

GEAR-FEN: GENERALIZED FEATURE REPRESENTATION FOR KINEMATIC HUMAN ACTIVITY RECOGNITION

Anonymous authors

Paper under double-blind review

ABSTRACT

This study addresses the challenge of efficient human activity recognition (HAR) with limited training data. We propose GEAR-FEN (**Generalized Activity Recognition Feature Extraction Network**), a novel transfer learning (TL) method that transforms kinematic motion signals into a generalized feature space. GEAR-FEN potentially outperforms the state-of-the-art in scenarios with limited training data. This was demonstrated through an evaluation across 11 public HAR datasets (encompassing number of activities ranging from 6 to 33 and number of samples per activity ranging from 8628 to 1140258), using a deep learning model based on convolutional neural networks (CNN), residual bi-directional long short-term memory (ResBiLSTM), and an attention mechanism. Furthermore, we established the generalizability of our method through performance comparisons on an independent dataset encompassing a distinct population and diverse kinematic modalities for 8 activities, and 26121 samples per activity. These findings highlight the potential of our proposed approach in robust feature representation for HAR tasks with limited data sizes.

1 INTRODUCTION

Human activity recognition (HAR) is fundamental to comprehending and interpreting behavioral patterns in everyday life, allowing for a deeper insight into the intricacies of human behavior Arshad et al. (2022). Recent decades have witnessed a significant evolution in HAR technology, underscored by its growing importance in patient monitoring and rehabilitation, promoting safety and independence for individuals across various populations (Ariza-Colpas et al., 2022; Meng et al., 2020). The advent of wearable sensors, such as inertial measurement units (IMUs), has further fueled this progress, especially for ambulatory and real-life scenarios Qiu et al. (2022).

HAR faces significant challenges due to the inherent variability of motion patterns Chen et al. (2021). This variability arises from: 1) the specific activity set being classified Trabelsi et al. (2022); 2) demographic factors influencing movement patterns, necessitating population-specific models; and 3) variations introduced by the composition of the data collection sensory setup. These factors collectively contribute to the limited generalizability and data-specificity of HAR models. While deep learning architectures have revolutionized HAR by automating feature extraction and superior performance, they have also exacerbated the data dependency challenge due to their inherent complexity (Jain & Kanhangad, 2018; Presotto et al., 2023; Yuan et al., 2024). State-of-the-art (SOTA) models relying on Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks Xia et al. (2020), require vast amounts of labeled data for optimal performance. Bidirectional LSTM (BiLSTM) models have further refined the encoding of temporal information within activities, yet have also increased data dependency Luwe et al. (2022). Recent advancements, such as the incorporation of deep reverse transformer-based attention mechanisms, address the challenge of extracting both global temporal and local spatial features through enhanced feature fusion Pramanik et al. (2023). The ConvTransformer model, which merges CNNs with Transformers, exemplifies this trend by focusing on key features to achieve improved HAR performance Zhang et al. (2023b). Consequently, these advancements have intensified the data dependency issue, presenting a significant bottleneck, as collecting large datasets is resource-intensive and time-consuming, particularly for novel activity sets or populations.

A prevalent approach to addressing data scarcity in HAR is transfer learning (TL), defined as a set of methodologies leveraging knowledge acquired from a related task and domain to enhance performance on the target task Dhekane & Ploetz (2024). In this study, we address the data dependency and scarcity challenge by leveraging a novel generalized TL approach. To summarize, our contributions are as follows:

- We introduce the **generalized activity recognition feature extraction network (GEAR-FEN)**. Trained on a collection of diverse HAR datasets, GEAR-FEN is designed to extract representative features from the kinematic time series inherent to human motion.
- We evaluate our proposed method, which combines GEAR-FEN for feature extraction with dataset-specific feature learning networks (FLNs) for classification learning. This approach demonstrably improves classification performance, especially for datasets with limited samples, compared to baseline methods and SOTA.
- We assess our method using an independent, unseen dataset from a novel population with a novel set of activities. This analysis confirms that the features learned by GEAR-FEN effectively represent the underlying human motions and activities.

2 RELATED WORK

Within wearable HAR, TL plays a crucial role in adapting predictive models to diverse and dynamic conditions. Strategic approaches, such as heterogeneous transfer, personalized TL, multi-source TL, and task-specific methods, effectively address challenges like sensor position variability, user-specific data adaptation, and the integration of heterogeneous datasets Dhekane & Ploetz (2024). Heterogeneous transfer enhances model reliability across various sensor modalities and positions, while personalized TL focuses on tailoring models to individual characteristics by fine-tuning based on personal activity patterns and physiological data. Multi-source TL exploits the diversity of multiple datasets to build robust models that generalize across novel environments and tasks. Additionally, adapting models to handle varying definitions and labels of activities helps accommodate significant discrepancies across datasets, ensuring the models’ applicability in real-world scenarios.

Recent advancements in HAR using wearable data have explored various methodologies to address challenges like data scarcity, sensor diversity, and robust feature representation. One notable recent study implemented multi-task self-supervised learning with a deep convolutional neural network (ResNet-V2, 18 layers) on an extensive unlabeled dataset from the UK Biobank, encompassing 700,000 person-days of accelerometer data, to recognize human activities Yuan et al. (2024). This network was pre-trained and subsequently evaluated via TL across eight public datasets to assess representation quality. However, the study encountered limitations such as the dataset’s demographic homogeneity—predominantly Caucasian data from the UK—which could hinder the generalizability of the findings. Additionally, the reliance solely on accelerometer data pointed to the necessity of integrating multimodal sensor data to improve the model’s robustness and applicability. Concerns were also raised about the self-supervised learning methods not achieving high-quality representations with free-living activity data, indicating a need for further methodological refinement and exploration of more sophisticated techniques.

Another study addressed the issue of labeled data scarcity in HAR by proposing a novel approach to combine multiple public datasets to create a generalized model that required less labeled data for effective fine-tuning on unseen domains Presotto et al. (2023). This model was trained specifically on data from waist-mounted devices and evaluated for its generalization ability across various datasets. The study revealed that focusing solely on waist-mounted data may not fully capture the variability of real-world scenarios, where devices may be placed at different body locations. Additionally, the performance variability across different datasets suggested that dataset-specific characteristics could significantly impact the effectiveness of the generalized model, and the advantage of pre-training might be limited when the target dataset diverges substantially from the other combined datasets.

