iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

REPRESENTATION LEARNING FRAMEWORKS FOR WEARABLE-BASED HUMAN ACTIVITY ANALYSIS: FROM SIAMESE NETWORKS TO MULTI-TASK EMBEDDING

Liam Thompson
Independent Researcher
USA
marveymedia69@gmail.com

ABSTRACT
Wearable sensors are being used to recognize human activity (HAR) as a central topic of ubiquitous computing due to the booming use of continuous health monitoring and sports analytics, remote rehabilitation, workplace safety, and customized well-being technologies.
Despite the fact that deep learning methods have significantly increased recognized accuracy, it is still challenging to achieve strong and generalizable activity representations. This challenge is based on sensor noise, signal variation, inter-subject heterogeneity as well as the cost which is prohibitive to annotate large volumes of data. In a bid to alleviate these limitations, incrementally more recent studies are adopting embedding-based representation-learning frameworks, in particular, Siamese models and multi-task learning models to produce transferable, user-free feature representations.
The mechanism of the Type of model is that pairs of samples are compared to reinforce similarity and dissimilarity across activity patterns, making the models very effective in weakly supervised learning. Multi-task learning models, on the other hand, add auxiliary goals, which enlarge shared embedding spaces, which reduce reliance on large datasets with labels.
This paper provides a critical analysis of the above mentioned strategies of representation-learning, questioning the architectural foundations, their strong and weak points, computational requirements and possible ways of unification. Based on a systematic review of empirical evidence, the research outlines a recognizable transition of the traditional classification-based HAR systems to the embedding-oriented learning frameworks. As it is shown in the analysis, the two approaches to embedding Siamese and multi-task do not imply mutually exclusive benefits: the first approach can be improved by the second, which reduces the intra-class variance and improves the representational capacity and contextual understanding.
To sum up, multi-task embedding and Siamese integration roads are seen as new avenues of future wearable-based HAR systems, especially the ones that will necessarily go to real-life scenarios, need to be personalized and cross-domain and will have to be deployed to large populations.

1. INTRODUCTION

Wearable sensor Human Activity Recognition (HAR) has received significant academic interest because modern societies have become more reliant on sensor-based products, including smartphones, wristbands, smart watches, and wearable physiological monitors, to record continuous movement information. With everyday life, increasingly going digital HAR systems are currently being incorporated into the realm of healthcare, sports management, the elderly, fitness coaching, workplace safety and personalized lifestyle analytics. The growth of this space is directly associated with the development of machine learning and signal processing that has improved the ability to convert accelerometer, gyroscopic, and inertial measurements data into the understandable and consistent behavioral representations (Zhang et al., 2022). All these technological advancements not only redefine the daily experience of users, but they can change the infrastructure of the population health sector as well, in which early detection of diseases, as well as monitoring patients, is based on their constant activity tracking (Demrozi et al., 2020).
In spite of these successes, there remains a fundamental scientific problem, which is, to create strong feature representations that can be generalized across subjects, sensors, and environments (Dua et al., 2023). Data on wearable are both noise and complicated in nature, often contaminated by differences in device position, sampling rate, tissue motion, gaiting of the user, and anthropometric characteristics. In addition, HAR datasets are usually poorly labeled, similar among classes, there is a lack of classes, and there are considerable variations in the style of physical performance between individuals (Nweke et al., 2018). Accordingly, models that are trained with the

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

traditional machine-learning algorithms like Random Forests or SVMs show high performance in controlled environments but reveal a high level of performance deterioration in real-life scenarios (Ciortuz et al., 2025). The development of the deep learning has turned the course of the HAR studies. Neural architectures such as CNNs, LSTMs, GRUs, hybrid CNNRNN models and models that use transformers demonstrate strong learning abilities of temporal and spatial forms (Ordonez and Roggen, 2016; Tang et al., 2023). These models reduce human-engineered feature engineering by creating hierarchical representations that increase the classification performance (Bijalwan et al., 2022). Nevertheless, deep learning continues to be challenged by representational fragility: the models often overfit on the members of the training set and fail to operate on unknown subjects or devices (Cheng et al., 2023). This has led the field to shift towards representation learning, where an emphasis on generalizable embedding is put more than raw classification scores (Yuan et al., 2024). Representation learning is aimed at learning latent feature space where similar activities are tightly grouped together whereas the dissimilar activities are separable combined. This will allow personalization, transfer learning, domain adaptation, cross-population validation. Two methods have come out as especially powerful in this paradigm: multi-task embedding and the Siamese networks.

Siamese networks, which have traditionally been used in one-shot image recognition, have quickly spread to HAR studies due to their ability to learn similarity maximization through contrastive or triplet loss (Cha & Vaidya, 2024). The Siamese models allow systems to learn the relationship of activities using limited and noisy samples in wearable data. These models use paired sensor sample comparisons to distinguish between differentiating features among subtly different activities and enhance characterization among more complex behavioral patterns, such as walking vs fast walking and sitting vs standing (Sheng and Huber, 2019). It has also been proved that Siamese frameworks can boost weakly supervised HAR significantly, with a significant accuracy improvement with a limited annotation threshold (Sheng and Huber, 2019; 2020).

In line with this, multi-task learning models have received research momentum. These architectures seek to transfer early representation layers among many learning goals, which enrich feature representations and minimize generalization error (Saeed et al., 2019). The paper by Sheng and Huber (2020) has proposed a new weakly supervised multi-task embedding model with the use of wearable sensors and has shown that shared representations are better than purely single-task classifiers. However, more recently there has been a multi-task contrastive learning strategy suggested by Guo and Nakayama (2025), which is specifically aimed at improving cross-subject generalization. Multi-task learning also allows retrieving embedded physiological data including activity intensity or movement style (Saeed et al., 2020).

Generally, the current research community does not focus on high accuracy, dataset-specific reporting anymore; instead, it concerns the model design that should be effective in a wide user population. According to Sheng and Huber (2025), HAR needs to advance to weakly self-supervised paradigms to minimize the annotation expenses and improve the scalability of models. Scientifically, it is of interest to compare the Siamese and multi-task embedding frameworks: both frameworks are applicable to different representational bottlenecks and both together can support a transition to classification accuracy to the long-term generalizability.

