DOMINO: A Dataset for Context-Aware Human Activity Recognition using Mobile Devices

1st Luca Arrotta
Dept. of Computer Science
University of Milan
Milan, Italy
luca.arrotta@unimi.it

2nd Gabriele Civitarese
Dept. of Computer Science
University of Milan
Milan, Italy
gabriele.civitarese@unimi.it

3rd Riccardo Presotto
Dept. of Computer Science
University of Milan
Milan, Italy
riccardo.presotto@unimi.it

4th Claudio Bettini
Dept. of Computer Science
University of Milan
Milan, Italy
claudio.bettini@unimi.it

***Abstract*—Human Activity Recognition (HAR) with mobile and wearable devices has been deeply studied in the last decades. Research groups working on this topic evaluated their proposed methods mostly on public datasets. However, most of the existing datasets only include inertial sensor data, while it is well-known that additional context data (e.g., semantic location) has the potential to significantly improve the recognition rate. Only a few datasets for context-aware HAR are publicly available, and their annotations were mostly self-reported in-the-wild by the subjects involved in data acquisition. This method harms the quality of annotations, thus discouraging the application of supervised models. In this paper, we propose DOMINO, a new public dataset for context-aware HAR. DOMINO includes 25 users (wearing a smartphone and a smartwatch) performing 14 activities. During data acquisition, the mobile devices recorded both inertial and high-level context data while our team monitored the quality of the self-reported annotations. Our experiments on DOMINO show the positive impact of considering high-level context information for Human Activity Recognition.**

***Index Terms*—human activity recognition, context-awareness, dataset**

I. Introduction

Sensor-based Human Activity Recognition (HAR) based on mobile and wearable devices has been widely studied in recent years due to its applications to healthcare and well-being. The majority of the existing works mainly consider inertial sensor data (e.g., accelerometer, gyroscope) to infer low-level physical activities with data-driven approaches [1]. The methods proposed in the literature are usually validated on public datasets. Most of these datasets were collected involving subjects wearing ad-hoc devices equipped with inertial sensors. Among these, a few datasets (e.g., UCI-HAR [2], LTMM [3], and KU-HAR [4]) considered unusual body positions for the devices (e.g., waist, center lower back) that may be considered unrealistic, at least for current applications. Other datasets like OPPORTUNITY [5], PAMAP2 [6], and others [7]–[12] performed data acquisition by placing such devices also on the hip and/or the wrist, positions where people typically carry personal smartphones and smartwatches. On the other hand, some datasets directly include the raw data collected by the inertial sensors installed on off-the-shelf mobile/wearable devices like smartphones and smartwatches. In these cases, smartphones are usually carried in one of the trouser's front pockets (e.g., in WISDM [13], MobiAct [14], and UniMiB SHAR [15]). Some datasets simulated the smartwatch by requiring subjects to wear a smartphone on the wrist, like in NTUT-HAR [16] and other datasets [17]–[19]. Finally, only the Heterogeneity Activity Recognition dataset directly includes smartphones' and smartwatches' sensor data [20].

However, the datasets presented above lack additional contextual information that mobile devices can collect (e.g., semantic position, noise level, weather). High-level context data have the potential to expand the set of recognizable activities, as well as to better discriminate them [21]. Unfortunately, only a few existing datasets include such information. For instance, the ExtraSensory dataset [22] includes inertial sensor measurements collected in-the-wild from the smartphone and the smartwatch of up to 60 users. From the smartphone, the authors also collected GPS, audio, and phone state information (e.g., battery status, WiFi connectivity). Moreover, ExtraSensory includes annotations about high-level context information, like the user's semantic place. The main disadvantage of ExtraSensory is the low quality and reliability of its annotations. Indeed, the subjects involved during data acquisition were also in charge of self-reporting the activities they performed (e.g., sitting, walking, running), as well as their current surrounding context (e.g., in a meeting, indoors, at home, with friends, phone in hand). In some cases, as explicitly stated by the authors [22], the subjects forgot the exact time of an activity they performed. More commonly, the users neglected to annotate relevant activities or contexts.