An additional concern in the current literature is the prevalent use of multi-channel data for training models. While this approach benefits from rich, multi-dimensional inputs, it inherently limits the models’ applicability to general kinematic signals that might be captured using simpler or different sensor configurations. This limitation raises questions about the scalability and flexibility of current HAR systems when deployed in real-world scenarios where sensor setups are not standardized and

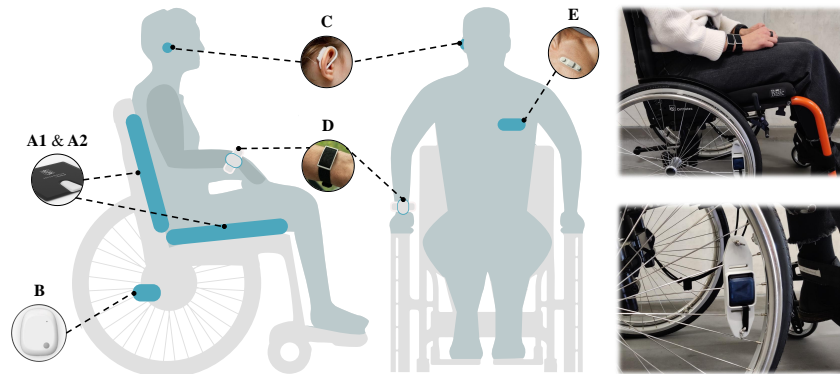


Figure 1: Composition of the sensory setup for the independent dataset. **A1 & A2**: Sensomative Wheelchair (pressure mats for backrest and bottom, respectively), **B**: mbientLab MMS+ (IMU), **C**: Cosinuss cm α (accelerometer), **D**: Vivalink Wearable ECG Monitor (accelerometer), **E**: Corsano CardioWatch 287-2 (accelerometer).

may vary greatly from one application to another (Gong et al., 2022; Suh et al., 2022; Sanabria et al., 2021; Al Hafiz Khan & Roy, 2022; Zhu et al., 2023; Alajaji et al., 2023; Qian et al., 2021; Varshney et al., 2022; Alinia et al., 2023; Lu et al., 2022; Faridee et al., 2022; Hu et al., 2023; Pavliuk et al., 2023; Haresamudram et al., 2022).

Despite the progress in this area, the HAR field yet faces notable limitations. Models frequently exhibit performance degradation when applied to domains that starkly differ from the training data, a phenomenon often observed in cross-domain applications Chen et al. (2019). Challenges such as overfitting and the disappearance of gradients in models trained on small datasets continue to hinder the reliability of these techniques Soleimani et al. (2022). Furthermore, the generalization of task-agnostic features often remains inadequate, leading to suboptimal performance when models are applied to new types of activities or different execution styles of the same activities Presotto et al. (2023). Another critical issue is the misclassification within activity clusters, where similar activities are often indistinguishable by the model, undermining its practical utility Kumar & Suresh (2023). Furthermore, the existing SOTA TL methods depend on a fixed sensor configuration or, at minimum, the same number of signals across all datasets, which limits their ability to leverage the diversity of available HAR datasets (Zhu et al., 2023; Alajaji et al., 2023).

3 MATERIALS AND METHOD

3.1 DATA

Eleven public HAR datasets were used to assess the generalization and classification efficacy of our proposed model: WISDM Kwapisz et al. (2011), MotionSense Malekzadeh et al. (2018), HHAR Stisen et al. (2015), REALWORLD Szttyler & Stuckenschmidt (2016), UniMiB SHAR Micucci et al. (2017), USC-HAD Zhang & Sawchuk (2012), MHEALTH Banos et al. (2014), PAMAP2 Reiss & Stricker (2012), WARD Yang et al. (2009), DSADS Barshan & Yüksesek (2014), and RealDISP Baños et al. (2012). These datasets ensured a wide spectrum of human activities, signal modalities, and age ranges. The employed modalities included linear acceleration and angular velocity. Further, we validated the generalization of our model using our independent dataset (Our Previous Work, 2023). The sensory setup for the independent data shown in Figure 1 was designed to monitor activities of daily living in wheelchair users via modalities from accelerometers, gyroscopes, and pressure mat sensors.

Table 1, gives an overview of the datasets included in this study. The time series from all datasets were resampled to a sampling rate of 20Hz, adequate for monitoring daily human activities Ejtehadi et al. (2023). Our experiments utilized 20s sliding windows with a 50% overlap. A 20-s time frame is

Table 1: Overview of the open-source and the independent HAR datasets utilized in this study, detailing the number of activities, signals, subjects, and samples per activity for each dataset.

Dataset	Num. of Activities	Num. of Signals	Num. of Subjects	Samples/Activity
DSADS (Barshan & Yüksek, 2014)	19	30	8	48000
HHAR (Stisen et al., 2015)	6	6	9	557644
Mhealth (Banos et al., 2014)	12	15	10	11439
Motionsense (Malekzadeh et al., 2018)	6	6	24	94192
PAMAP2 (Reiss & Stricker, 2012)	12	18	9	32382
REALDISP (Baños et al., 2012)	33	54	17	8628
REALWORLD (Sztyley & Stuckenschmidt, 2016)	8	6	15	1140258
UniMiB SHAR (Micucci et al., 2017)	9	3	30	20923
USCHAD (Zhang & Sawchuk, 2012)	12	6	14	46861
WARD (Yang et al., 2009)	13	25	20	44169
WISDM (Kwapisz et al., 2011)	6	3	36	312579
Independent Dataset	8	39	20	26121

sufficient for most HAR applications. For convenience, we refer to the collective set of all eleven processed open-source datasets as datapool.

3.2 FRAMEWORK OVERVIEW

Figures 2, 3, and 4 provides a detailed depiction of the proposed framework for the study. To evaluate the generalizability and efficacy of our proposed method, we conducted a comparative analysis including 2 other baseline methods. Baseline method 1 (detailed in Figure 2), the standard practice, trained each dataset independently using a multi-channel FEN to process all signals concurrently,

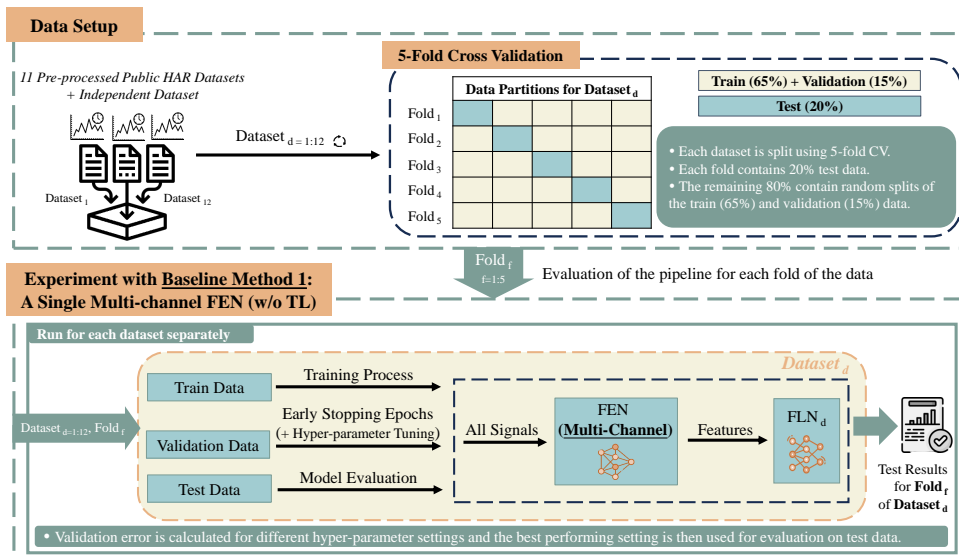


Figure 2: Flowchart of the experimental setup for baseline method 1, illustrating the feature learning process for human motion kinematic signals, with training performed independently for each dataset. This method utilizes a multi-channel FEN to learn features from all signals collectively. Indices d and f refer to the active dataset and fold respectively. Subscript L indicates the number of available signals in the respective $dataset_d$