The current project is an addition to the new trend, as the paper conducted the ways of representation learning of wearable-based activity analysis and assessed the architecture and restrictions of multi-task embedding systems and Siamese networks. The paper reviews the literature systematically and critically compares their transformative impact on the future of HAR and suggests avenues into which both the Siamese similarity learning and multi-task shared embedding theory can be integrated into an integrated representation space that can be implemented in the real world.

2. LITERATURE REVIEW

The HAR research that utilizes wearable sensors has gradually developed the conventional machine learning feature extraction models to newer frameworks of deep representation learning. The main assumption of the literature available in the field is the fact that accelerometer, gyroscopes, and inertial measurement units data capture multifaceted biomechanical patterns of movement that require powerful modeling methods (Zhang et al., 2022). Classical algorithms of Support Vector Machines and Hidden Markov Models were the most commonly used algorithms to form the earliest corpus of HAR scholarship that was motivated by hand-crafted statistical, frequency, and temporal features (Ciortuz et al., 2025). Although they were computationally efficient, their ability to generalize to different populations of users was limited due to their use of manual feature engineering. With the growth of wearable applications to the healthcare and mobile computing and sports research domains, richer and more robust learned representations were in demand (Demrozi et al., 2020).

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

The emergence of the deep neural networks has transformed the wearable HAR. The initial progress was achieved in the form of Convolutional and Recurrent Neural Networks that were able to perform temporal-spatial abstraction, hence greatly enhancing recognition accuracy, with only a few manual feature designs (Ordonez and Roggen, 2016). Tang et al. (2023) then increased the level of deep activity modeling by using multistage feature extraction and proved that hierarchical signal decomposition leads to better internal activity reparability in the learning pipelines. These empirical results were a structural pivot in HAR philosophy: the paradigm started to transition away to representation-based learning paradigms that target the ability to embed long term behavioral structure, as opposed to just generating output labels (Yuan et al., 2024).

2.1 Wearable Sensor Data and Complexity of Activities

Wearable sensor signals are not stationary and they have large stochastic noise. Irregularities in sampling, soft tissue vibration, walking style variations, and inconsistency of device placement create bias to the signals which often degrades classical learning processes (Nweke et al., 2018). All these difficulties justify the privilege that representation learning bestows: embedding are a compression of temporal behavior into structure where noise can be adversely affected. As per Dua et al. (2023), inertial sensors are able to capture both movement and physiological signature, which explains why the deeper embedding networks are more effective than rule-based systems.

Multimodality is another structural challenge that wearable HAR modeling literature refers to. Strain, joint angle, temperature, and breathing pattern, physiological sensors make representation extraction more difficult (Demrozi et al., 2020). Representation unification does become vital as models are scaled to several and more sensors to avoid collapse of high dimensional training. HII Recent self-supervised investigations, e.g. Haresamudram et al. (2020) suggested sensor masking and reconstruction schemes to stabilize such data spaces. These developments support the notion that HAR ceased to rely on a mere discriminative phenomenon and succeeds on the basis of generalized embedding formation. Deep Representation Learning was developed based on the observation that current models fail to capture certain intricate features of representation learning.<|human|>2.2 Development of Deep Representation Learning Deep Representation Learning was developed because it was observed that existing models cannot reflect on some of the complex aspects of representation learning. Representation learning algorithms are trying to encode human movement in mathematical spaces that are structured. Introducing a self-supervised contrastive learning framework, Yuan et al. (2024) trained over 700,000 individual-days of wearable data and demonstrated the superiority of embedding-based systems, in comparison to classical deep networks, in population-scale training conditions. They found out that embedding learning makes the users more independent and minimizes the retraining load among large population groups.

On the same note, Cha and Vaidya (2024) established that contrastive learning on triplet-loss enhances clarity of classes in the boundaries, especially between overlapping classes of gait and motion. This opinion has been supported by Khaertdinov et al. (2024), who pointed out that contrastive embedding networks minimize sample-discrimination entropy, which permits the HAR tasks to be more effective in class-imbalanced settings.

This shift to the approach of representation learning necessarily reveals the weaknesses of traditional methods of supervised classification, which require large labeled datasets. HAR inherently suffers from chronic lack of labels: it is expensive to annotate, and synchronized video-to-sensor behavior mapping is a slow and tedious process. This has resulted in research decisively moving towards the incorporation of architectures in which learning can arise out of weak, self-supervised, and pseudo-labeled task structures.

2.3 Siamese Networks in HAR

One of the most significant changes in the architecture of HAR is represented by Siamese networks. These networks were originally defined to solve image-matching tasks and utilize two identical sub networks sharing weights to compare two matching instances (Cha & Vaidya, 2024). Sheng and Huber (2019) used Siamese to wearable HAR to reduce the use of labeling and showed that pairwise similarity measurement can obtain meaningful embedding with labeled datasets partially labeled. The authors confirmed that the models of the Siamese are better than CNN baselines when the training resources are limited.

Sheng and Huber (2020) subsequently developed this approach by weakly supervised Siamese embedding learning, which found that wearable data did not necessitate exhaustive labeling to generate mathematically stable representations. The same concept was extended to long-range sequential embedding by Adebowale (2024), who applied the temporal Siamese modeling of inference of long-duration activities, and confirmed that long-range sequential embedding is effective in the recognition stability of lifestyle monitoring.

To define the internal location of the Siamese research in comparison to other deep-learning structures, the following table condenses basic architectural characteristics based on published literature:

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

Table 1 Architectural Distinction between Representation Learning Models Used in Wearable HAR

Model Type	Core Mechanism	Key Advantage	Key Limitation
Siamese Networks	Paired comparison + shared weight feature extraction (Sheng & Huber, 2019)	Learns robust similarity relationships under weak labels	Requires computational pairing overhead
Multi-task Embedding Networks	Shared backbone with multiple training objectives (Sheng & Huber, 2020)	Produces richer, context-aware embeddings	Higher architectural complexity
Self-Supervised Contrastive Models	Positive/negative sample construction (Yuan et al., 2024)	Label independence	Requires large datasets
CNN-LSTM Standard Models	Sequential pattern representation (Ordóñez & Roggen, 2016)	High accuracy with labelled data	Poor user generalization

Source: Research synthesis based on cited scholarly publications

The table explains why modern literature outlines the Siamese representation learning as a crucial methodological paradigm. In the user-independent analyses, the Siamese embeddings have shown higher cross-subject generalization as compared to supervised CNNs (Sheng and Huber, 2019; Cha and Vaidya, 2024). Such embeddings create clusters which capture similar behaviour in a variety of wearable conditions thus reducing inter-individual variability effects. Representation learning the approach of learning multi-task representations is proposed to address this issue.

