Towards Deep Clustering of Human Activities from Wearables

Alireza Abedin
The University of Adelaide
alireza.abedinvaramin@adelaide.edu.au

Farbod Motlagh
The University of Adelaide
farbod.motlagh@student.adelaide.edu.au

Qinfeng Shi
The University of Adelaide
javen.shi@adelaide.edu.au

Hamid Rezatofighi
The University of Adelaide
hamid.rezatofighi@adelaide.edu.au

Damith Ranasinghe
The University of Adelaide
damith.ranasinghe@adelaide.edu.au

ABSTRACT

Our ability to exploit low-cost wearable sensing modalities for critical human behaviour and activity monitoring applications in health and wellness is reliant on supervised learning regimes; here, deep learning paradigms have proven extremely successful in learning activity representations from annotated data. However, the costly work of gathering and annotating sensory activity datasets is labor intensive, time consuming and not scalable to large volumes of data. While existing unsupervised remedies of deep clustering leverage network architectures and optimization objectives that are tailored for static image datasets, deep architectures to uncover cluster structures from raw sequence data captured by on-body sensors remains largely unexplored. In this paper, we develop an unsupervised end-to-end learning strategy for the fundamental problem of human activity recognition (HAR) from wearables. Through extensive experiments, including comparisons with existing methods, we show the effectiveness of our approach to jointly learn *unsupervised representations* for sensory data and generate *cluster assignments* with strong semantic correspondence to distinct human activities.

KEYWORDS

Activity Recognition; Deep learning; Clustering; Wearable Sensors

1 INTRODUCTION

Accurately and precisely understanding human activities is the basis for applications ranging from assessing our cognitive decline, physical and mental health to performance in sporting activities [10–13, 17, 24, 27, 32]. Increasing plethora of wearables are providing the opportunity to conveniently and at low-cost collect *fine-grained* physiological information to understand human activities. However, the premise for realizing the multitude of applications is our ability to build accurate and, often personalized, models for recognizing human activities from wearables.

Problem. Human activity recognition problems have relied predominantly on supervised learning regimes where deep learning paradigms are extremely successful in learning activity representations from annotated data. While the process of collection and annotation may be retrospective with vision based sensing modalities where visual inspections of, for example, video frames provides the basis for ground truth, the parallel task with wearables is nearly impossible. Moreover, such methods cannot be easily scaled to gather large datasets often necessary for deep neural networks (DNNs). In comparison to other domains, generating labelled data to benefit from supervised learning methods to build HAR applications in the absence of a reliable visualisation to establish ground truth is a unique HAR problem with wearable sensors.

Our Motivation. Although unsupervised methods provide avenues for learning from unlabelled data, investigations of unsupervised learning from wearable multi-channel time-series data remains dominantly limited to *pre-training* [3, 36] or *unsupervised representation learning* [6, 19]. Unsupervised alternatives without requiring any labels, such as *deep clustering*, exist for image data, however, these frameworks are tailored for still images and lack the *inherent* capability to learn representations and clusters from raw sequential data captured by wearables. Therefore, our motivation is to investigate and develop a deep clustering architecture that:

- Leverages the inherently sequential nature of sensory data.
- Learns *clustering friendly* representations of activity features in the multi-sensor and multi-channel input signals that offer separability of activity classes in the feature space.
- Promotes the formation of highly discriminative clusters with high semantic correspondence to human activities.

Our Contributions. In this study, we propose *Deep Sensory Clustering*—a deep clustering architecture that learns highly discriminative representations using self-supervision with *reconstruction and future prediction tasks informed by feedback from a clustering objective to guide the network towards clustering-friendly representations.* The self-supervised tasks intend to incentivize the network to learn salient activity features that offer semantic separation in the feature space while simultaneously reducing the risk of collapsing clusters. Further, we augment the optimization objective with a clustering-oriented criterion to further refine the feature representations and gradually promote clustering-friendliness in the feature space. We validate our design concepts through extensive experiments; we summarize our key contributions below:

(1) We develop an *unsupervised* deep learning network architecture for clustering human activities from raw sequences of wearable sensor data streams. Our approach, to the best of knowledge, provides the *first* standalone, end-to-end, deep clustering method for *raw* sequential data from wearables.
(2) Through a systematic experimental regime conducted on three diverse HAR benchmark datasets (`UCI HAR`, `Skoda`, `MHEALTH`), we demonstrate the effectiveness of our proposed approach. Further, we compare our method with closely related approaches, including traditional clustering methods.