Finally, datasets like RealWorld [23] (scripted, involving 15 subjects) and Daily Log [24] (in-the-wild, involving 7 subjects) collected measurements from inertial sensors (i.e., accelerometer, gyroscope, magnetometer) and context data (e.g., GPS, microphone, and luminosity sensors) considering smartphones and smartwatches worn on different body positions (including one of the trousers' front pocket and wrist). The subjects were also in charge of self-reporting through a smartwatch app their current semantic location (e.g., home, office), low-level activity (e.g., running, sitting, standing), and high-level activity with the corresponding sub-activity (e.g., sport coupled with gym or basketball, transportation with bicycle or tram). The main drawbacks of these datasets are that participating subjects self-reported the activity annotations, with a negative impact on the correctness of the dataset's labels, thus discouraging the application of supervised machine learning solutions.

To the best of our knowledge, there are no public datasets for context-aware HAR ensuring a high quality of their annotations. In this work, we propose DOMINO, a novel *Dataset for cOntext-aware huMan actIvity recogNitiOn*. Our dataset contains context-aware HAR data collected from 25 different users wearing a smartphone and a smartwatch, including 14 different activities. We collected ≈ 9 hours of both inertial and context data. Since our dataset was acquired in a scripted scenario, the activity and context labels are accurate, thus allowing researchers to evaluate the impact of context data in activity recognition. However, the scripted setting did not allow us to collect a wide variety of different context data. Hence, we also augmented the dataset with synthetically generated context data that could be automatically acquired in real scenarios.

The contributions of this work are the following:

- We present a novel dataset for context-aware HAR[1].
- The dataset includes high-level context information that can be used to improve HAR purely based on inertial sensor data.
- We provide benchmarks that show the positive impact of context data, indicating the need to acquire accurate context data in real-world scenarios.

II. THE DOMINO DATASET

A. Dataset design

We designed the data collection campaign of DOMINO considering that the combination of inertial sensors data with the information about the user's surrounding context (e.g., proximity to public transportation routes) has the potential to expand the set of activities that an activity classifier can recognize. At the same time, context data may also be useful to better discriminate activities with similar motion patterns but typically performed in different contexts. For instance, even if going downstairs and walking share similar physical movements, when the user is outdoors, it is more likely that she is walking. For this reason, DOMINO includes a combination of inertial sensor and context data provided by the user's mobile devices (i.e., smartphone and smartwatch).

Overall, we planned to acquire data about 14 activities: *Brushing Teeth*, *Cycling*, *Elevator Down*, *Elevator Up*, *Lying*, *Moving by Car*, *Running*, *Sitting*, *Sitting on Transport*, *Stairs Down*, *Stairs Up*, *Standing*, *Standing on Transport*, and *Walking*. We designed 4 different scripted scenarios for data collection, represented in Table I. Each scenario (identified by a letter) is a template that presents the ordered sequence of activities a subject should perform during data acquisition. The flow of time is represented vertically from top to bottom, while horizontal dashed lines represent transitions between subsequent activities. The table also shows whether the activities should be performed indoors or outdoors and their suggested execution duration.

[1]The dataset can be downloaded here: https://tinyurl.com/domino-dataset

The raw measurements derived by the mobile devices' inertial sensors (i.e., accelerometer, gyroscope, and magnetometer) are crucial to monitor the user's physical movements. At the same time, mobile devices can collect data useful to derive high-level context information about the user's surroundings. For instance, the smartphone's barometer and GPS measure the user's height variations and speed, respectively. On the other hand, the microphone can reveal the environment's noise level. Moreover, additional context information can be derived by combining the smartphone's built-in sensors with public web services. *Google's Places API* provides the user's closest semantic places (e.g., university); *OpenWeatherMap* supplies current local weather conditions (e.g., rainy), while *Transitland* provides information about the public transportation routes and stops closest to the user. Finally, some context information can be obtained by post-processing the data collected from mobile devices. For instance, to detect whether a user is following a public transportation route, it is possible to combine the information provided by the GPS and *Transitland*.