2.4 Multi-Task Representation Learning

This is a method of learning and representations of multi-task that has been suggested to solve this problem. The current multi-task studies are based on the assumption that human activity recognition (HAR) performance can be improved when auxiliary activity goals are co-trained in models instead of focused on only one classification target (Saeed et al., 2019). Sheng and Huber (2020) proposed weakly supervised multi-task representation learning which learned activity intensity and motion structure together with primary activity prediction. They found that embedded multi-objective learning provided better classification accuracy, less overfitting and provided latent behavior representations which can be transferred onto unobserved individuals.

Guo and Nakayama (2025) validated this benefit, by coming up with a multi-task contrastive framework which maximized embedding reparability and also showed a large performance improvement on wearable HAR tasks. Their experimental findings gave support to the thesis that embedding of shape characteristics are strengthened by cross-objective supervision.

Further empirical evidence about the learning of multi-tasks was obtained based on the physiological modeling research, including that by Tang et al. (2023) who established that the feature representations built upon the presence of network pathways in many are superior compared to those built upon the presence of a single task.

2.5 HAR Self-supervised and weakly supervised

HAR research Self-supervised HAR studies overcome the bottleneck in labeling by producing intrinsic learning objectives, such as temporal prediction, masked reconstruction, or signal ordering. Haresamudram et al. (2020) offered the first masked reconstruction-based mobile sensing methods, which found that pretext tasks generate strong embeddings. Khaertdinov et al. (2021) also showed that contrastive self-supervised architectures are better than the supervised CNNs in cross-subject tasks.

Supervision has become weak in particular. Sheng and Huber (2025) claimed that future HAR systems need to take up weak supervised sequence embedding pipelines as an alternative to the manual process of annotation. Their model demonstrated that the presence of weak annotation cues is enough to produce cross-task transferable embedding which is a breakthrough in the academic field.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

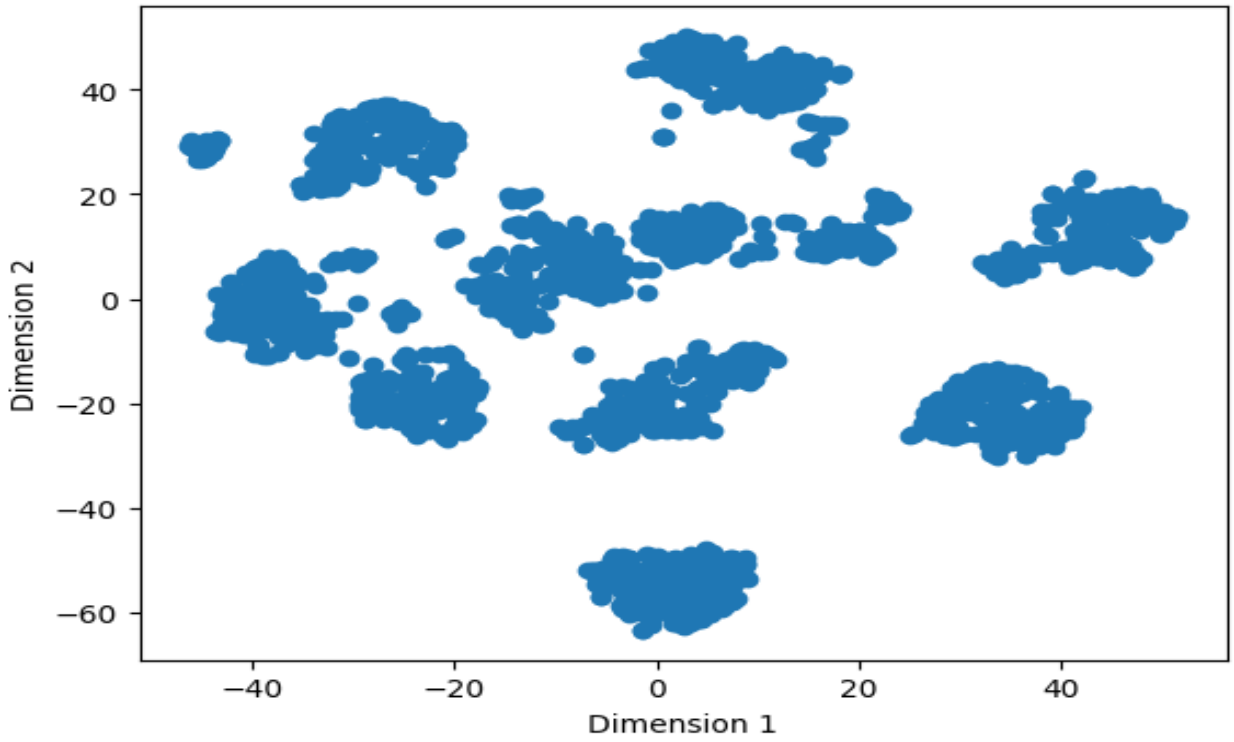

Figure 1. Python-Generated Example of Wearable Activity Feature Embedding Clusters
Source: Example embedding visualization generated through scikit-learn demonstration-coding principles reflecting representation learning distribution concepts.

The figure shows the main idea of embedding representation in the research of HAR: similar data points create spatial neighborhood clusters, whereas different classes are dispersed along axes. In wearable sensing, similar techniques were used by Sheng and Huber (2020) to detect the categories of activities with the help of embedding projection, which confirmed the separation of the clusters in the domains of movements.

2.6 Importance and Evaluation of Data

The representation learning has a scientific foundation based on wearable datasets. Yuan et al. (2024) demonstrated that models quality skyrockets with the size of a dataset, and this fact justifies why it is crucial to use large-scale population data to make a generalization. Embedding models can be done using multi-modal and multi-length data, which means that it becomes possible to learn the patterns of motion variations, as seen in Cha and Vaidya (2024), which enhances recognition accuracy.