2 RELATED WORKS

HAR with Wearable Sensors. The superior performance of supervised deep neural networks in classification tasks has led to a shift towards the adoption of deep learning paradigms for recognizing human activities from raw wearable data [2, 20, 23, 30]. Researchers have explored CNNs [8, 33, 39–41], RNNs [15, 18], and a combination of convolutional and recurrent layers [1, 29, 31] to effectively model the temporal dependencies inherent in sequences captured with sensors. However, acquisition of labeled sensory data is labor-intensive and time-consuming. But, in the sequential sensor data domain, unsupervised learning has merely been investigated as a means for weight initialization [3, 19], unsupervised feature learning prior to supervised fine tuning with labels [6, 36] or clustering of handcrafted features [22], rather than a standalone end-to-end approach for exploiting cheaply accessible raw unlabelled data.

Clustering with Deep Neural Networks. Recently, representation learning power of DNNs has been leveraged to achieve clustering-friendly representations and cluster assignments simultaneously for still image data; a shift towards *Deep Clustering* paradigms [28]. In this regard, the feature space for representing images are initialized using deep autoencoders and iteratively refined to obtain cluster assignments [25, 37]. Following similar ideas, Chen et al. [9] propose a locality preserving criteria to learn structure preserving image representations, and Dizaji et al. [14] encourage balanced cluster assignments during training. In another study, a CNN is trained with agglomerative clustering objective in a recurrent process [38]. Although these methods achieve impressive results for computer vision applications, existing deep clustering frameworks are tailored for still image datasets and suffer from their inability to exploit the sequential nature of wearable sensor data streams to learn representations and generate clusters of activities as substantiated in Section 4.1.

3 THE PROPOSED FRAMEWORK

We consider the problem of clustering a set of n unlabelled segments of sensory readings $\{x_i\}_{i=1}^{n}$ into k clusters, each representing a semantic human activity category. These segments are obtained by applying a sliding window of fixed temporal duration δt over d sensor channels of recorded datastreams. We propose our unsupervised two-staged *Deep Sensory Clustering* framework illustrated in Fig. 1 for the problem and detail our approach in what follows.

3.1 Stage I: Pretraining a Multi-Task Autoencoder

In order to facilitate learning clustering-friendly representations, we initialize the feature space by pretraining a recurrent autoencoder to accomplish *auxiliary tasks* in an unsupervised fashion. In accomplishing the delegated tasks, the network is *forced to extract enriched representations* from the multi-channel sensor sequences.

Recurrent Encoder (Enc_θ). The encoder component of our network consumes a windowed excerpt of a raw multi-channel sensory sequence and learns a compact fixed length representation as a holistic summary of the input. In particular, we adopt a bi-directional GRU that reads through the partitioned sensory sequence $\mathbf{x}$ in both forward and backward directions and updates its internal hidden state in each time step according to the received input. The final hidden state obtained after scanning the entire input sequence is reduced in dimensionality through a fully connected layer. The resulting low-dimensional embedded feature, denoted by $\mathbf{z} \in \mathbb{R}^z$, encodes contextual activity information by modeling the temporal dependencies present in the input sequence of sensory measurements $\mathbf{x}$. We summarize the parameterized operations associated with encoding the input sequence $\mathbf{x}_i$ as $\mathbf{z}_i = \mathrm{Enc}_\theta(\mathbf{x}_i)$.

Conditional Recurrent Decoders (Dec_ϕ). The decoder modules of the framework are structured symmetrically to the encoder component. First, a context vector is achieved by back projecting the embedded representation from the encoder into a higher-dimensional space such that it can be used to initialize the hidden states for the decoders. *Two recurrent decoders then jointly exploit the generated context vector to accomplish different self-supervised tasks without requiring any manual supervision.* Inspired by [34], we share the encoder network between decoders with two different expertise; one decoder is specialized to reconstruct the temporally inversed input sequence, while the other one learns to *anticipate the future sensory measurements* that should follow after, conditioned on the encoded input representation. Hence, the network has to not only learn a representation enriched with sufficient information to reproduce the input sequence, but also features that allow extrapolating future measurements. We summarize the parameterized decoding process as $(\hat{\mathbf{y}}_i^{\mathrm{rec}}, \hat{\mathbf{y}}_i^{\mathrm{fut}}) = \mathrm{Dec}_\theta(\mathbf{z}_i)$, where $\hat{\mathbf{y}}_i^{\mathrm{rec}}$ and $\hat{\mathbf{y}}_i^{\mathrm{fut}}$ respectively denote the reconstructed and the anticipated sequences generated from the input $\mathbf{x}_i$ to satisfy the tasks.