216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269

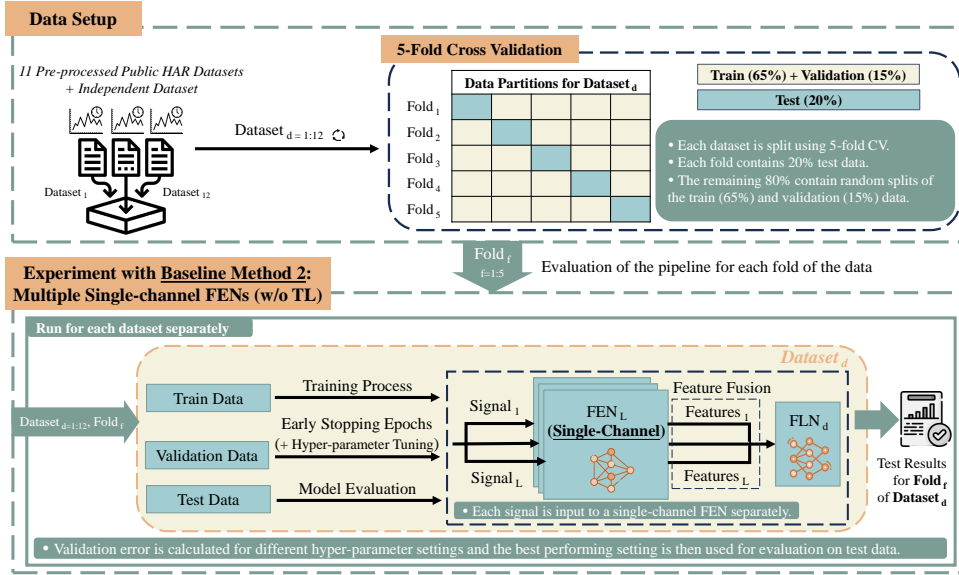


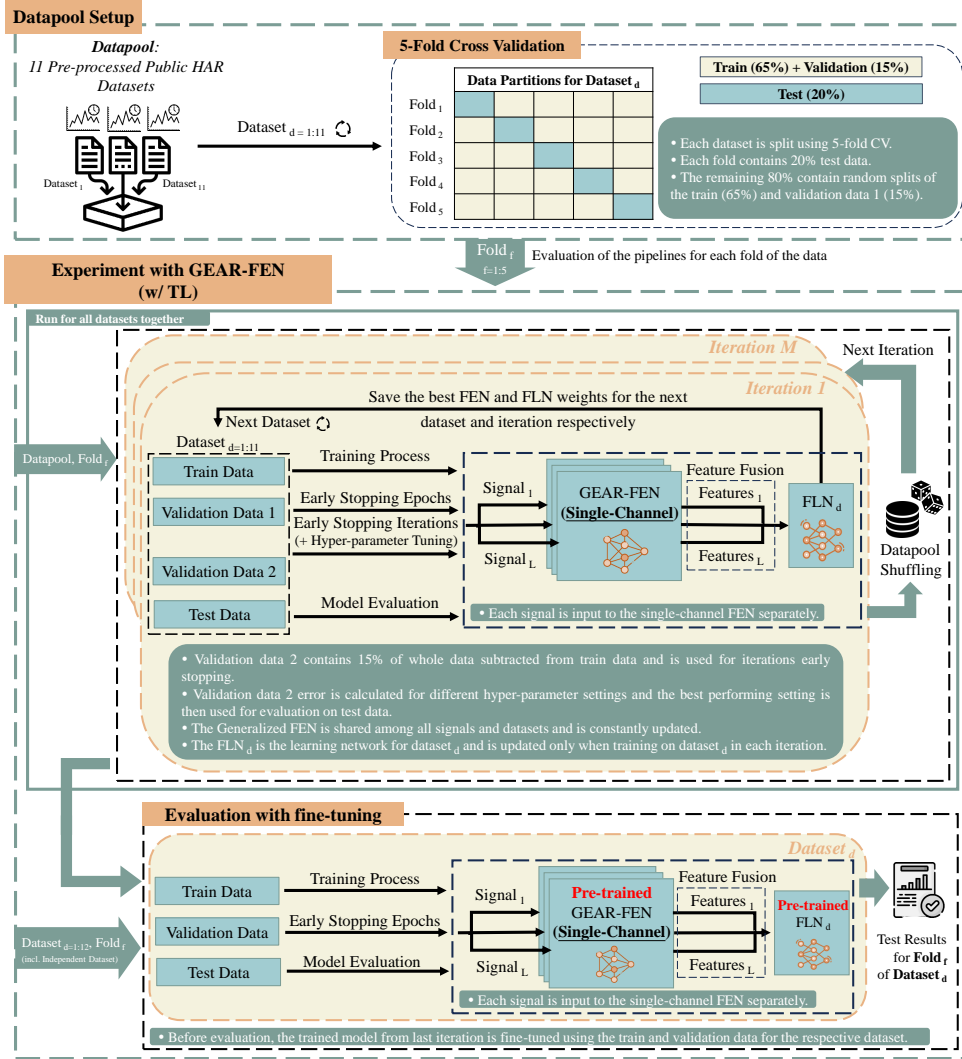
Figure 3: Flowchart of the experimental setup for baseline method 2, illustrating the feature learning process for human motion kinematic signals, with training performed independently for each dataset. This method utilizes a single-channel signal-specific FEN to learn features from the input signal. Indices d and f refer to the active dataset and fold respectively. Subscript L indicates the number of available signals in the respective $dataset_d$

followed by the FLN for classification and feature learning. Baseline method 2 (detailed in 3), also a standard practice, utilized a distinct single-channel FEN for each signal, followed by the FLN for classification of the concatenated feature representation of the signals. Lastly, the GEAR-FEN method (detailed in Figure 4), structurally similar to method 2, employed one shared, single-channel FEN (GEAR-FEN) across all datasets and signals, aimed to achieve transferable feature representations from human motion signals. While the GEAR-FEN weights were continuously updated across all datasets and iterations, the FLN weights were only updated when training on the corresponding dataset within each iteration. The one-dimensional structure of GEAR-FEN would ensure its efficacy despite data heterogeneity, such as variations in the number of sensors and signals. This is despite the standard SOTA TL pipelines where a multi-channel FEN with fixed number of input signals is pre-trained and fine-tuned, which limits the pipeline to using the datasets with identical sensor and signal settings.

As discussed in section 3.1, all data underwent pre-processing for consistency. A 5-fold cross-validation (50-65% training, 15-30% validation, 20% test) ensured unbiased experimental results. To maintain the inter-subject variability and to avoid data leakage in the evaluation, we split the data by subjects. The performance was evaluated on the held-out test data of each fold without and with fine-tuning. In the former approach, the model from the final iteration was directly evaluated on each dataset. In the latter, the final model was further fine-tuned on the respective dataset before evaluation. The detail description of method 3 is provided in Algorithm 1.

3.3 NETWORK STRUCTURE & HYPER-PARAMETER TUNING

Inspired by Zhang et al. (2023a), the network architecture for all the baseline and GEAR-FEN methods incorporated CNN, ResBiLSTM, and an attention mechanism, effectively capturing the intrinsic patterns of human activity. For a detailed overview of the network structure, refer to Figure 5. The FEN architectures consisted of a 1-D CNN layer responsible for capturing low-level temporal dependencies, followed by ReLU activation and then max-pooling for dimensionality reduction. The FLN architectures consisted of a ResBiLSTM layer for learning long-term dependencies and temporal relationships from the dataset-specific features, followed by an attention mechanism and



305
306
307
308
309
310
311
312

Figure 4: A Flowchart of the experimental setup employing TL through GEAR-FEN, illustrating the generalized feature learning process for human motion kinematic signals. This method utilizes a generalized single-channel FEN to learn features from any input signal. Indices d and f refer to the active dataset and fold respectively. Subscript L indicates the number of available signals in the respective $dataset_d$

313
314
315
316

then a fully connected layer. The attention mechanism would allow the network to focus on the most relevant parts of the input sequence, improving its ability to identify significant patterns and filter out less important information. The fully connected layer would aggregate the attention-weighted features and produce the final classification output.