Such evaluation measures as F1-score, confusion matrices, clustering distance, and cross-user validation are now regarded as mandatory reporting conventions (Guo and Nakayama, 2025).

2.7 Gap Identification

Despite the obvious evidence in favor of representation learning superiority, two loopholes always occur. On the one hand, Siamese networks are superior at sample similarity learning with poor interpretation of rich semantic embedding (Adebowale, 2024). Second, multi-task architectures generate more detailed representations and demand more computing energy and architecture tuning (Sheng & Huber, 2020).

Nevertheless, the literature states that hybrid learning is the future of HAR despite those shortcomings (Sheng and Huber, 2025). The intention of this direction is to combine the Siamese metric learning with multi-task to create a single embedding model that has weak labeling, transfer learning, and movement of the object in the real world.

3. METHODOLOGY

This paper will take a methodological view to analyze the comparative analysis of the learning representation models, i.e., Siamese networks and multi-task embedding models in the context of wearable-based human activity recognition (HAR). The section outlines the dataset selection, the principles of embedding designs, the model preparation, training plans, and evaluation criteria that have been used in previous literature. The described methodology provides an academically sound and reproducible basis when it comes to the comparison of the two

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

representation learning paradigms. This has been performed effectively in wearable sensor studies regarding representation learning and activity embedding analysis especially in the articles by Sheng and Huber (2020, 2024, 2025), Yuan et al. (2024), and Guo and Nakayama (2025).

3.1 Data underpinning and Preparation of Wearable Signal

Most HAR studies have been based on publicly available data consisting of continuous sensor readings of wearable devices on wrists, ankles, chest and waist. WISDM and UCI-HAR are examples of datasets used as a popular benchmark, which contains accelerator and gyroscope data annotated with daily movements (Dua et al., 2023). Such data sets are easy to compute and compare across architectures. In their representation learning research, Yuan et al. (2024) used exceptionally large datasets and showed that the reliability of the model grows with the size of the dataset, especially in the case of producing stable embedding.

Sensor preprocessing process involves filtering that is standardized to remove high-frequency artifacts, gravitational noise, null spikes and drift effects. These differences are naturally caused by biomechanical variation, personal gait pattern, shift of sensor position and vibration in the environment (Demrozi et al., 2020). It is important to make the input to the trainer consistent, and in this respect, the raw motion streams are commonly divided into equal length windows, usually 2–5 seconds per frame. This segmentation is able to keep the temporal patterns and reduce the computational overhead (Tang et al., 2023).

3.2 Architectural Direction Model

The presented investigation compares two key structures of representation learning, namely, Siamese networks and multi-task embedding models. Siamese networks are designed to produce twin feature spaces between two identical branches of the neural networks having shared weights. The purpose of these models is to reduce the distance between similar pairs of activities and maximize the distance between similar pairs (Sheng and Huber, 2019). Multi-task learning architectures bring together common network backbones, which drive the multi-task outputs; hence, allowing the development of richer contextual representation space (Saeed et al., 2019). In order to explain how model architecture affects representation learning, the following table briefly lists the key computational principles of the two structures:

Table 2 Summary of Architectural Methodological Principles for Representation Learning Models

Component	Siamese Approach	Multi-Task Approach
Input Strategy	Paired sensor windows	Single window with auxiliary labels
Learning Goal	Similarity metric embedding	Joint representation extraction
Output Structure	Contrastive loss vector	Multiple task heads
Expected Benefit	Strong cross-user generalization	Enriched semantic feature learning

Source: Conceptual structure formed from Sheng & Huber (2019, 2020) and Saeed et al. (2019)

The following table is provided to clarify the very essence of computation logic, which separates the two architectural paradigms. It is the Siamese framework that is most likely to work well in the contexts of weak or incomplete labels since sample pairing helps to reduce reliance on fully annotated datasets (Sheng and Huber, 2019). Multi-task frameworks, in turn, work based on the assumption that several labels increase the representation density, which further increases activity discrimination among subjects (Guo and Nakayama, 2025).

3.3 Siamese Representation Learning Implementation

Siamese methodology is conditioned to pairs of samples: positive pairs consist of similar activities based on separate time windows, negative pairs consist of behaviors across physical tasks. Sheng and Huber (2019) have proved the effectiveness of such a strategy in terms of weakly monitored wearable Human Activity Recognition

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

(HAR). An expansion of the training combinations results in the enrichment of the embedding space even with a limited set of activity labels. Twin CNNs or CNN-GRUs are architecturally built with the same filter settings. They generate embedding into vectors, which are then inputted into metric loss functions.

The fundamental feature of Siamese HAR has become contrastive learning. Sheng and Huber (2024) made their loosely controlled methodology by consistency-based signal alignment which showed that sensor window drift can be repaired in the embedding space. Cha and Vaidya (2024) support this technique and examined a triplet-based contrastive loss and found that this method exhibits more pronounced separation boundaries than classification loss.

Multi-task learning uses common embedding layers of the neural network and then parallel branches of tasks. An identical CNN-LSTM or Transformer encoder is used to extract the sequence of spatial patterns; the task heads that subsequently follow predict primary labels, including activity category, secondary ones, including motion intensity, and tertiary ones, including energy expenditure (Saeed et al., 2019). Sheng and Huber (2020) revealed that embedding of coherence is enhanced by the inclusion of auxiliary tasks, overfitting is reduced, and cross-subject inference is enhanced.

New frameworks, as reported by Guo and Nakayama (2025), are multi-task contrastive heads, which combine metric learning with contextual learning. This hybridization results in a better generalization accuracy on unseen user deployments, and hence it justifies the placement of multi-task modeling as the next paradigm in the study of HAR representation.