Pre-training Objective. We pre-train the entire recurrent autoencoder with a joint objective function,

$$\mathcal{L}_{\mathrm{AE}}^{(i)} = \mathcal{L}_{\mathrm{rec}}^{(i)} + \mathcal{L}_{\mathrm{fut}}^{(i)} = \underbrace{\|\mathbf{y}_i^{\mathrm{rec}} - \hat{\mathbf{y}}_i^{\mathrm{rec}}\|^2}_{\text{reconstruction loss}} + \underbrace{\|\mathbf{y}_i^{\mathrm{fut}} - \hat{\mathbf{y}}_i^{\mathrm{fut}}\|^2}_{\text{future prediction loss}}, \quad (1)$$

where $\mathcal{L}_{\mathrm{rec}}$ and $\mathcal{L}_{\mathrm{fut}}$ denote the mean square error between each decoder's generated output sequence (*i.e.*, $\hat{\mathbf{y}}^{\mathrm{rec}}$ and $\hat{\mathbf{y}}^{\mathrm{fut}}$) and the expected ground-truth target sequences (*i.e.*, $\mathbf{y}^{\mathrm{rec}}$ and $\mathbf{y}^{\mathrm{fut}}$). Once the training is complete and the discrepancy between the generated outputs and their corresponding target sequences is minimized, the optimal network parameters, *i.e.*, $(\theta^*, \phi^*) = \min_{\theta,\phi} \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}_{\mathrm{AE}}^{(i)}$, serve as an initialization point for the second stage.

3.2 Stage II: Representation Refinement with a Clustering Criterion

Once the autoencoder becomes proficient in accomplishing the auxiliary tasks, we expect the feature space to find a semantic orientation. We further, extend our framework with a parameterized clustering network $f_\omega(.)$ capable of estimating cluster assignment distributions and iteratively optimize a clustering objective $\mathcal{L}_{\mathrm{C}}$ to refine the feature space and guide the network towards yielding clustering-friendly representations. In this paper, we incorporate *Cluster Assignment Hardening* [37] as a representative centroid-based approach for further refinement of the established feature space. During the refinement stage, both the clustering loss $\mathcal{L}_{\mathrm{C}}$ and the autoencoding objectives $\mathcal{L}_{\mathrm{AE}}$ are jointly incorporated to be optimized. Hence, the aggregated optimization criterion, for

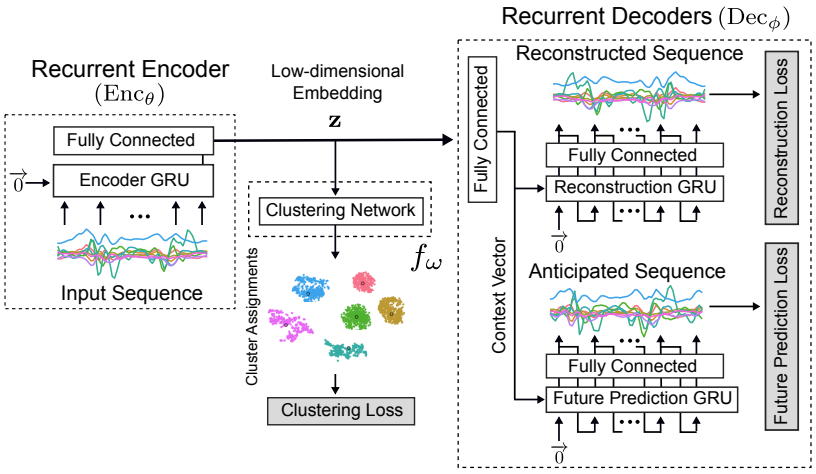

Figure 1: An Overview of *Deep Sensory Clustering* pipeline.

instance i, is formulated as

$$\mathcal{L}^{(i)} = \gamma \mathcal{L}_{\mathrm{C}}^{(i)} + \mathcal{L}_{\mathrm{AE}}^{(i)}, \tag{2}$$

where the coefficient $\gamma \in [0, 1]$ controls contribution of the clustering objective. Notably, we chose not to discard the decoding tasks during the refinement step to preserve the local data structure and allow a smoother manipulation of the feature space without distorting the previously established one. Once the network parameters are optimized with respect to the global criterion, $(\theta^*, \phi^*, \omega^*) = \min_{\theta,\phi,\omega} \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}^{(i)}$, the clustering network of our framework directly delivers cluster assignments *without* requiring a separate clustering algorithm to be run on the representations in a decoupled process. We describe the clustering criterion utilized next.