317
318
319
320
321
322
323

To ensure a fair comparison between the baseline methods and GEAR-FEN, hyper-parameter tuning was carried out for all three methods with the grid shown in Figure 5. Hyper-parameter tuning prioritized method 3 due to its access to a large data pool, supporting complex model structures without overfitting. The optimal hyper-parameters identified for GEAR-FEN method set the initial search space for baseline methods 1 and 2. Given GEAR-FEN method’s extensive use of data, baseline methods 1 and 2 required a simplification of the model parameters to avoid overfitting. Moreover, since the FEN in the first baseline method employs a multi-channel approach, its output dimensions were the same as those of the single signals in the second baseline and GEAR-FEN

Algorithm 1 GEAR-FEN Training Process

Input: HAR datasets D_1, D_2, \dots, D_N , split into train, validation 1 (early stopping for epochs), validation 2 (early stopping for iterations), and test data

Parameters: Learning rate: η , Batch size: n

Output: Model weights W for FEN and W_d for FLN_d for D_d

```

1: Initialize  $W$ 
2: Initialize  $W_d$  for  $d = 1$  to  $N$ 
3: while true do
4:   Shuffle the order of datasets  $D_1$  to  $D_N$ 
5:   for each dataset  $D_d$  in shuffled order do
6:     for each signal  $X$  in train data do
7:       Compute feature representation for each signal:  $\hat{\Theta} = FEN(X; W)$ 
8:     end for
9:     Concatenate feature representations from all signals
10:    Update  $W$  and  $W_d$  based on early stopping criteria for epochs
11:   end for
12:   if Early stopping criteria for iterations then
13:     break
14:   end if
15: end while
16: for each dataset  $D_d$  do
17:   Combine validation data 2 with train data for  $D_d$ 
18:   Update  $W$  and  $W_d$  based on early stopping criteria for epochs
19: end for
20: return  $W$  and  $W_d$  for  $d = 1$  to  $N$ 

```

Table 2: Mean and standard deviation of f1-score performance comparison of the baseline methods, GEAR-FEN method, and SOTA performance (w/ or w/o TL). For the SOTA performances, the standard deviations are shown as n/a when not reported.

Dataset	Average F1-Score			
	Method 1	Method 2	Method 3	SOTA
DSADS	0.80 ± 0.08	0.85 ± 0.04	0.89 ± 0.03	0.82 ± n/a (Su et al., 2022)
HHAR	0.70 ± 0.04	0.70 ± 0.05	0.72 ± 0.04	0.52 ± n/a (Bock et al., 2022)
Mhealth	0.83 ± 0.06	0.88 ± 0.05	0.92 ± 0.03	0.94 ± 0.04 (Suh et al., 2023)
Motionsense	0.93 ± 0.05	0.95 ± 0.01	0.96 ± 0.01	0.92 ± n/a (Tahir et al., 2022)
PAMAP2	0.83 ± 0.10	0.93 ± 0.04	0.95 ± 0.03	0.86 ± n/a (Essa & Abdelmaksoud, 2023)
REALDISP	0.83 ± 0.02	0.90 ± 0.01	0.92 ± 0.01	0.88 ± 0.17 (Suh et al., 2023)
REALWORLD	0.73 ± 0.03	0.74 ± 0.02	0.76 ± 0.01	0.78 ± n/a (Kwon et al., 2020)
UniMib SHAR	0.67 ± 0.04	0.72 ± 0.04	0.82 ± 0.06	0.77 ± n/a (Al-qaness et al., 2023)
USCHAD	0.80 ± 0.06	0.81 ± 0.04	0.84 ± 0.05	0.83 ± n/a (Essa & Abdelmaksoud, 2023)
WARD	0.89 ± 0.03	0.93 ± 0.02	0.95 ± 0.03	0.91 ± n/a (Yang et al., 2009)
WISDM	0.80 ± 0.04	0.80 ± 0.05	0.82 ± 0.04	0.85 ± n/a (Essa & Abdelmaksoud, 2023)
Independent Dataset	0.56 ± 0.07	0.73 ± 0.07	0.85 ± 0.04	–

methods after the concatenation of features. Further, for a fair comparison between the feature extraction pipelines, the FLNs for all methods shared the same hyper-parameters. The batch size was set at 64 to effectively manage computational resources and model performance.

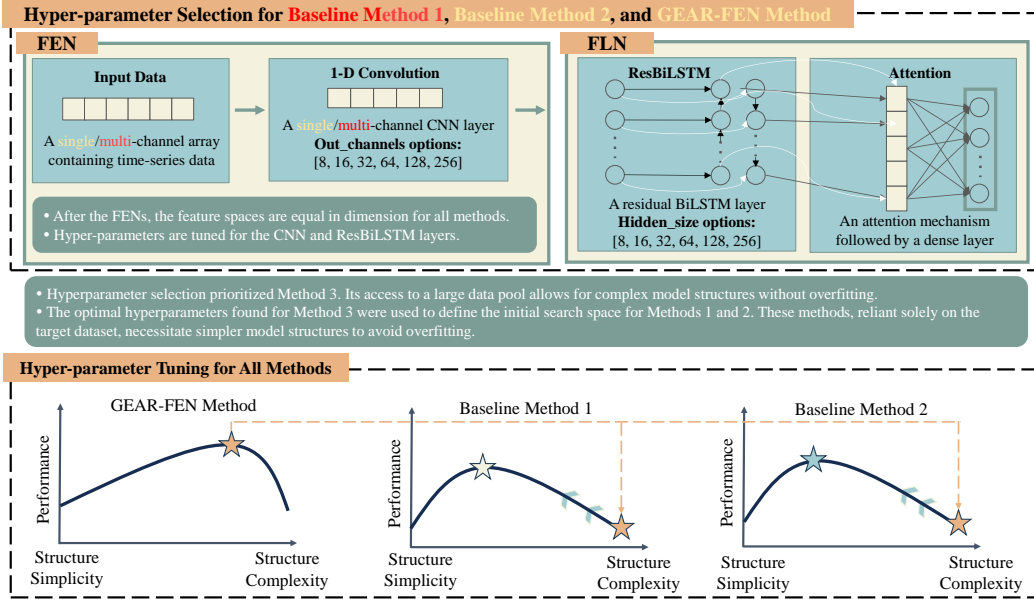


Figure 5: Network architectures and hyper-parameter grids for the methods used in this study. GEAR-FEN method’s access to ample data allowed for increased complexity without overfitting. The optimal settings for GEAR-FEN method defined the initial hyper-parameter grid space for baseline methods 1 and 2.