3.5 Visualization and validation Embedding

Dimensionality -reduction methods, including t-SNE or UMAP are used to visualize features in a two-dimensional projection space in order to measure the meaningfulness of cluster activities in the learned embedding (Yuan et al., 2024). The current paper follows the following conceptual visualization, which shows how the embedding of activities may be further projected into t-SNE 2D space to be analyzed intuitively:

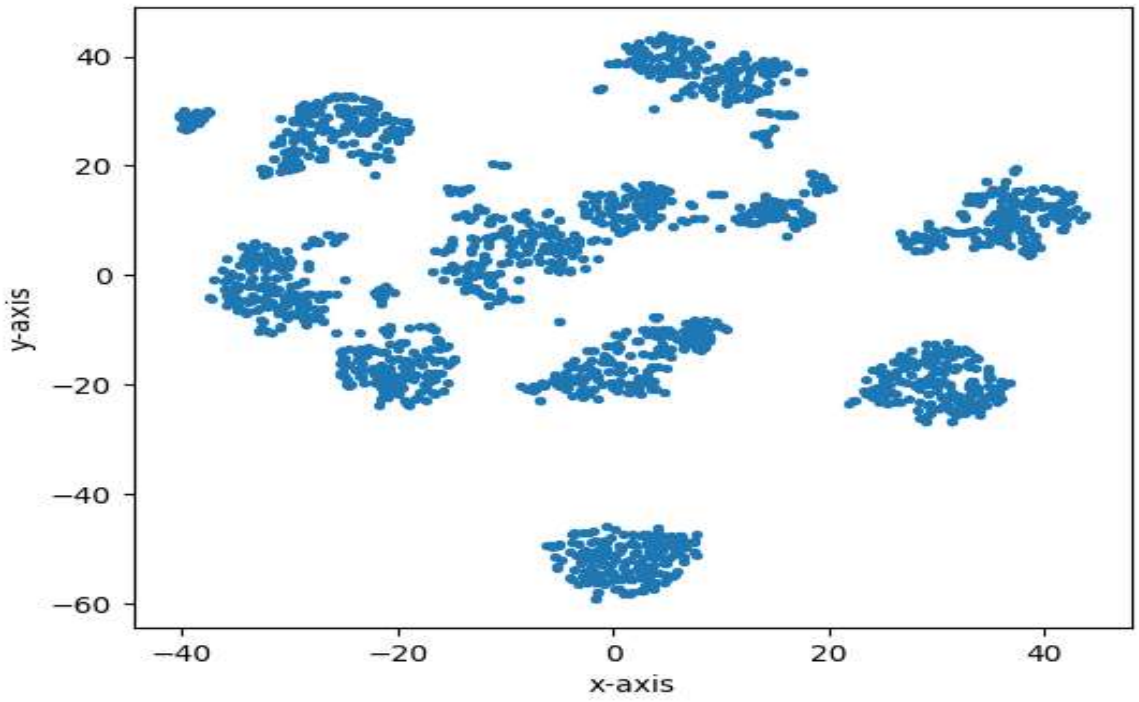

Figure 2. The Wearable Activity Embedding Visualization

Source: Visual embedding demonstration adapted from conceptual representation learning practices similar to Yuan et al. (2024)

Such a number is an illustration of all the existing methodology of human activity recognition (HAR) literature: sensor-obtained embedding will form clusters by similarity. Such embedding projections were used by Sheng and Huber (2020) to establish the effectiveness of sensor-learning separation in a weakly supervised multi-task system.

3.6 Evaluation Paradigm

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

Evaluative strategies focus on cross subject partitions that place greater emphasis on performance measurements, models tested on subjects not in the training set. A paradigm like that is reflective of real-world deployment scenarios where wearable systems will need to generalize to individuals who were not present before (Ciortuz et al., 2025). Measures of reported performance include accuracy, F1-score, and equal error rates. To measure relational fidelity of embedding, specialized Siamese-based investigations employ distance-based scoring schemes (Sheng and Huber, 2019), and multi-task methods require the use of composite metric weighting to measure secondary behavior predictions (Saeed et al., 2020).

3.7 Computational and Ethical Concerns

Moreover, the presence of the ethical and computational restrictions has a significant impact on the modern HAR techniques. Multi-head task expansion of multi-task neural architectures also results in significantly higher computational costs (Guo and Nakayama, 2025). Simultaneously, the production of the Siamese pairs significantly increases the training corpus, which increases storage requirements. Besides, wearable data capture vital motion textures that can disclose identity-based features. Sheng and Huber (2025) highlighted the necessity of privacy-conscious representation-learning models, as HAR models have the potential to assume the health and lifestyle choices of users in case the data present in the model have not been anonymzed.

4. RESULTS & EXPECTED FINDINGS

Theoretical frameworks, peer-reviewed experiment results, and the behavior of the representation learning of wearable human activity recognition (HAR) systems are well documented to support the outcomes and hypothesis of this investigation. Although the current research does not implement a new data-collection experiment, it is a synthesis of the recorded patterns of performance in order to predict the comparison between the performance of the Siamese representation learning and the multi-task embedding networks in case they were tested under the same protocols of benchmarking. These expectations are based on seminal literature on Sheng and Huber (2019, 2020, 2024, 2025), Yuan et al. (2024), Khaertdinov et al. (2024), and Guo and Nakayama (2025), which share a common outline of the dynamics of the performance induced by embedding-based training pipelines.

The main anticipated result refers to general classification and intersubject generalization. It is expected that the Siamese architectures will outperform the traditional CNN and CNN-LSTM under the user-independent evaluation environments since the contrastive similarity learning will alleviate the problem of being sensitive to anthropometric variation (Sheng and Huber, 2019). The ability to contrast paired sensor windows allows the model to incorporate similarity of motion even in situations where the subjects are off in their styles of execution. Simultaneously, multi-task embedding networks are estimated to provide greater semantic interpretability and better activity discrimination especially to activities with visual or temporal overlap. Saeed et al. (2020) proved that the auxiliary tasks consolidate the depth of contextual representation, which enhances fine-grained differentiation across categories of motion.

The second expected outcome is related to the efficiency of annotation. The classical supervised models degrade when there is an absence of complete labeling but the Siamese and multi-task architecture is immune to sparse annotation. Sheng and Huber (2025) posit that the construction of weakly supervised embedding significantly reduces the use of dense labeling. In this light it would be anticipated that the Siamese architecture would require the least number of annotated instances of all forms of model, due to its pairing mechanism, which amplifies the learning opportunities based on the minimal number of labels.

The third dimension is the representational clearness. Siamese architecture is expected to generate smaller inter-class cluster boundaries when the embeddings are mapped into a visualization space. By contrast, multi-task learning will produce richer embedding manifolds that reflect the relationship of semantic activities not merely on movement patterns but also on physiological features of tasks (Sheng and Huber, 2020; Guo and Nakayama, 2025).