B. Devices

Each subject involved in the dataset collection carried a smartphone (*LG Nexus 5X*) in the trousers' front pocket and wore a smartwatch (*LG G-watch R*) on the dominant hand's wrist. To allow data interchange during data acquisition, a Bluetooth (BT) connection is established between the smartphone and the smartwatch. All the data collected and annotated as will be explained in Section II-C were formatted by the mobile devices using JSON before being sent to a REST server implemented in Java. This server was finally in charge of storing the data in a MongoDB database.

C. Data collection and annotation

We acquired the dataset from the 20th of December 2017 to the 16th of January 2018 in Milan. Specifically, we considered several indoor and outdoor locations that were nearby our department building in Milan. During the dataset collection process, 25 subjects performed one or more of the 4 scripted scenarios we previously presented. 19 subjects performed only one scenario. On the other hand, 4 subjects performed two scenarios, while 1 subject performed three scenarios, and 1 subject performed all the scenarios we scripted. To increase data variability, the subjects were allowed to slightly modify the scenarios we assigned to them. For instance, the subjects could go upstairs/downstairs by taking the stairs instead of the elevator, and vice versa. Almost all the recruited subjects were students or researchers at our university. Overall, 68% were males, while the remaining 32% were females. All the recruited subjects were right-handed. Additional statistics about the subjects involved during data acquisition are described in Table II. Instead, Table III shows, for each scenario, the number of times it has been performed during data acquisition, as well as its average duration in minutes. Overall, Scenario A is the longest one since it includes moving multiple times by public transport. On the other hand, the shortest scenario is Scenario D because it does not contain long activities

TABLE I: The four scripted scenarios of DOMINO

A	B	C	D
Sitting (indoor) 4 min	Moving by Car (outdoor) 4 min	Walking (indoor) 0.5 min	Walking (indoor) 0.5 min
Standing (indoor) 2 min	Walking (outdoor) 1 min	Elevator Down (indoor) 0.5 min	Standing (indoor) 0.5 min
Walking (indoor) 1 min	Stairs Up (indoor) 0.5 min	Walking (indoor-outdoor) 1 min	Elevator Down (indoor) 0.5 min
Stairs Down (indoor) 0.5 min	Sitting (indoor) 6 min	Running (outdoor) 3 min	Stairs Up (indoor) 0.5 min
Walking (indoor-outdoor) 2 min	Standing (indoor) 1 min	Cycling (outdoor) 3 min	Walking (indoor) 0.5 min
Standing (outdoor) 4/5 min	Walking (indoor) 0.5 min	Walking (outdoor) 2 min	Sitting (indoor) 4 min
Standing/Sitting on Transport (outdoor) 3 min	Brushing Teeth (indoor) 2 min	Sitting (outdoor) 2 min	Walking (indoor) 4 min
Walking (outdoor) 2 min	Walking (indoor) 0.5 min	Standing (outdoor) 1 min	Stairs Down (indoor) 0.5 min
Stairs Down (outdoor) 0.5 min	Lying (indoor) 4 min	Walking (outdoor-indoor) 2 min	Standing (indoor) 1 min
Standing (outdoor) 4/5 min		Elevator Up (indoor) 0.5 min	Elevator Down (indoor) 0.5 min
Standing/Sitting on Transport (outdoor) 2 min		Walking (indoor) 0.5 min	
		Sitting (indoor) 3 min	

TABLE II: Users' statistics

	Range	Mean (Standard deviation)
Age (years)	20-59	26.6 ($\pm$ 9.8)
Height (cm)	157-192	174.2 ($\pm$ 8.5)
Weight (kg)	48-92	66.2 ($\pm$ 13.0)

TABLE III: Scenarios' statistics

Scenario ID	Number of instances	Average recorded minutes (standard deviation)
A	9	36.9 ($\pm$ 3.8)
B	8	25.0 ($\pm$ 5.7)
C	7	23.9 ($\pm$ 2.8)
D	10	13.2 ($\pm$ 9.5)

that involve moving with a vehicle (e.g., *standing/sitting on transport*).