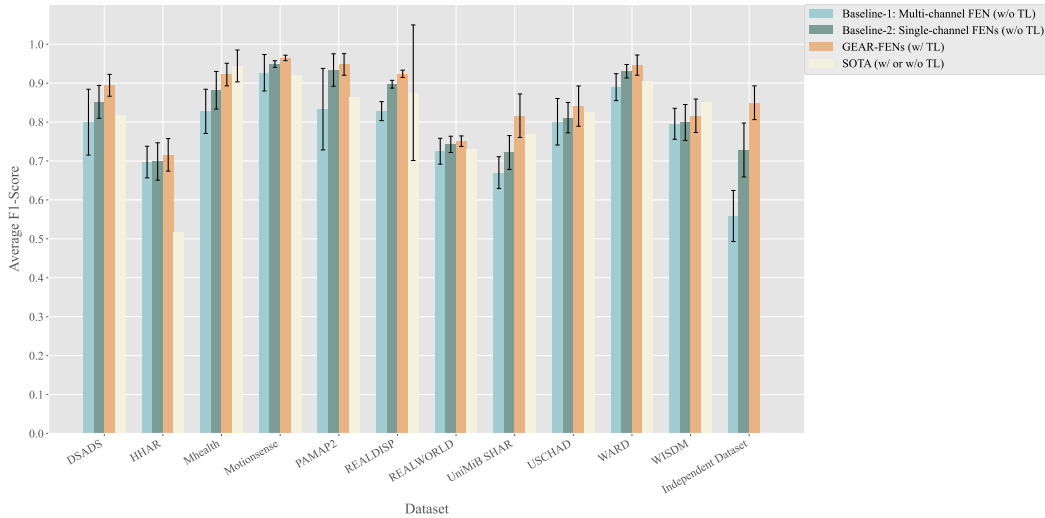


Figure 6: Barplot visualization of the mean and standard deviation of f1-score performance comparison of the baseline methods, GEAR-FEN method, and SOTA performance (w/ or w/o TL). For the SOTA performances, the standard deviations are visualized when reported.

4 EXPERIMENTS AND RESULTS

Table 2 presents the mean and standard deviation of f1-scores, averaged across all five experimental folds, for all methods and datasets in this study. The table also compares the baseline and GEAR-FEN performances with the SOTA performance for each dataset. Further, Figure 6 the barplot depicts

432 the barplot representation for the comparative f1-scores of the baseline and GEAR-FEN methods,
433 alongside the SOTA performances.

434 The results suggest that the GEAR-FEN method consistently achieves better or comparable results
435 in terms of f1-score, always outperforming the other baseline methods and often outperforming the
436 SOTA. In 9 out of 12 datasets, GEAR-FEN achieves the highest average f1-score. For instance on
437 DSADS, GEAR-FEN achieves a score of 0.89 ± 0.03 , significantly better than baseline method 1
438 (0.80 ± 0.08), baseline method 2 (0.85 ± 0.04), and SOTA (0.82). On UniMiB SHAR, GEAR-FEN
439 outperforms the baselines with a score of 0.82 ± 0.06 , compared to baseline 1 (0.67 ± 0.04) and
440 baseline 2 (0.72 ± 0.04), and surpasses SOTA (0.77). Similarly, GEAR-FEN consistently outperforms
441 on datasets like PAMAP2, REALDISP, and WARD, with higher mean f1-scores and relatively low
442 standard deviations, indicating robust performance. In some cases, like Mhealth and WISDM, the
443 GEAR-FEN method’s performance is slightly below SOTA, but it still shows competitive results with
444 a narrower margin for Mhealth. In the case of the independent dataset, the GEAR-FEN method (0.85
445 ± 0.04) outperforms the baseline methods (0.56 ± 0.07 and 0.73 ± 0.07) by a considerable margin.
446 Also the standard deviation has a narrower margin which shows the robustness of GEAR-FEN against
447 the baseline methods.

448 449 5 DISCUSSION

451 The results demonstrate that the GEAR-FEN method consistently outperforms the baseline methods
452 and often exceeds SOTA performance across majority of the investigated HAR datasets. GEAR-FEN
453 demonstrates strong generalization, particularly in more challenging datasets with limited sample size,
454 and shows greater robustness, as indicated by lower variance in f1-scores. Even in cases where it does
455 not surpass SOTA, GEAR-FEN remains competitive, highlighting its effectiveness in feature learning
456 for human motion signals. The observed improvements in test accuracy across iterations for a variety
457 of datasets, substantiate the effectiveness of our model and training approach. Importantly, the ability
458 of our model to generalize across such a wide range of HAR datasets supports the hypothesis that
459 kinematic signals e.g., linear acceleration and angular velocity and motion-related kinetic modalities
460 such as those in pressure mats exhibit similar patterns, which GEAR-FEN can capture.

461 Moreover, the superior performance of the proposed models on an independent dataset indicates that
462 the feature representation learned by GEAR-FEN is highly transferable for human motion patterns.
463 This transferability has resulted in significant improvements when applied to datasets with small
464 sample sizes, and with new populations, sensor modalities, and sensor locations, as demonstrated by
465 the independent dataset’s performance.

466 Additionally, the results show that feature learning using multiple single-channel FENs consistently
467 outperforms feature representation from a multi-channel FEN, highlighting the structural superiority
468 of GEAR-FEN compared to SOTA transfer learning networks. It’s also important to note that GEAR-
469 FEN’s generalized single-channel structure addresses issues related to incompatible signal counts
470 and sensor locations in transfer learning. This design enables highly customizable fine-tuning of the
471 model, making GEAR-FEN adaptable to diverse datasets and sensor configurations

472 473 6 CONCLUSION

475 In this paper, we demonstrated a pipeline for automated feature representation of the kinematic
476 signals related to human motion. This pipeline can be used for transferring the signals into a domain
477 of representative features, capable of distinguishing between human activities. The generalized
478 feature extraction pipeline, combined with feature fusion models, contributes to HAR research,
479 outperforming most of the benchmark models in the field.

480 Further, the proposed activity classification pipeline outperformed most of the SOTA scores for
481 different HAR datasets. It is recognized that SOTA models in HAR heavily rely on large and diverse
482 datasets, posing significant challenges in data acquisition. Our research addresses these issues by
483 focusing on generalizing across all kinematic signals. While aiming for a model that generalizes
484 across all kinematic signals, we acknowledge that some datasets show lower accuracy compared
485 to certain SOTA models. This trade-off is acceptable in our goal to create a universally adaptable
system. It is noteworthy that the proposed GEAR-FEN pipeline can be further fine-tuned using

486 the HAR datasets with more inherent similarities to the target dataset. The customized fine-tuning
 487 can aid the model in capturing the local and intricate patterns and features specific to a set of
 488 activities or populations. Moreover, validation of the independent dataset confirms the effectiveness
 489 of GEAR-FEN’s generalization capability.

490 Our research revealed that the feature representation pipeline exhibits superior performance compared
 491 to the benchmark deep learning models when applied to an independent dataset. This dataset
 492 encompassed a diverse population (wheelchair users) and incorporated various sensor modalities,
 493 including pressure distribution. The findings suggest that the feature learning pipeline effectively
 494 captures the essential characteristics of motion from the signals, potentially enhancing classification
 495 performance when integrated with a classification model. The balance between broad applicability
 496 and dataset-specific precision highlights the potential of our models in diverse real-world applications
 497 and sets a direction for future research in HAR technology.

498 In our study, we acknowledge a limitation of the feature learning pipeline’s applicability to kinematic
 499 signals associated with human motion. Further, we acknowledge that the GEAR-FEN methodology
 500 has been validated only with a specific neural network structure (comprising CNN, ResBiLSTM,
 501 and attention mechanism). The pipeline should be further validated for different neural network
 502 structures. In forthcoming research, we intend to also employ the Local Interpretable Model-Agnostic
 503 Explanations (LIME) technique to elucidate the features extracted by the pipeline. This will help
 504 assess whether the feature representation pipeline accurately captures features relevant to the target
 505 output. Given the pipeline’s capability to transform any signal into the feature domain, conducting an
 506 activity-wise analysis of these features will enable us to delineate the contribution of each input signal
 507 modality to the recognition of various activities. Furthermore, it should be investigated whether the
 508 generalized feature representation pipeline can be extended to physiological signals with repeatable
 509 patterns, such as heart rate and blood pressure.