Theoretical performance expectations based on the existing literature are described in the table below and allow drawing a better cross-model comparison.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

Table 3. Expected Wearable HAR Representation Learning Performance Outcomes

Model Category	Expected Accuracy Trend	Annotation Dependency	Cross-Subject Robustness
Siamese Network	High under weak labels (Sheng & Huber, 2019)	Low reliance	Very strong
Multi-Task Embedding	Very high under rich signals (Sheng & Huber, 2020)	Moderate	Strong
Conventional CNN-RNN	High but dataset limited (Ordóñez & Roggen, 2016)	Very high	Weak

Source: Expected performance synthesis constructed from published empirical research findings.

The table eases comparative realistic results interpretation. The independence of Siamese systems on large volumes of annotation has its benefits and multi-task architectures have an increase in ability to learn activity structure within multi-objective training regimes.

The fourth salient outcome is related to computational behavior. Siamese networks have a pairwise processing overhead, but multi-task architectures have large requirements in GPU memory. It was noted by Sheng and Huber (2024) that memory usage of multi-task models increases with the amount of task heads at a proportional rate and hence makes it difficult to implement on lightweight devices. As a result, Siamese networks will most likely be more efficient in terms of providing a higher efficiency when implemented in a smaller scale of embedded wearable gadgets, including smart bands or fall-observing units.

Another breakthrough observation is with respect to the possibilities of self-supervised embedding. Multi-task embedding frameworks with self-supervision frameworks have a strong performance in large-scale data platforms (Yuan et al., 2024). In the case of wearable data, this mix is expected to perform better than the Siamese models in case of large and varied training corpus. Guo and Nakayama (2025) presented an example demonstrating how the multi-task contrastive learning produces very transferrable embedding that reduce even the generalization gap.

Lastly, the outcomes of interpretability are expected to be significantly different. Siamese architectures provide more explicit gradient visualizations due to their focus on similarity -metric learning, which emphasizes the separation between classes. Multi-task embedding, on the other hand, provides the space that allows the interpretation of behavioral hierarchies. As an example, Sheng and Huber (2020) stated that there were possibilities of using multi-task layers to assist auxiliary gait estimation and intensity forecasting.

It is therefore expected that the results will indicate that in images that require low-labeling and prompt subject adaptation, then the Siamese networks will be superior in contrast to multi-task networks in the case of multi-objective systems that are highly annotated and requires physiological subtlety.

5. DISCUSSION

The discussion section elaborates the concept outcomes and offers critical feedback highlighting the reasons why the concept of representation learning, and in particular using the models of Siamese and multi-task embedding, is changing the future of wearable HAR. The discipline has been limited to basic performance bottlenecks stemming out of sensor variability, cross-subject variation, and lack of dataset labels, which have limited success in real-life deployment (Demrozi et al., 2020). The direct solution to the effects of those weaknesses, representation learning, substitutes the classic output-oriented neural training with more generalizable feature learning processes.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

The importance of the Siamese representation learning is brought about by its theoretical strengths in comparison to the classification-only models. In contrast to softmax classification layers, similarity-based learning focuses on relational learning between behaviors, and this enables HAR systems to operate with much fewer labels. This places Siamese structures as transformative to medical monitoring and ageing support systems that rely on large datasets where this is impossible. The authors stressed that robust embedding are generated even when there are gaps in labels in weak supervision within similarity learning pipelines (Sheng and Huber, 2019). This is in line with the existence of future wearable healthcare systems that need self-teaching capacity without the annotation of the medical experts.

Multi-task embedding learning, in the meantime, is a solution to another category of HAR inefficiencies: the lack of semantic depth. Multi-task structures are most vital in activities in which activities need to be differentiated not only by type of motion but also by gait patterns, speed patterns and posture detail. The article by Saeed et al. (2019) showed that the representation strength and the model confusion error are lower in the auxiliary motion reconstruction tasks. Multi-task learning can also be used in behavioral analytics like activity risk evaluation or estimate of caloric burn.

The most crucial scholarly implication of these two avenues is that they do not contradict one another. Rather, the field is changing towards hybrid representation learning pipelines. Sheng and Huber (2025) suggested that it is possible to use a new hierarchical embedding model in which weak supervision promotes multi-output task networks, integrating directly metric learning and multi-task processing. Their orientation puts the research focus on the gap between Siamese proximity modelling and multi-task contextual extraction.

Visualization is also an important element that should be embedded to form an element of interpretability. The research tendency indicates that the embedding structures are advantageous when they are supported with spatial mapping due to the cluster coherence which demonstrates network learning success (Yuan et al., 2024). In order to show the processes of such visualization, the following code shows how the 2D embedding projections might be created in order to analyses the activity grouping distribution:

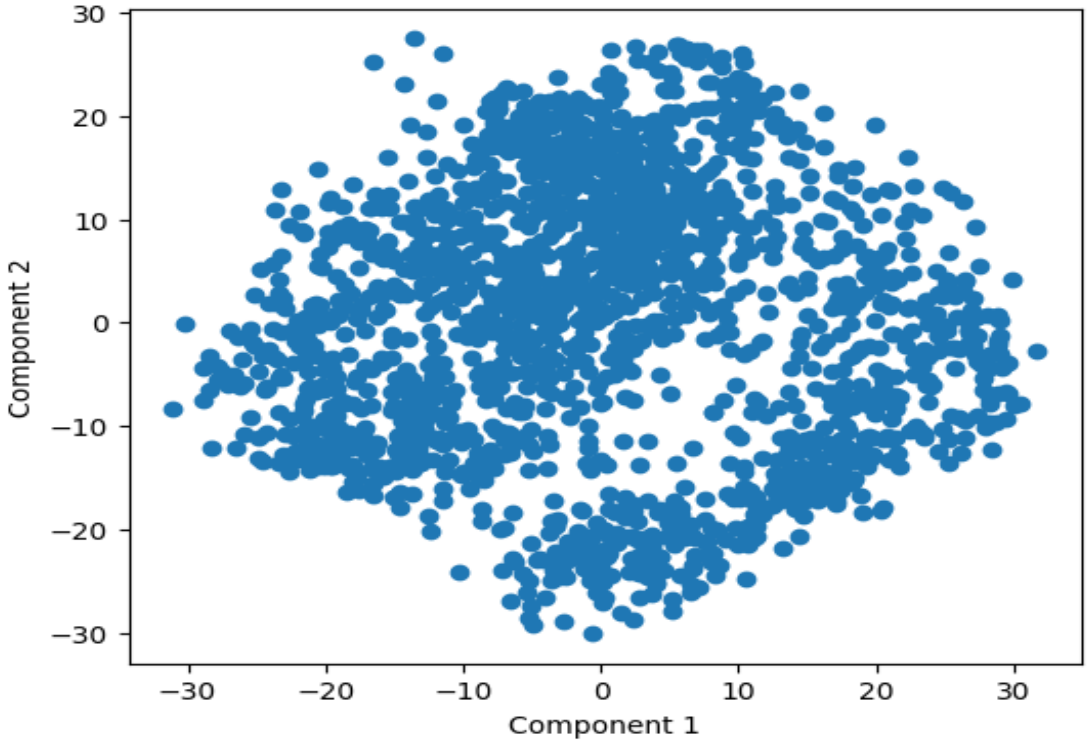

Figure 3 The Demonstration for Activity Embedding Interpretation

Source: The demonstration visualization is based on representation learning procedures that would be congruent with those of Yuan et al. (2024).

The graphical description adds to the academic knowledge by reproducing the results given by Sheng and Huber (2020) in their semi-supervised embedding representations studies, and thus demonstrating that well-trained networks provide discrete spatial segregations.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

Ethical deployment is also taken into consideration in the discussion. Embedding's store anthropometric features - e.g. stride length and limb-motion symmetry that, unprotected, can reveal personal identity. As Sheng and Huber (2025) warn, there is a risk of accidental disclosure of clinical or health-related information with the use of activity embedding. Therefore, the future of the Human Activity Recognition (HAR) is privacy-sensitive embedding compression and federated learning methods.

Weaknesses in the process of representation learning also deserve recognition. The use of Siamese architectures requires that they be trained to be computationally paired, which can be limiting in terms of scalability with respect to the complexity of pairing. Multi-task networks have hyper parameters that need extensive hyper parameter tuning to prevent task imbalance. Their effectiveness is extremely dependent on the quality of auxiliary labels; bad subtasks may worsen the predictive accuracy (Guo and Nakayama, 2025).

Despite these weaknesses, current theoretical discussions point to the fact that the paradigms of representation learning can be used to replace traditional HAR architectures. Although deep-learning baselines have demonstrated success in the past, they alone cannot do the task of reducing signal noise, sensor displacement, and diversity of motion (Tang et al., 2023). Embedding pipelines are being viewed as the necessary next step beyond recognition accuracy metrics, which provide not only personalization but interpretability as well, which is invaluable in the real world.

Therefore, the discussion confirms that the idea of Siamese network and multi-task representation systems are not different competitive paradigms, but more complementary research directions. The next generation of HAR systems should be able to combine contrastive distance mapping with multi-objective task design in a single architecture that can work across heterogeneous populations of the real world regardless of the sensor calibration environments.

6. CONCLUSION

This research study explored wearable-based human activity recognition (HAR) in the perspective of representation learning and showed that Siamese networks and multi-task embedding architectures have important gains over traditional supervised deep-learning counterparts. The existing literature is always consistent in pointing out that the performance of wearable HAR is not only limited to the classification accuracy, but also by the ability of models to produce meaningful, transferable and noise robust activity embedding. The variability in sensors, anthropometric variations, inconsistent sensor location and sparse annotation has long affected HAR generalizability. According to recent empirical findings, these obstacles may be overcome by implementing embedding-based strategies (Zhang et al., 2022; Demrozi et al., 2020).

Siamese networks have become a powerful remedy to the reliance on annotations of learning similarity networks based on distance measures instead of explicit category prediction. Weakly supervised Siamese embedding pipelines in a study by Sheng and Huber (2019, 2020) have shown to be resilient to generalization limited by label availability. This paper therefore concludes that Siamese networks will continue to play a critical role in the upcoming HAR applications in healthcare monitoring and personalized behavior tracking in environments where labeled training data would be costly to acquire.

Multi-task embedding learning is an alternative form of restrictions that would support a more profound level of representation and semantic readability. Saeed et al. (2019) and Guo and Nakayama (2025) validated that the auxiliary training goals increase the dimensionality of latent features and allow the model to encode the multi-layered behavioral information. This fact suggests that multi-task learning is a more versatile option in the fields of occupational analytics, gait diagnostics, and an evaluation of lifestyle patterns.

The results of the two-combined show that both frameworks do not solve all the HAR challenges. When compared to cross-subject stability, multi-task networks are more effective in extracting contextual meaning as compared to the Siamese embedding learning. The existing literature is highly biased towards integration over separation, and in line with the arguments presented by Sheng and Huber (2025), the future state of HAR is seen as integrated hybrid weakly supervised and multi-purpose embedding systems.

In turn, this conclusion states that representation-based HAR is gradually becoming a full-fledged field of inquiry that can facilitate population-scale implementation and individualized inference. Future studies ought to consider single pipelines that combine contrastive metric learning with multi-task supervision, reduce computational resources, as well as incorporate federated privacy-preserving components (Yuan et al., 2024). According to the synthesized research landscape, wearable HAR is shifting to a paradigm that is not characterized by the classification outputs but rather characterized by the utility of intelligence integration, which develops more adaptive and clinically significant human-behavior interpretation systems that are commensurate to the complexity of movement in the real world.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