We installed a mobile application in charge of collecting and annotating data on the smartphone and the smartwatch of the user. To collect inertial sensor data, we considered the maximum sampling rate of such mobile devices, i.e., 200 Hz for the smartphone and 140 Hz for the smartwatch. At each second, the smartwatch sent its measurements to the smartphone through the BT channel. Hence, every 3 seconds, the smartphone sent to the server all the measurements it collected, coupled with the ones received by the smartwatch. Moreover, the smartphone app was also in charge of collecting context data. Since context information does not change frequently, it was collected and sent to the server every 15 seconds.

During data acquisition, the subjects also annotated in real-time the activities they were performing. Each person interacted with the smartwatch app to specify the type of activity she was going to execute before actually performing it. Figure 1 shows a screenshot of the user interface we implemented in the smartwatch app that enabled the users to annotate their activities. During data collection, a person of our team followed the subject to verify the correctness of her annotations and to annotate through another smartphone information about her current surrounding context. Unfortunately, the barometer and GPS readings through the mobile app were sometimes unreliable due to software issues. Despite an intensive data cleaning process after the dataset collection campaign, we decided to include in DOMINO only annotations of context data manually inserted by the research team members during data acquisition instead of the noisy ones actually collected through the mobile app.

Fig. 1: A screenshot of the app used to annotate activities data

Moreover, since we collected activity data through scripted scenarios, it is challenging to acquire data in all the possible context conditions where an activity can take place. Consider, for instance, the different weather conditions in which a person can go *running* outdoors. Figure 2 shows the distribution of two context conditions (i.e., indoors/outdoors and weather) in which *running* was performed during the data collection campaign. First, the involved subjects only went *running*

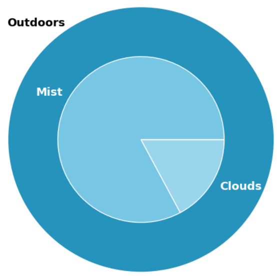

Fig. 2: Subset of actual context conditions (i.e., weather, indoors/outdoors) of *running*

outdoors. Moreover, this activity was performed only under two weather conditions, i.e., misty or cloudy. This happened since we collected DOMINO during winter in Milan. Since the actual context data we collected were not sufficiently diverse, we decided to include in DOMINO two alternative annotated context data: the actual context that surrounded the subjects during data acquisition and augmented context data we simulated to cover a larger amount of context situations in which each activity can take place. We generated such context data by considering common-sense knowledge about the HAR domain (e.g., *running* typically takes place outdoors, but the user could also run at home or inside a gym). For instance, Figure 3 shows the distribution of the indoors/outdoors and weather context conditions in which *running* takes place in the augmented version of the dataset. In this case, *running*

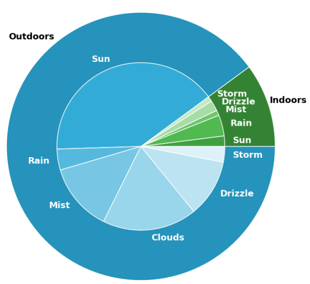

Fig. 3: Subset of augmented context conditions of *running*

still takes place outdoors in most cases ($\approx 90\%$), but it can also occur indoors. In the former case, it typically occurs when it is sunny. In the latter, it mostly takes place in the gym (90% of the time), but also at home.