511 REFERENCES

- 512 Md Abdullah Al Hafiz Khan and Nirmalya Roy. Cross-domain unseen activity recognition using
 513 transfer learning. In *2022 IEEE 46th Annual Computers, Software, and Applications Conference*
 514 *(COMPSAC)*, pp. 684–693, 2022. doi: 10.1109/COMPSAC54236.2022.00117.
- 515 Mohammed A. A. Al-qaness, Abdelghani Dahou, Mohamed Abd Elaziz, and A. M. Helmi. Multi-
 516 resatt: Multilevel residual network with attention for human activity recognition using wearable
 517 sensors. *IEEE Transactions on Industrial Informatics*, 19(1):144–152, 2023. doi: 10.1109/TII.
 518 2022.3165875.
- 519 Abdulaziz Alajaji, Walter Gerych, Luke Buquicchio, Kavin Chandrasekaran, Hamid Mansoor,
 520 Emmanuel Agu, and Elke Rundensteiner. Domain adaptation methods for lab-to-field human
 521 context recognition. *Sensors*, 23(6), 2023. ISSN 1424-8220. doi: 10.3390/s23063081. URL
 522 <https://www.mdpi.com/1424-8220/23/6/3081>.
- 523 Parastoo Alinia, Asiful Arefeen, Zhila Esna Ashari, Seyed Iman Mirzadeh, and Hassan Ghasemzadeh.
 524 Model-agnostic structural transfer learning for cross-domain autonomous activity recognition.
 525 *Sensors*, 23(14), 2023. ISSN 1424-8220. doi: 10.3390/s23146337. URL <https://www.mdpi.com/1424-8220/23/14/6337>.
- 526 Paola Patricia Ariza-Colpas, Enrico Vicario, Ana Isabel Oviedo-Carrascal, Shariq Butt Aziz, Mar-
 527 lon Alberto Piñeres-Melo, Alejandra Quintero-Linero, and Fulvio Patara. Human activity recogni-
 528 tion data analysis: History, evolutions, and new trends. *Sensors*, 22(9), 2022. ISSN 1424-8220.
 529 doi: 10.3390/s22093401. URL <https://www.mdpi.com/1424-8220/22/9/3401>.
- 530 Muhammad Haseeb Arshad, Muhammad Bilal, and Abdullah Gani. Human activity recognition:
 531 Review, taxonomy and open challenges. *Sensors*, 22(17), 2022. ISSN 1424-8220. doi: 10.3390/
 532 s22176463. URL <https://www.mdpi.com/1424-8220/22/17/6463>.
- 533 Oresti Baños, Miguel Damas, Héctor Pomares, Ignacio Rojas, Máté Attila Tóth, and Oliver Amft.
 534 A benchmark dataset to evaluate sensor displacement in activity recognition. In *Proceedings*
 535 *of the 2012 ACM Conference on Ubiquitous Computing, UbiComp ’12*, pp. 1026–1035, New
 536 York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450312240. doi:
 537 10.1145/2370216.2370437. URL <https://doi.org/10.1145/2370216.2370437>.

- 540 Oresti Banos, Rafael Garcia, Juan A. Holgado-Terriza, Miguel Damas, Hector Pomares, Ignacio
541 Rojas, Alejandro Saez, and Claudia Villalonga. mhealthdroid: A novel framework for agile
542 development of mobile health applications. In Leandro Pecchia, Liming Luke Chen, Chris Nugent,
543 and José Bravo (eds.), *Ambient Assisted Living and Daily Activities*, pp. 91–98, Cham, 2014.
544 Springer International Publishing. ISBN 978-3-319-13105-4.
- 545
546 Billur Barshan and Murat Cihan Yükses. Recognizing daily and sports activities in two open source
547 machine learning environments using body-worn sensor units. *The Computer Journal*, 57(11):
548 1649–1667, 2014. doi: 10.1093/comjnl/bxt075.
- 549 M. Bock, A. Hoelzemann, M. Moeller, and K. Van Laerhoven. Investigating (re)current state-of-the-
550 art in human activity recognition datasets. *Frontiers in Computer Science*, 4:924954, 2022. doi:
551 10.3389/fcomp.2022.924954.
- 552
553 Kaixuan Chen, Dalin Zhang, Lina Yao, Bin Guo, Zhiwen Yu, and Yunhao Liu. Deep learning
554 for sensor-based human activity recognition: Overview, challenges, and opportunities. *ACM*
555 *Comput. Surv.*, 54(4), may 2021. ISSN 0360-0300. doi: 10.1145/3447744. URL <https://doi.org/10.1145/3447744>.
- 556
557 Yiqiang Chen, Jindong Wang, Meiyu Huang, and Han Yu. Cross-position activity recognition with
558 stratified transfer learning. *Pervasive and Mobile Computing*, 57:1–13, 2019. ISSN 1574-1192. doi:
559 <https://doi.org/10.1016/j.pmcj.2019.04.004>. URL <https://www.sciencedirect.com/science/article/pii/S1574119218303432>.
- 560
561
562 Sourish Gunesh Dhekane and Thomas Ploetz. Transfer learning in human activity recognition: A
563 survey, 2024.
- 564
565 Mehdi Ejtehadi, Sabrina Amrein, Inge Eriks Hoogland, Robert Riener, and Diego Paez-Granados.
566 Learning activities of daily living from unobtrusive multimodal wearables: Towards monitoring
567 outpatient rehabilitation. In *2023 International Conference on Rehabilitation Robotics (ICORR)*,
568 pp. 1–6, 2023. doi: 10.1109/ICORR58425.2023.10304743.
- 569
570 Ehab Essa and Islam R. Abdelmaksoud. Temporal-channel convolution with self-attention network
571 for human activity recognition using wearable sensors. *Knowledge-Based Systems*, 278:110867,
572 2023. ISSN 0950-7051. doi: <https://doi.org/10.1016/j.knosys.2023.110867>. URL <https://www.sciencedirect.com/science/article/pii/S0950705123006172>.
- 573
574 Abu Zaher Md Faridee, Avijoy Chakma, Archan Misra, and Nirmalya Roy. Strangan: Adversarially-
575 learnt spatial transformer for scalable human activity recognition. *Smart Health*, 23:100226,
576 2022. ISSN 2352-6483. doi: <https://doi.org/10.1016/j.smhl.2021.100226>. URL <https://www.sciencedirect.com/science/article/pii/S2352648321000477>.
- 577
578 Taesik Gong, Yeonsu Kim, Ryuhaerang Choi, Jinwoo Shin, and Sung-Ju Lee. Adapting to unknown
579 conditions in learning-based mobile sensing. *IEEE Transactions on Mobile Computing*, 21(10):
580 3470–3485, 2022. doi: 10.1109/TMC.2021.3061130.
- 581
582 Harish Haresamudram, Irfan Essa, and Thomas Ploetz. Investigating enhancements to contrastive
583 predictive coding for human activity recognition, 2022.
- 584
585 Rong Hu, Ling Chen, Shenghuan Miao, and Xing Tang. Swl-adapt: an unsupervised domain
586 adaptation model with sample weight learning for cross-user wearable human activity recognition.
587 In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth*
588 *Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on*
589 *Educational Advances in Artificial Intelligence*, AAAI’23/IAAI’23/EAAI’23. AAAI Press, 2023.
590 ISBN 978-1-57735-880-0. doi: 10.1609/aaai.v37i5.25743. URL <https://doi.org/10.1609/aaai.v37i5.25743>.
- 591
592 Ankita Jain and Vivek Kanhangad. Human activity classification in smartphones using accelerometer
593 and gyroscope sensors. *IEEE Sensors Journal*, 18(3):1169–1177, 2018. doi: 10.1109/JSEN.2017.2782492.