Cluster Assignment Hardening (CAH). This clustering objective leverages the similarities between the data representations and the cluster centroids as a kernel to compute soft cluster assignments. Placing emphasis on the high confidence assignments, it then purifies the clusters and forces the assignments to have stricter probabilities. To incorporate this method, our clustering network f_ω comprises a single layer which maintains the cluster centroids $(\omega_j \in \mathbb{R}^{\mathrm{z}})_{j=1}^{k}$ as tunable network parameters and generates assignment distributions $Q_i = f_\omega(\mathbf{z}_i)$ for each instance i. This layer follows the Student's t-distribution to measure the similarity of embedded sequence representation $\mathbf{z}_i \in \mathbb{R}^{\mathrm{z}}$ to the k cluster centroids and therefore, obtains the normalized similarities $Q_i = (q_{ij})_{j=1}^{k}$,

$$q_{ij} = \frac{(1 + \|\mathbf{z}_i - \omega_j\|^2)^{-1}}{\sum_{j'=1}^{k} (1 + \|\mathbf{z}_i - \omega'_j\|^2)^{-1}}. \tag{3}$$

Through squaring this distribution and then normalizing it, an auxiliary target distribution $P_i = (p_{ij})_{j=1}^{k}$ that leverages high confidence assignments is then defined to point the learning process towards stricter cluster assignments.

$$p_{ij} = \frac{q_{ij}^2 / \sum_{i=1}^{n} q_{ij}}{\sum_{j'=1}^{k} (q_{ij'}^2 / \sum_{i=1}^{n} q_{ij'})}. \tag{4}$$

Table 1: A summary of the datasets explored in this work.

Dataset	UCI HAR	Skoda	MHEALTH
Sensor Sampling Rate	50Hz	33Hz	50Hz
Sliding Window Duration (δt)	2.56s	1s	2.56s
Number of Sensor Channels (d)	9	60	23
Number of Activity Categories (k)	6	10	12
Number of Training Segments	7352	5448	4088
Number of Testing Segments	2947	718	1022

Subsequently, the soft assignment distribution Q_i is iteratively purified through minimizing the Kullback-Leilbler (KL) divergence between the soft labels and the auxiliary target distribution via training the network parameters, $\mathcal{L}_{\mathrm{C}}^{(i)} = \mathrm{KL}(P_i || Q_i) = \sum_{j=1}^{k} p_{ij} \log \frac{p_{ij}}{q_{ij}}$. This centroid-based approach requires the cluster centers to be initialized *once* at the beginning of the refinement stage. The initial centers are obtained by applying classical clustering algorithms on the acquired representations from the optimal pretrained parameters; *i.e.*, $\{\mathbf{z}_i = \mathrm{Enc}_{\theta^*}(\mathbf{x}_i)\}_{i=1}^{n}$.

4 EXPERIMENTS

We ground our study by evaluating on three diverse HAR benchmark datasets: UCI HAR [4]; Skoda [35]; and MHEALTH [7] employing standard train and holdout test splits (as summarized in Table 1). Datastreams are initially rescaled using per-channel normalization. After adopting the sliding window segmentation technique to partition the continuous data-streams, we consider the first 50% of sensory measurements in each segment to constitute the input sequences to our framework. Accordingly, the temporally inversed version of the input is used as the target sequence for the reconstruction task while the remaining sensory measurements are considered as the target sequence for the future prediction task.

Network Architecture. We leverage a 2-layer bi-directional GRU with 256 hidden units for the encoder. The decoders have an identical structure but utilize uni-directional connections. Considering the lower input dimension for UCI HAR compared with Skoda and

Table 2: A quantitative comparison of clustering performance on three HAR benchmark datasets in accuracy (ACC) and NMI.