Table IV shows the different types of context information included in DOMINO with the corresponding possible values, considering both actual and augmented context data. In particular, the subject's presence in an indoor or an outdoor environment was derived from the information about the user's closest semantic place. Moreover, we discretized the speed values provided by the GPS sensor of the subject's smartphone into four possible values, i.e., *null*, *low*, *medium*, and *high*. Intuitively, a low speed was typically measured while the subjects were walking. On the other hand, a medium speed was the maximum speed that occurred while the subjects were running. Higher speeds were measured only while moving by bike, car, or public transportation. Moreover, the combination of the information provided by the GPS sensor and by *Transitland* was used to detect whether the subject was following a public transportation route. Finally, we discretized the measurements provided by the smartphone's barometer into three different height variations: *negative*, *null*, and *positive*.

TABLE IV: Type of context data included in DOMINO

Context information	Augmented context values	Actual context values
Indoor / Outdoor	Indoor, Outdoor	Indoor, Outdoor
Semantic place	Home, Office, University, Mall, Station, Museum, Gym, Shop, Bar, Restaurant, Barbershop, Bank, Church, Null	Home, University, Null
Speed	Null, Low, Medium, High	Null, Low, Medium, High
Weather	Sun, Rain, Mist, Clouds, Drizzle, Storm	Rain, Mist, Clouds, Drizzle
Following Public Transportation Route	True, False	True, False
Height Variation	Negative, Null, Positive	Negative, Null, Positive
Audio Level	Low, Medium, High	Low, Medium, High

TABLE V: Activities' statistics

Activity	Overall recorded minutes	Average recorded minutes per instance (standard deviation)	Instances
Brushing Teeth	14.4	1.4 (± 0.7)	10
Cycling	21.8	2.7 (± 1.0)	8
Elevator Down	12.9	0.4 (± 0.1)	29
Elevator Up	8.4	0.4 (± 0.1)	22
Lying	33.1	1.7 (± 1.5)	19
Moving by Car	12.8	2.6 (± 1.6)	5
Running	22.9	1.4 (± 0.6)	16
Sitting	129.9	2.0 (± 0.8)	64
Sitting on Transport	14.4	2.4 (± 1.4)	6
Stairs Down	17.9	0.5 (± 0.1)	37
Stairs Up	12.8	0.5 (± 0.1)	25
Standing	127.6	1.8 (± 1.7)	69
Standing on Transport	22.3	1.9 (± 1.0)	12
Walking	100.5	1.1 (± 1.1)	92
Total	551.7		414

Finally, Table V describes statistics about the activities included in DOMINO. Overall, we acquired 551.7 minutes (≈ 9 hours) of labeled data for 414 different activity instances[2].

III. Benchmarks

In this section, we provide some benchmarks on DOMINO that researchers could consider for future work on context-aware HAR.

A. Data pre-processing

In our experimental setup, we consider state-of-the-art hyper-parameters to pre-process sensor data [21]. We segment

[2]An activity instance is a specific occurrence of the activity in the dataset (e.g., John performed *walking* from 10:20AM to 10:40AM). By using segmentation, it is possible to derive several samples from the same instance.

each mobile device's inertial stream into non-overlapping windows of 4 seconds. To reduce the intrinsic noise of raw sensor data, for each segmentation window, we apply (on each axis of each inertial sensor) a median filter of size 3 to the measurements provided. Finally, since the Android OS does not guarantee a constant sampling rate, we downsample inertial sensor data to 50 Hz, a widely adopted sampling rate in many public HAR datasets [2], [19], [23]. Hence, for each mobile device, each window is represented as a matrix of shape $(200, 9)$[3]. Context data collected by the smartphone are also segmented in time windows of 4 seconds. In particular, the context values described in Table IV are one-hot encoded and concatenated to obtain vectors of 34 elements.