- 594 Prabhat Kumar and S. Suresh. Deeptranshar: a novel clustering-based transfer learning approach
595 for recognizing the cross-domain human activities using grus (gated recurrent units) networks.
596 *Internet of Things*, 21:100681, 2023. ISSN 2542-6605. doi: [https://doi.org/10.1016/j.iot.](https://doi.org/10.1016/j.iot.2023.100681)
597 2023.100681. URL [https://www.sciencedirect.com/science/article/pii/](https://www.sciencedirect.com/science/article/pii/S2542660523000045)
598 S2542660523000045.
- 599 Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A. Moore. Activity recognition using cell
600 phone accelerometers. *SIGKDD Explor. Newsl.*, 12(2):74–82, mar 2011. ISSN 1931-0145. doi:
601 10.1145/1964897.1964918. URL <https://doi.org/10.1145/1964897.1964918>.
- 602 Hyeokhyen Kwon, Catherine Tong, Harish Haresamudram, Yan Gao, Gregory D. Abowd, Nicholas D.
603 Lane, and Thomas Ploetz. Imutube: Automatic extraction of virtual on-body accelerometry from
604 video for human activity recognition, 2020. URL <https://arxiv.org/abs/2006.05675>.
- 605 Wang Lu, Jindong Wang, Yiqiang Chen, Sinno Jialin Pan, Chunyu Hu, and Xin Qin. Semantic-
606 discriminative mixup for generalizable sensor-based cross-domain activity recognition. *Proc.*
607 *ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(2), jul 2022. doi: 10.1145/3534589. URL
608 <https://doi.org/10.1145/3534589>.
- 609 Yee Jia Luwe, Chin Poo Lee, and Kian Ming Lim. Wearable sensor-based human activity recognition
610 with hybrid deep learning model. *Informatics*, 9(3), 2022. ISSN 2227-9709. doi: 10.3390/
611 informatics9030056. URL <https://www.mdpi.com/2227-9709/9/3/56>.
- 612 Mohammad Malekzadeh, Richard G. Clegg, Andrea Cavallaro, and Hamed Haddadi. Protecting
613 sensory data against sensitive inferences. In *Proceedings of the 1st Workshop on Privacy by Design*
614 *in Distributed Systems, W-P2DS’18*, New York, NY, USA, 2018. Association for Computing
615 Machinery. ISBN 9781450356541. doi: 10.1145/3195258.3195260. URL [https://doi.org/](https://doi.org/10.1145/3195258.3195260)
616 10.1145/3195258.3195260.
- 617 Zhaozong Meng, Mingxing Zhang, Changxin Guo, Qirui Fan, Hao Zhang, Nan Gao, and Zonghua
618 Zhang. Recent progress in sensing and computing techniques for human activity recognition and
619 motion analysis. *Electronics*, 9(9), 2020. ISSN 2079-9292. doi: 10.3390/electronics9091357.
620 URL <https://www.mdpi.com/2079-9292/9/9/1357>.
- 621 Daniela Micucci, Marco Mobilio, and Paolo Napoletano. Unimib shar: A dataset for human activity
622 recognition using acceleration data from smartphones. *Applied Sciences*, 7(10), 2017. ISSN
623 2076-3417. doi: 10.3390/app7101101. URL [https://www.mdpi.com/2076-3417/7/](https://www.mdpi.com/2076-3417/7/10/1101)
624 10/1101.
- 625 Olena Pavliuk, Myroslav Mishchuk, and Christine Strauss. Transfer learning approach for human
626 activity recognition based on continuous wavelet transform. *Algorithms*, 16(2), 2023. ISSN 1999-
627 4893. doi: 10.3390/a16020077. URL <https://www.mdpi.com/1999-4893/16/2/77>.
- 628 Rishav Pramanik, Ritodeep Sikdar, and Ram Sarkar. Transformer-based deep reverse attention
629 network for multi-sensory human activity recognition. *Engineering Applications of Artificial*
630 *Intelligence*, 122:106150, 2023. ISSN 0952-1976. doi: [https://doi.org/10.1016/j.engappai.](https://doi.org/10.1016/j.engappai.2023.106150)
631 2023.106150. URL [https://www.sciencedirect.com/science/article/pii/](https://www.sciencedirect.com/science/article/pii/S0952197623003342)
632 S0952197623003342.
- 633 Riccardo Presotto, Sannara Ek, Gabriele Civitarese, François Portet, Philippe Lalanda, and Claudio
634 Bettini. Combining public human activity recognition datasets to mitigate labeled data scarcity. In
635 *2023 IEEE International Conference on Smart Computing (SMARTCOMP)*, pp. 33–40, 2023. doi:
636 10.1109/SMARTCOMP58114.2023.00022.
- 637 Hangwei Qian, Sinno Jialin Pan, and Chunyan Miao. Latent independent excitation for generalizable
638 sensor-based cross-person activity recognition. *Proceedings of the AAAI Conference on Artificial*
639 *Intelligence*, 35(13):11921–11929, May 2021. doi: 10.1609/aaai.v35i13.17416. URL [https://](https://ojs.aaai.org/index.php/AAAI/article/view/17416)
640 ojs.aaai.org/index.php/AAAI/article/view/17416.
- 641 Sen Qiu, Hongkai Zhao, Nan Jiang, Zhelong Wang, Long Liu, Yi An, Hongyu Zhao, Xin Miao,
642 Ruichen Liu, and Giancarlo Fortino. Multi-sensor information fusion based on machine learning
643 for real applications in human activity recognition: State-of-the-art and research challenges. *In-*
644 *formation Fusion*, 80:241–265, 2022. ISSN 1566-2535. doi: <https://doi.org/10.1016/j.inffus>.