	UCI HAR Dataset				Skoda Dataset				MHEALTH Dataset			
	Train Split		Test Split		Train Split		Test Split		Train Split		Test Split	
	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC
Traditional Clustering on Input Data Space												
k-means	44.28%	48.25%	42.28%	42.14%	43.41%	41.01%	46.01%	40.67%	49.71%	39.55%	48.37%	42.37%
AC-Average	1.38%	19.16%	1.93%	18.29%	4.61%	14.34%	30.98%	26.04%	4.44%	9.21%	7.55%	9.31%
AC-Complete	3.97%	19.56%	20.04%	31.69%	30.85%	27.48%	39.01%	37.47%	16.42%	11.82%	17.84%	11.15%
AC-Ward	41.07%	42.26%	48.21%	43.26%	46.55%	44.68%	46.92%	41.78%	54.06%	45.16%	56.99%	45.99%
Traditional Clustering on Autoencoding Space												
k-means	51.93%	60.19%	45.49%	55.62%	53.75%	47.56%	50.64%	42.62%	54.86%	43.96%	55.75%	48.24%
AC-Average	45.18%	37.57%	46.41%	34.61%	18.88%	16.96%	38.59%	30.22%	34.54%	20.47%	47.01%	29.26%
AC-Complete	40.66%	40.03%	40.81%	43.67%	32.55%	32.47%	41.57%	35.93%	42.23%	35.05%	44.42%	36.51%
AC-Ward	75.27%	74.78%	52.83%	60.33%	55.81%	51.51%	54.41%	45.96%	61.07%	48.91%	57.04%	46.28%
End-to-End Deep Clustering												
DEC [37]	52.85%	50.45%	53.00%	49.85%	45.32%	40.46%	47.06%	40.25%	51.86%	43.64%	52.38%	44.91%
IDEC [16]	54.86%	51.14%	50.47%	50.15%	49.54%	47.41%	47.47%	45.96%	50.89%	42.49%	53.44%	44.72%
(Ours) Deep Sensory Clustering (k-means Init.)	64.75%	64.54%	61.58%	61.28%	56.91%	50.97%	57.01%	50.28%	**62.65%**	**57.19%**	**63.06%**	**56.85%**
(Ours) Deep Sensory Clustering (Ward Init.)	**76.43%**	**78.79%**	**71.25%**	**75.41%**	**56.97%**	**52.9%**	**59.06%**	**53.48%**	59.42%	51.57%	60.91%	53.33%

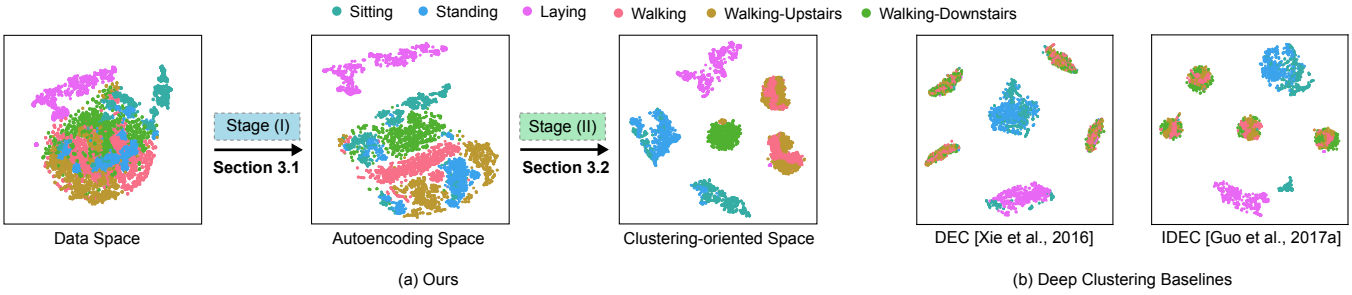

Figure 2: t-SNE visualizations of data representations for UCI HAR dataset achieved with (a) our proposed framework and, (b) deep clustering baselines. Sequence representations are color-coded with their corresponding ground-truth activity labels.

MHEALTH datasets, we impose a bottleneck embedding dimension of 64 for the former and 256 for the latter in our autoencoders. The clustering network $f_\omega(.)$ for integrating CAH uses a single layer that generates soft cluster assignments according to Eq. (3).

Optimization Settings. In mini-batches of size 256, the network parameters are updated using the ADAM optimizer with the initial learning rate set to 10^{-3} and decayed by a factor of 10 after 70 epochs. The network is pretrained for 100 epochs, and fine-tuned with the clustering objective until the cluster assignment changes between two consecutive epochs is less than 0.1%. The weighting coefficient γ is set to 0.1. All above parameters are held constant across all datasets to refrain from unrealistic parameter tuning.