B. Adopted activity models

For the sake of this work, we rely on a Convolutional Neural Network (CNN) to capture spatio-temporal dependencies from sensor data as usually proposed for sensor-based HAR [25]–[27]. Even though more complex models have been recently proposed in the literature [1], we considered a standard solution to evaluate the impact of context data. The architecture of our network is inspired by the one recently proposed in [28]. Specifically, our CNN is fed with three different input flows: 1) time windows of inertial sensor data from the smartphone, 2) time windows of inertial sensor data from the smartwatch, and 3) the one-hot encoded context data. Each inertial data flow starts with two convolutional layers with 8 3×3 and 64 2×2 filters, respectively. Each of these layers is followed by a 2×2 max pooling layer. Then, we added a flatten and fully connected layer with 128 neurons. On the other hand, the context data flow is processed by a single fully connected layer composed of 8 neurons. The features automatically extracted from the three input flows are then concatenated and provided to a dropout layer with a dropout rate of 0.1 and then to a fully connected layer with 256 neurons. Finally, the classification is obtained through the last fully connected layer, which relies on the softmax activation function.

C. Evaluation methodology

In order to evaluate the impact of context data in HAR, we designed four different evaluation strategies:

- *inertial only*: a variant of the CNN that is only fed with the inertial sensors' data provided by the smartphone and the smartwatch
- *actual*: the CNN trained and tested on inertial data and actual context data included in DOMINO.
- *augmented*: the CNN trained and tested on inertial data and augmented context data of DOMINO.
- *actual-augmented*: the CNN trained with inertial data and actual context data included in DOMINO and tested on inertial data and augmented context data; the goal of this methodology is to evaluate if the CNN trained on the actual context data can generalize over the more realistic context situations encoded in the augmented context data of DOMINO.

[3]200 represents 50 samples for each second, on a time window of 4 seconds. 9 is the number of axes considering the sensors on a mobile device (i.e., accelerometer, gyroscope, and magnetometer)

For each evaluation strategy, we adopt a leave-one-subject-out cross-validation technique. At each fold, the test set includes one user's data, while the training and the validation set consist of the 90% and the 10% of all the remaining data, respectively. The recognition rate is evaluated in terms of F1 score on the test set of each fold. We consider 200 epochs with batches of 32 samples during training and an early stopping strategy that terminates the learning process when the validation loss does not improve for 5 consecutive epochs.

D. Results

Figure 4 compares the previously described evaluation strategies.

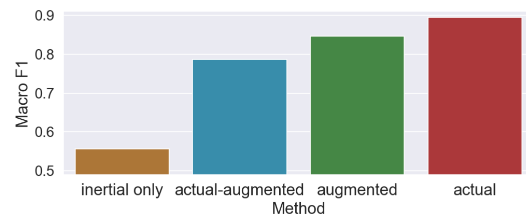

Fig. 4: Comparison between the considered methodologies

In general, considering context information leads to a significant improvement with respect to considering inertial sensors only. The *actual* strategy is the one reaching the highest F1 score (89%); however, it is based on training and testing the CNN on the actual context data of DOMINO, which does not cover a wide range of context situations due to its scripted nature. Hence, this result does not guarantee that the classifier would reach satisfying performance also considering more realistic context scenarios. Indeed, among the evaluation methodologies including context data, the lowest F1 score (79%) is obtained by *actual-augmented*. This strategy trains the CNN with the actual context data but tests it on the augmented data that better captures the different context conditions in which the activities can take place. This result shows the low generalization capabilities of the CNN when it is trained on the actual context data. Hence, we believe that the most reliable benchmark is represented by the results obtained with the *augmented* strategy (F1 score is 85%), which considers more realistic context data both for training and testing.