- 648 2021.11.006. URL <https://www.sciencedirect.com/science/article/pii/S1566253521002311>.
- 649
- 650
- 651 Attila Reiss and Didier Stricker. Introducing a new benchmarked dataset for activity monitoring. In
- 652 *2012 16th International Symposium on Wearable Computers*, pp. 108–109, 2012. doi: 10.1109/
- 653 ISWC.2012.13.
- 654 Andrea Rosales Sanabria, Franco Zambonelli, Simon Dobson, and Juan Ye. Contrasgan: Un-
- 655 supervised domain adaptation in human activity recognition via adversarial and contrastive
- 656 learning. *Pervasive and Mobile Computing*, 78:101477, 2021. ISSN 1574-1192. doi:
- 657 <https://doi.org/10.1016/j.pmcj.2021.101477>. URL [https://www.sciencedirect.com/](https://www.sciencedirect.com/science/article/pii/S1574119221001103)
- 658 [science/article/pii/S1574119221001103](https://www.sciencedirect.com/science/article/pii/S1574119221001103).
- 659 Elnaz Soleimani, Ghazaleh Khodabandelou, Abdelghani Chibani, and Yacine Amirat. Generic
- 660 semi-supervised adversarial subject translation for sensor-based activity recognition. *Neuro-*
- 661 *computing*, 500:649–661, 2022. ISSN 0925-2312. doi: [https://doi.org/10.1016/j.neucom.](https://doi.org/10.1016/j.neucom.2022.05.075)
- 662 [2022.05.075](https://doi.org/10.1016/j.neucom.2022.05.075). URL [https://www.sciencedirect.com/science/article/pii/](https://www.sciencedirect.com/science/article/pii/S0925231222006592)
- 663 [S0925231222006592](https://www.sciencedirect.com/science/article/pii/S0925231222006592).
- 664 Allan Stisen, Henrik Blunck, Sourav Bhattacharya, Thor Siiger Prentow, Mikkel Baun Kjærgaard,
- 665 Anind Dey, Tobias Sonne, and Mads Møller Jensen. Smart devices are different: Assessing
- 666 and mitigating mobile sensing heterogeneities for activity recognition. In *Proceedings of the*
- 667 *13th ACM Conference on Embedded Networked Sensor Systems*, SenSys ’15, pp. 127–140, New
- 668 York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450336314. doi:
- 669 [10.1145/2809695.2809718](https://doi.org/10.1145/2809695.2809718). URL <https://doi.org/10.1145/2809695.2809718>.
- 670
- 671 Jie Su, Zhenyu Wen, Tao Lin, and Yu Guan. Learning disentangled behaviour patterns for wearable-
- 672 based human activity recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(1),
- 673 March 2022. doi: [10.1145/3517252](https://doi.org/10.1145/3517252). URL <https://doi.org/10.1145/3517252>.
- 674 Sungho Suh, Vitor Fortes Rey, and Paul Lukowicz. Adversarial deep feature extraction network for
- 675 user independent human activity recognition. In *2022 IEEE International Conference on Pervasive*
- 676 *Computing and Communications (PerCom)*, pp. 217–226, 2022. doi: [10.1109/PerCom53586.2022.](https://doi.org/10.1109/PerCom53586.2022.9762387)
- 677 [9762387](https://doi.org/10.1109/PerCom53586.2022.9762387).
- 678 Sungho Suh, Vitor Fortes Rey, and Paul Lukowicz. Tasked: Transformer-based adversarial learning
- 679 for human activity recognition using wearable sensors via self-knowledge distillation. *Knowledge-*
- 680 *Based Systems*, 260:110143, 2023. ISSN 0950-7051. doi: [https://doi.org/10.1016/j.knosys.](https://doi.org/10.1016/j.knosys.2022.110143)
- 681 [2022.110143](https://doi.org/10.1016/j.knosys.2022.110143). URL [https://www.sciencedirect.com/science/article/pii/](https://www.sciencedirect.com/science/article/pii/S0950705122012394)
- 682 [S0950705122012394](https://www.sciencedirect.com/science/article/pii/S0950705122012394).
- 683
- 684 Timo Szytler and Heiner Stuckenschmidt. On-body localization of wearable devices: An investigation
- 685 of position-aware activity recognition. In *2016 IEEE International Conference on Pervasive*
- 686 *Computing and Communications (PerCom)*, pp. 1–9, 2016. doi: [10.1109/PERCOM.2016.7456521](https://doi.org/10.1109/PERCOM.2016.7456521).
- 687 Sheikh Badar ud din Tahir, Abdul Basit Dogar, Rubia Fatima, Affan Yasin, Muhammad Shafiq,
- 688 Javed Ali Khan, Muhammad Assam, Abdullah Mohamed, and El-Awady Attia. Stochastic
- 689 recognition of human physical activities via augmented feature descriptors and random forest
- 690 model. *Sensors*, 22(17), 2022. ISSN 1424-8220. doi: [10.3390/s22176632](https://doi.org/10.3390/s22176632). URL [https:](https://www.mdpi.com/1424-8220/22/17/6632)
- 691 [//www.mdpi.com/1424-8220/22/17/6632](https://www.mdpi.com/1424-8220/22/17/6632).
- 692 Imen Trabelsi, Jules Francoise, and Yacine Bellik. Sensor-based activity recognition using deep
- 693 learning: A comparative study. In *Proceedings of the 8th International Conference on Movement*
- 694 *and Computing*, MOCO ’22, New York, NY, USA, 2022. Association for Computing Machinery.
- 695 ISBN 9781450387163. doi: [10.1145/3537972.3537996](https://doi.org/10.1145/3537972.3537996). URL [https://doi.org/10.1145/](https://doi.org/10.1145/3537972.3537996)
- 696 [3537972.3537996](https://doi.org/10.1145/3537972.3537996).
- 697 N. Varshney, B. Bakariya, and A. K. S. Kushwaha. Human activity recognition using deep transfer
- 698 learning of cross position sensor based on vertical distribution of data. *Multimedia Tools and*
- 699 *Applications*, 81:22307–22322, 2022. doi: [10.1007/s11042-021-11131-4](https://doi.org/10.1007/s11042-021-11131-4).
- 700
- 701 Kun Xia, Jianguang Huang, and Hanyu Wang. Lstm-cnn architecture for human activity recognition. *IEEE Access*, 8:56855–56866, 2020. doi: [10.1109/ACCESS.2020.2982225](https://doi.org/10.1109/ACCESS.2020.2982225).

702 Allen Yang, Roozbeh Jafari, Shankar Sastry, and Ruzena Bajcsy. Distributed recognition of human
703 actions using wearable motion sensor networks. *JAISE*, 1:103–115, 01 2009. doi: 10.3233/
704 AIS-2009-0016.

705
706 H. Yuan, S. Chan, A. P. Creagh, et al. Self-supervised learning for human activity recognition
707 using 700,000 person-days of wearable data. *npj Digital Medicine*, 7:91, 2024. doi: 10.1038/
708 s41746-024-01062-3. URL <https://doi.org/10.1038/s41746-024-01062-3>.

709 Junjie Zhang, Yuanhao Liu, and Hua Yuan. Attention-based residual bilstm networks for human
710 activity recognition. *IEEE Access*, 11:94173–94187, 2023a. doi: 10.1109/ACCESS.2023.3310269.

711
712 Mi Zhang and Alexander A. Sawchuk. Usc-had: a daily activity dataset for ubiquitous activity
713 recognition using wearable sensors. In *Proceedings of the 2012 ACM Conference on Ubiquitous*
714 *Computing*, UbiComp ’12, pp. 1036–1043, New York, NY, USA, 2012. Association for Computing
715 Machinery. ISBN 9781450312240. doi: 10.1145/2370216.2370438. URL [https://doi.org/](https://doi.org/10.1145/2370216.2370438)
716 [10.1145/2370216.2370438](https://doi.org/10.1145/2370216.2370438).

717 Zhanpeng Zhang, Wenting Wang, An Aimin, Yuwei Qin, and Fazhi Yang. A human activity
718 recognition method using wearable sensors based on convtransformer model. *Evolving Systems*,
719 14:1–17, 01 2023b. doi: 10.1007/s12530-022-09480-y.

720
721 Yida Zhu, Haiyong Luo, Song Guo, and Fang Zhao. Dmstl: A deep multi-scale transfer learning
722 framework for unsupervised cross-position human activity recognition. *IEEE Internet of Things*
723 *Journal*, 10(1):787–800, 2023. doi: 10.1109/JIOT.2022.3204542.

724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755