4.1 Clustering

We base our evaluations for clustering on the two widely adopted metrics of *unsupervised clustering accuracy* (*ACC*) and Normalized Mutual Information (*NMI*) [28]. Our approach is compared against popular centroid-based k-means clustering [5] as well as representative hierarchical algorithms including agglomerative clustering with average linkage (AC-Average) [21], agglomerative clustering with complete linkage (AC-Complete) and Ward agglomerative clustering (AC-Ward). Further, we compare against end-to-end deep clustering methods proposed in [16, 37] for still images and show their inability to cater for the sequential nature of time-series data.

Clustering Performance. In Table 2, we evaluate the clustering performance of the traditional baselines on both the: *i*) *data space* using raw input representations; *ii*) *autoencoding space* using the embedded features $\{\mathbf{z}_i = \text{Enc}_{\theta^*}(\mathbf{x}_i)\}_{i=1}^n$ attained by optimizing $\mathcal{L}_{\text{AE}}$ in the pretraining stage; and *iii*) compare with the *end-to-end cluster assignments* generated by deep clustering baselines and our proposed *Deep Sensory Clustering*. As required by the CAH objective, we report results over two different strategies to initialize the cluster centers *only once* before commencing the refinement stage: *i*) we run k-means clustering on the embedded features to obtain k centroids; and *ii*) we perform Ward clustering and use the mean representation of the obtained clusters as initial centers.

Our results demonstrate that our end-to-end approach not only outperforms traditional clustering algorithms applied on both input data and auto-encoding spaces, but also offers a large performance margin over representative deep clustering baselines proposed for image data. Without any manual supervision, our proposed unsupervised approach can directly deliver cluster assignments with high correspondence to activities of interest in the explored datasets; we can observe accuracy (ACC) performance of 78.79%, 52.9% and 57.19%, respectively on UCI HAR, Skoda and MHEALTH datasets. In addition, the *consistent improvement* of unsupervised metrics across all three HAR datasets using our proposed framework demonstrates its generalizability to different HAR problems.

Space Visualization. In Fig. 2, we illustrate: *i*) the evolution of the feature space towards the ultimate clustering-oriented embedding space achieved with our framework; and *ii*) the deep clustering baselines by visualizing the data representation for the sequences in UCI HAR using t-SNE [26]. For our framework, we show the original dataset (*data space*), the dataset embedded by the encoder after the pretraining stage (*autoencoding space*) and the final representations after optimizing for the aggregated objective function $\mathcal{L}$ in Eq. 2 (*clustering-oriented space*) with Ward initialization. We can observe that *our framework discovers well-defined and clearly separated clusters of activity segments with strong correspondence to the ground-truth labels without manual supervision.* In contrast, the feature spaces achieved by the baseline deep clustering methods fail to correctly discover activity clusters; *e.g.* static activities of *sitting* and *standing* are recognized as a single cluster, and different *walking* variations are completely intermingled. These visualizations highlight: *i*) necessity to leverage recurrent structures within the network; and *ii*) effectiveness of incorporating self-supervised tasks when dealing with time-series data from wearables.

5 CONCLUSIONS

This study tackles the hitherto unexplored problem of *end-to-end clustering* of human actions from raw unlabelled multi-channel time-series data captured by wearables using a deep learning paradigm. To the best of knowledge, ours is the *first* to investigate and develop a novel deep clustering architecture for HAR problems with raw sensor data. Our systematic experiments demonstrate: *i*) the effectiveness; and *ii*) generalizability of our proposed approach for clustering of human activities across three diverse HAR benchmark datasets. We believe our study creates new opportunities for recognition of human activities from unlabelled raw data that can be conveniently and cheaply collected from wearables.

REFERENCES

[1] Alireza Abedin, Mahsa Ehsanpour, Qinfeng Shi, Hamid Rezatofighi, and Damith C Ranasinghe. 2020. Attend And Discriminate: Beyond the State-of-the-Art for Human Activity Recognition using Wearable Sensors. *arXiv preprint arXiv:2007.07172* (2020).

[2] Alireza Abedin, S. Hamid Rezatofighi, Qinfeng Shi, and Damith C. Ranasinghe. 2019. SparseSense: Human Activity Recognition from Highly Sparse Sensor Data-streams Using Set-based Neural Networks. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. 5780–5786. https://doi.org/10.24963/ijcai.2019/801

[3] Mohammad Abu Alsheikh, Ahmed Selim, Dusit Niyato, Linda Doyle, Shaowei Lin, and Hwee-Pink Tan. 2016. Deep activity recognition models with triaxial accelerometers. In *Workshops at the 30th AAAI Conference on Artificial Intelligence*.