In particular, the *augmented* evaluation methodology exhibits a significant increment of $+29\%$ compared to *inertial only* in terms of macro F1 score. Figure 5 shows the same result at the activity granularity. The major improvements are related to those activities that are hard to discriminate only based on inertial data since they share similar motion patterns, but they are executed in different contexts. For instance, without information about height variations, it is challenging to distinguish *standing* from *elevator up/down* since the user stands still in both situations. At the same time, without knowing the user's speed and whether she is following

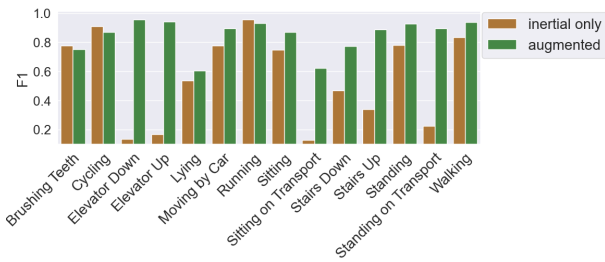

Fig. 5: Activity level results

a public transportation route, it is non-trivial to distinguish *standing/sitting* from *standing/sitting on transport*. As a drawback, context data may slightly reduce the recognition rates of poorly represented activities that share context conditions with strongly represented activities. For instance, by including context information in the network, the *running* activity is slightly more confused with *walking*.

IV. Conclusion and Future Work

In this paper, we presented DOMINO, a novel dataset for context-aware HAR. We believe that researchers could adopt this dataset to design new methods combining inertial and context sensor data. For instance, hybrid knowledge-based and data-driven approaches are a promising direction in this area [21], [28]. The major limitation of this dataset is its scripted nature. In future work, we will consider performing a large-scale in-the-wild campaign to collect such data while maintaining high-quality annotations.

Acknowledgement

Part of this research was supported by the MUSA – Multilayered Urban Sustainability Action – project, funded by the European Union – NextGenerationEU, under the National Recovery and Resilience Plan (NRRP) Mission 4 Component 2 Investment Line 1.5: Strenghtening of research structures and creation of R&D "innovation ecosystems", set up of "territorial leaders in R&D".

References

[1] F. Gu, M.-H. Chung, M. Chignell, S. Valaee, B. Zhou, and X. Liu, "A survey on deep learning for human activity recognition," *ACM Computing Surveys (CSUR)*, vol. 54, no. 8, pp. 1–34, 2021.

[2] D. Anguita, A. Ghio, L. Oneto, X. Parra Perez, and J. L. Reyes Ortiz, "A public domain dataset for human activity recognition using smartphones," in *21th ESANN Symposium*, 2013, pp. 437–442.

[3] E. A. Ihlen, A. Weiss, J. L. Helbostad, and J. M. Hausdorff, "The discriminant value of phase-dependent local dynamic stability of daily life walking in older adult community-dwelling fallers and nonfallers," *BioMed research international*, vol. 2015, 2015.

[4] N. Sikder and A.-A. Nahid, "Ku-har: An open dataset for heterogeneous human activity recognition," *Pattern Recog Lett*, vol. 146, pp. 46–54, 2021.

[5] D. Roggen, A. Calatroni, M. Rossi, T. Holleczek, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, A. Ferscha *et al.*, "Collecting complex activity datasets in highly rich networked sensor environments," in *Seventh INSS conference*. IEEE, 2010, pp. 233–240.

[6] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," in *2012 16th international symposium on wearable computers*. IEEE, 2012, pp. 108–109.

[7] M. Zhang and A. A. Sawchuk, "Usc-had: A daily activity dataset for ubiquitous activity recognition using wearable sensors," in *Proceedings of the 2012 ACM conference on ubiquitous computing*, 2012, pp. 1036–1043.

[8] O. Banos, C. Villalonga, R. Garcia, A. Saez, M. Damas, J. A. Holgado-Terriza, S. Lee, H. Pomares, and I. Rojas, "Design, implementation and validation of a novel open framework for agile development of mobile health applications," *Biomedical engineering online*, vol. 14, no. 2, pp. 1–20, 2015.

[9] B. Bruno, F. Mastrogiovanni, and A. Sgorbissa, "A public domain dataset for adl recognition using wrist-placed accelerometers," in *the 23rd IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2014, pp. 738–743.

[10] C. Chen, R. Jafari, and N. Kehtarnavaz, "Utd-mhad: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor," in *IEEE ICIP conference)*. IEEE, 2015, pp. 168–172.

