org/10.1016/j.inffus.2024.102464
24. Ordóñez, F.J., Roggen, D.: Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. Sensors **16**(1) (2016). https://doi.org/10.3390/s16010115
25. Özdemir, A.T., Barshan, B.: Detecting falls with wearable sensors using machine learning techniques. Sensors **14**(6), 10691–10708 (2014). https://doi.org/10.3390/s140610691
26. Quesada, F.J., Moya, F., Espinilla, M., Martínez, L., Nugent, C.D.: Feature sub-set selection for activity recognition. In: Inclusive Smart Cities and e-Health (ICOST 2015). Lecture Notes in Computer Science, vol. 9102, pp. 307–312. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19312-0_27
27. Ravi, N., Dandekar, N., Mysore, P., Littman, M.L.: Activity recognition from accelerometer data. In: Proceedings of the 17th Conference on Innovative Applications of Artificial Intelligence - Volume 3. pp. 1541–1546. IAAI'05, AAAI Press (2005)
28. Rokoko Electronics ApS: Smartsuit pro ii — inertial motion capture suit (2021), https://www.rokoko.com/products/smartsuit-pro, Product page, last accessed 2026-07-01
29. Stisen, A., Blunck, H., Bhattacharya, S., Prentow, T.S., Kjærgaard, M.B., Dey, A., Sonne, T., Jensen, M.M.: Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition. In: Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems (SenSys '15). pp. 127–140. Association for Computing Machinery, New York, NY, USA (2015). https://doi.org/10.1145/2809695.2809718
30. Tao, W., Chen, H., Moniruzzaman, M., Leu, M.C., Yi, Z., Qin, R.: Attention-based sensor fusion for human activity recognition using imu signals (2021). https://doi.org/10.48550/arXiv.2112.11224, Preprint arXiv:2112.11224, last accessed 2026-07-01
31. Tapia, E.M., Intille, S.S., Larson, K.: Activity recognition in the home using simple and ubiquitous sensors. In: Ferscha, A., Mattern, F. (eds.) Pervasive Computing. pp. 158–175. Springer Berlin Heidelberg, Berlin, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24646-6_10
32. Teng, S., Kim, J.Y., Jeon, S., Gil, H.W., Lyu, J., Chung, E.H., Kim, K.S., Nam, Y.: Analyzing optimal wearable motion sensor placement for accurate classification of fall directions. Sensors **24**(19) (2024). https://doi.org/10.3390/s24196432
33. Wang, G., Li, Q., Wang, L., Wang, W., Wu, M., Liu, T.: Impact of sliding window length in indoor human motion modes and pose pattern recognition based on smartphone sensors. Sensors **18**(6), 1965 (2018). https://doi.org/10.3390/s18061965
34. Wang, Y., Fehr, K.H., Adamczyk, P.G.: Impact-aware foot motion reconstruction and ramp/stair detection using one foot-mounted inertial measurement unit. Sensors **24**(5) (2024). https://doi.org/10.3390/s24051480
35. Yang, J.B., Nguyen, M.N., San, P.P., Li, X.L., Krishnaswamy, S.: Deep convolutional neural networks on multichannel time series for human activity recognition.

In: Proceedings of the 24th International Conference on Artificial Intelligence. pp. 3995–4001. IJCAI'15, AAAI Press (2015)
36. Yao, S., Hu, S., Zhao, Y., Zhang, A., Abdelzaher, T.: Deepsense: A unified deep learning framework for time-series mobile sensing data processing. In: Proceedings of the 26th International Conference on World Wide Web. pp. 351–360. WWW '17, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE (2017). https://doi.org/10.1145/3038912.3052577
37. Yin, X., Liu, Z., Liu, D., Ren, X.: A novel cnn-based bi-lstm parallel model with attention mechanism for human activity recognition with noisy data. Scientific Reports **12**(1), 7878 (May 2022). https://doi.org/10.1038/s41598-022-11880-8
38. Zeng, M., Nguyen, L.T., Yu, B., Mengshoel, O.J., Zhu, J., Wu, P., Zhang, J.: Convolutional neural networks for human activity recognition using mobile sensors. In: 6th International Conference on Mobile Computing, Applications and Services. pp. 197–205 (2014). https://doi.org/10.4108/icst.mobicase.2014.257786
39. Zhao, Y., Yang, R., Chevalier, G., Xu, X., Zhang, Z.: Deep residual bidir-lstm for human activity recognition using wearable sensors. Mathematical Problems in Engineering **2018**, 1–13 (2018). https://doi.org/10.1155/2018/7316954
40. Zhou, Y., Zhao, H., Huang, Y., Riedel, T., Hefenbrock, M., Beigl, M.: Tinyhar: A lightweight deep learning model designed for human activity recognition. In: Proceedings of the 2022 ACM International Symposium on Wearable Computers. pp. 89–93. ISWC '22, Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3544794.3558467