[4] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge Luis Reyes-Ortiz. 2013. A public domain dataset for human activity recognition using smartphones.. In *Esann*.

[5] David Arthur and Sergei Vassilvitskii. 2007. K-means++: The Advantages of Careful Seeding. In *the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*. 1027–1035.

[6] Lu Bai, Chris Yeung, Christos Efstratiou, and Moyra Chikomo. 2019. Motion2Vector: Unsupervised Learning in Human Activity Recognition Using Wrist-Sensing Data. In *Proceedings of the ACM International Symposium on Wearable Computers*. 537âĂŞ542. https://doi.org/10.1145/3341162.3349335

[7] Oresti Banos, Rafael Garcia, Juan A Holgado-Terriza, Miguel Damas, Hector Pomares, Ignacio Rojas, Alejandro Saez, and Claudia Villalonga. 2014. mHealthDroid: a novel framework for agile development of mobile health applications. In *International workshop on ambient assisted living*. 91–98.

[8] Sourav Bhattacharya and Nicholas D Lane. 2016. Sparsification and separation of deep learning layers for constrained resource inference on wearables. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*. 176–189.

[9] Dongdong Chen, Jiancheng Lv, and Yi Zhang. 2017. Unsupervised multi-manifold clustering by learning deep representation. In *Workshops at the 31st AAAI Conference on Artificial Intelligence*.

[10] Michael Chesser, Asangi Jayatilaka, Renuka Visvanathan, Christophe Fumeaux, Alanson Sample, and Damith C. Ranasinghe. 2019. Super Low Resolution RF Powered Accelerometers for Alerting on Hospitalized Patient Bed Exits. In *IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 1–10.

[11] Jordana Dahmen, Alyssa La Fleur, Gina Sprint, Diane Cook, and Douglas L Weeks. 2017. Using wrist-worn sensors to measure and compare physical activity changes for patients undergoing rehabilitation. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. 667–672.

[12] Jordan Frank, Shie Mannor, and Doina Precup. 2010. Activity and gait recognition with time-delay embeddings. In *AAAI Conference on Artificial Intelligence*.

[13] N. Gao, W. Shao, and F. D. Salim. 2019. Predicting Personality Traits From Physical Activity Intensity. *Computer* 52, 7 (2019), 47–56.

[14] Kamran Ghasedi Dizaji, Amirhossein Herandi, Cheng Deng, Weidong Cai, and Heng Huang. 2017. Deep Clustering via Joint Convolutional Autoencoder Embedding and Relative Entropy Minimization. In *The IEEE International Conference on Computer Vision*.

[15] Yu Guan and Thomas Plötz. 2017. Ensembles of deep lstm learners for activity recognition using wearables. *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (2017), 11.

[16] Xifeng Guo, Long Gao, Xinwang Liu, and Jianping Yin. 2017. Improved deep embedded clustering with local structure preservation.. In *the 26th International Joint Conference on Artificial Intelligence*. 1753–1759.

[17] Nils Yannick Hammerla, James Fisher, Peter Andras, Lynn Rochester, Richard Walker, and Thomas Plötz. 2015. PD disease state assessment in naturalistic environments using deep learning. In *AAAI conference on artificial intelligence*.

[18] Nils Y. Hammerla, Shane Halloran, and Thomas Plötz. 2016. Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables. In *the 25th International Joint Conference on Artificial Intelligence*. 1533–1540.

[19] Harish Haresamudram, David Anderson, and Thomas Plötz. 2019. On the Role of Features in Human Activity Recognition. In *Proceedings of International Symposium on Wearable Computers*. 78–88.

[20] HM Sajjad Hossain, MD Abdullah Al Haiz Khan, and Nirmalya Roy. 2018. DeActive: scaling activity recognition with active deep learning. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 1–23.

[21] Anil K Jain, M Narasimha Murty, and Patrick J Flynn. 1999. Data clustering: a review. *ACM computing surveys* 31, 3 (1999), 264–323.

[22] Zhuxi Jiang, Yin Zheng, Huachun Tan, Bangsheng Tang, and Hanning Zhou. 2017. Variational Deep Embedding: An Unsupervised and Generative Approach to Clustering. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. 1965âĂŞ1972.