[11] X. Chao, Z. Hou, and Y. Mo, "Czu-mhad: A multimodal dataset for human action recognition utilizing a depth camera and 10 wearable inertial sensors," *IEEE Sensors Journal*, vol. 22, no. 7, pp. 7034–7042, 2022.

[12] K. Altun and B. Barshan, "Human activity recognition using inertial/magnetic sensor units," in *International workshop on human behavior understanding*. Springer, 2010, pp. 38–51.

[13] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SigKDD Explorations Newsletter*, vol. 12, no. 2, pp. 74–82, 2011.

[14] C. Chatzaki, M. Pediaditis, G. Vavoulas, and M. Tsiknakis, "Human daily activity and fall recognition using a smartphone's acceleration sensor," in *International Conference on Information and Communication Technologies for Ageing Well and e-Health*. Springer, 2016, pp. 100–118.

[15] D. Micucci, M. Mobilio, and P. Napoletano, "Unimib shar: A dataset for human activity recognition using acceleration data from smartphones," *Applied Sciences*, vol. 7, no. 10, p. 1101, 2017.

[16] C. Betancourt, W.-H. Chen, and C.-W. Kuan, "Self-attention networks for human activity recognition using wearable devices," in *IEEE SMC conference*. IEEE, 2020, pp. 1194–1199.

[17] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. Havinga, "Complex human activity recognition using smartphone and wrist-worn motion sensors," *Sensors*, vol. 16, no. 4, p. 426, 2016.

[18] M. Shoaib, H. Scholten, and P. J. Havinga, "Towards physical activity recognition using smartphone sensors," in *IEEE 10th UIC and 2013 IEEE 10th ATC conferences*. IEEE, 2013, pp. 80–87.

[19] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. Havinga, "Fusion of smartphone motion sensors for physical activity recognition," *Sensors*, vol. 14, no. 6, pp. 10 146–10 176, 2014.

[20] A. Stisen, H. Blunck, S. Bhattacharya, T. S. Prentow, M. B. Kjærgaard, A. Dey, T. Sonne, and M. M. Jensen, "Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition," in *Ih ACM SenSys conference*, 2015, pp. 127–140.

[21] C. Bettini, G. Civitarese, and R. Presotto, "Caviar: Context-driven active and incremental activity recognition," *Knowl-Based Syst*, vol. 196, p. 105816, 2020.

[22] Y. Vaizman, K. Ellis, and G. Lanckriet, "Recognizing detailed human context in the wild from smartphones and smartwatches," *IEEE pervasive computing*, vol. 16, no. 4, pp. 62–74, 2017.

[23] T. Sztyler and H. Stuckenschmidt, "On-body localization of wearable devices: An investigation of position-aware activity recognition," in *IEEE PerCom Conference*. IEEE, 2016, pp. 1–9.

[24] T. Sztyler, J. Carmona, J. Völker, and H. Stuckenschmidt, "Self-tracking reloaded: applying process mining to personalized health care from labeled sensor data," in *Transactions on Petri nets and other models of concurrency XI*. Springer, 2016, pp. 160–180.

[25] B. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu, "Convolutional neural networks for time series classification," *Journal of Systems Engineering and Electronics*, vol. 28, no. 1, pp. 162–169, 2017.

[26] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Exp Sys Appl*, vol. 59, pp. 235–244, 2016.

[27] S. Ha, J.-M. Yun, and S. Choi, "Multi-modal convolutional neural networks for activity recognition," in *2015 IEEE International conference on systems, man, and cybernetics*. IEEE, 2015, pp. 3017–3022.

[28] L. Arrotta, G. Civitarese, and C. Bettini, "Knowledge infusion for context-aware sensor-based human activity recognition," in *2022 IEEE International Conference on Smart Computing (SMARTCOMP)*. IEEE, 2022, pp. 1–8.