[23] Md Abdullah Al Hafiz Khan, Nirmalya Roy, and Archan Misra. 2018. Scaling human activity recognition via deep learning-based domain adaptation. In *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 1–9.

[24] Matthias Kranz, Andreas MöLler, Nils Hammerla, Stefan Diewald, Thomas PlöTz, Patrick Olivier, and Luis Roalter. 2013. The mobile fitness coach: Towards individualized skill assessment using personalized mobile devices. *Pervasive and Mobile Computing* 9, 2 (2013), 203–215.

[25] Fengfu Li, Hong Qiao, and Bo Zhang. 2018. Discriminatively boosted image clustering with fully convolutional auto-encoders. *Pattern Recognition* 83 (2018), 161–173.

[26] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9 (2008), 2579–2605.

[27] Andrea Mannini, Mary Rosenberger, William L Haskell, Angelo M Sabatini, and Stephen S Intille. 2017. Activity recognition in youth using single accelerometer placed at wrist or ankle. *Medicine and science in sports and exercise* 49, 4 (2017), 801.

[28] E. Min, X. Guo, Q. Liu, G. Zhang, J. Cui, and J. Long. 2018. A Survey of Clustering With Deep Learning: From the Perspective of Network Architecture. *IEEE Access* 6 (2018), 39501–39514.

[29] Vishvak S. Murahari and Thomas Plötz. 2018. On Attention Models for Human Activity Recognition. In *ACM International Symposium on Wearable Computers*. 100–103.

[30] Dzung Tri Nguyen, Eli Cohen, Mohammad Pourhomayoun, and Nabil Alshurafa. 2017. SwallowNet: Recurrent neural network detects and characterizes eating patterns. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 401–406.

[31] Francisco Ordóñez and Daniel Roggen. 2016. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors* 16, 1 (2016), 115.

[32] Thomas Plötz, Nils Y Hammerla, Agata Rozga, Andrea Reavis, Nathan Call, and Gregory D Abowd. 2012. Automatic assessment of problem behavior in individuals with developmental disabilities. In *Proceedings of the 2012 ACM conference on ubiquitous computing*. 391–400.

[33] Charissa Ann Ronao and Sung-Bae Cho. 2015. Deep convolutional neural networks for human activity recognition with smartphone sensors. In *International Conference on Neural Information Processing*. 46–53.

[34] Nitish Srivastava, Elman Mansimov, and Ruslan Salakhudinov. 2015. Unsupervised learning of video representations using lstms. In *International conference on machine learning*. 843–852.

[35] Thomas Stiefmeier, Daniel Roggen, Georg Ogris, Paul Lukowicz, and Gerhard Tröster. 2008. Wearable activity tracking in car manufacturing. *IEEE Pervasive Computing* 2 (2008), 42–50.

[36] Alireza Abedin Varamin, Ehsan Abbasnejad, Qinfeng Shi, Damith C Ranasinghe, and Hamid Rezatofighi. 2018. Deep Auto-Set: A Deep Auto-Encoder-Set Network for Activity Recognition Using Wearables. In *the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. 246–253.

[37] Junyuan Xie, Ross Girshick, and Ali Farhadi. 2016. Unsupervised deep embedding for clustering analysis. In *International conference on machine learning*. 478–487.

[38] Jianwei Yang, Devi Parikh, and Dhruv Batra. 2016. Joint unsupervised learning of deep representations and image clusters. In *the IEEE Conference on Computer Vision and Pattern Recognition*. 5147–5156.

[39] Jian Bo Yang, Minh Nhut Nguyen, Phyo Phyo San, Xiao Li Li, and Shonali Krishnaswamy. 2015. Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition. In *Proceedings of the 24th International Conference on Artificial Intelligence* (Buenos Aires, Argentina). 3995–4001. http://dl.acm.org/citation.cfm?id=2832747.2832806

[40] Rui Yao, Guosheng Lin, Qinfeng Shi, and Damith C. Ranasinghe. 2018. Efficient dense labelling of human activity sequences from wearables using fully convolutional networks. *Pattern Recognition* 78 (2018), 252 – 266.

[41] Ming Zeng, Le T Nguyen, Bo Yu, Ole J Mengshoel, Jiang Zhu, Pang Wu, and Joy Zhang. 2014. Convolutional neural networks for human activity recognition using mobile sensors. In *6th International Conference on Mobile Computing, Applications and Services*. 197–205.