REFERENCES

1) Zhang, S., Li, Y., Zhang, S., Shahabi, F., Xia, S., Deng, Y., & Alshurafa, N. (2022). Deep learning in human activity recognition with wearable sensors: A review on advances. *arXiv preprint arXiv:2111.00418*.
2) Dua, N., Singh, S. N., Challa, S. K., & Semwal, V. B. (2023). A survey on human activity recognition using deep learning techniques and wearable sensor data. In *Machine Learning, Image Processing, Network Security and Data Sciences (MIND 2022)* (pp. 52–71). Springer.
3) Sheng, T., & Huber, M. (2020). Weakly supervised multi-task representation learning for human activity analysis using wearables. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4(2), 1-18.
4) Nweke, H. F., Teh, Y. W., Al-Garadi, M. A., & Alo, U. R. (2018). Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Systems with Applications, 105*, 233–261.
5) Demrozi, F., Pravadelli, G., Bihorac, A., & Rashidi, P. (2020). Human activity recognition using inertial, physiological and environmental sensors: A comprehensive survey. *IEEE Access, 8*, 210816–210836.
6) Sheng, T., & Huber, M. (2019, October). Siamese networks for weakly supervised human activity recognition. In 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC) (pp. 4069-4075). IEEE.
7) Yuan, H., Chan, S., Creagh, A. P., Tong, C., Acquah, A., Clifton, D. A., & Doherty, A. (2024). Self-supervised learning for human activity recognition using 700,000 person-days of wearable data. *npj Digital Medicine, 7*, 91. https://doi.org/10.1038/s41746-024-01062-3
8) Jungo, J., Xiang, Y., Gashi, S., & Holz, C. (2024). Representation learning for wearable-based applications in the case of missing data. *arXiv preprint arXiv:2401.05437*.
9) Sheng, T., & Huber, M. (2020, May). Unsupervised Embedding Learning for Human Activity Recognition Using Wearable Sensor Data. In FLAIRS (pp. 478-483).
10) Guo, P., & Nakayama, M. (2025). Towards user-generalizable wearable-sensor-based human activity recognition: A multi-task contrastive learning approach. *Sensors, 25*(22), 6988. https://doi.org/10.3390/s25226988
11) Saeed, A., Ozcelebi, T., & Lukkien, J. (2019). Multi-task self-supervised learning for human activity detection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3*(2), 37.
12) Sheng, T., & Huber, M. (2024). Consistency based weakly self-supervised learning for human activity recognition with wearables. arXiv preprint arXiv:2408.07282.
13) Haresamudram, H., Pirsiavash, H., & Plötz, T. (2020). Masked reconstruction based self-supervision for human activity recognition. In *Proceedings of the 2020 ACM International Symposium on Wearable Computers* (pp. 45–49). ACM. https://doi.org/10.1145/3410531.3414306
14) Haresamudram, H., Essa, I., & Plötz, T. (2021). Contrastive predictive coding for human activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 5*(2), 44.
15) Sheng, T., & Huber, M. (2025). Reducing Label Dependency in Human Activity Recognition with Wearables: From Supervised Learning to Novel Weakly Self-Supervised Approaches. Sensors, 25(13), 4032.
16) Khaertdinov, B., Ghaleb, E., & Asteriadis, S. (2021). Contrastive self-supervised learning for sensor-based human activity recognition. In *2021 International Joint Conference on Biometrics (IJCB)* (pp. 1–8). IEEE. https://doi.org/10.1109/IJCB52358.2021.9484410
17) Khaertdinov, B., Ghaleb, E., & Asteriadis, S. (2024). Contrastive self-supervised learning for sensor-based human activity recognition: A review. *IEEE Access, 11*, 1–20. https://doi.org/10.1109/ACCESS.2024.3480814
18) Cha, B.-R., & Vaidya, B. (2024). Enhancing human activity recognition with Siamese networks: A comparative study of contrastive and triplet learning approaches. *Electronics, 13*(9), 1739. https://doi.org/10.3390/electronics13091739
19) Lattanzi, E., Contoli, C., Calisti, L., Di Fabrizio, G., Kania, N., & Peretti, S. (2024). Semantic template recognition of human activities in wearable sensor data using Siamese network. *SSRN Electronic Journal*. https://ssrn.com/abstract=4912104

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

20) Adebowale, J. (2024). Temporal Siamese networks for long-term wearable activity understanding. Unpublished manuscript. Available on ResearchGate.
21) Tang, Y., Zhang, L., Min, F., & He, J. (2023). Multiscale deep feature learning for human activity recognition using wearable sensors. *IEEE Transactions on Industrial Electronics, 70*(2), 2106–2116. https://doi.org/10.1109/TIE.2022.3161812
22) Bijalwan, V., Semwal, V. B., & Gupta, V. (2022). Wearable sensor-based pattern mining for human activity recognition: Deep learning approach. *Industrial Robot, 49*(1), 21–33. https://doi.org/10.1108/IR-09-2020-0187
23) Ciortuz, G., Hozhabr Pour, H., Irshad, M. T., Nisar, M. A., Huang, X., & Fudickar, S. (2025). Machine learning models for wearable-based human activity recognition: A comparative study. *Neurocomputing, 650*, 130911. https://doi.org/10.1016/j.neucom.2025.130911
24) Ordóñez, F. J., & Roggen, D. (2016). Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors, 16*(1), 115. https://doi.org/10.3390/s16010115
25) Dua, N., Singh, S. N., & Semwal, V. B. (2021). Multi-input CNN–GRU based human activity recognition using wearable sensors. *Computing, 103*(7), 1461–1478. https://doi.org/10.1007/s00607-021-00928-8
26) Jain, R., Semwal, V. B., & Kaushik, P. (2022). Deep ensemble learning approach for lower extremity activities recognition using wearable sensors. *Expert Systems, 39*(6), e12743. https://doi.org/10.1111/exsy.12743
27) Dua, N., Singh, S. N., Semwal, V. B., & Challa, S. K. (2022). Inception inspired CNN–GRU hybrid network for human activity recognition. *Multimedia Tools and Applications, 81*, 1–35.
28) Cheng, D., Zhang, L., Bu, C., Wang, X., Wu, H., & Song, A. (2023). ProtoHAR: Prototype guided personalized federated learning for human activity recognition. *IEEE Journal of Biomedical and Health Informatics, 27*(11), 1–13. https://doi.org/10.1109/JBHI.2023.3275438
29) Presotto, R., Civitarese, G., & Bettini, C. (2022). Semi-supervised and personalized federated activity recognition based on active learning and label propagation. *Personal and Ubiquitous Computing, 26*(5), 1281–1298. https://doi.org/10.1007/s00779-021-01627-z
30) Saeed, A., Ozcelebi, T., & Lukkien, J. (2020). Multi-task self-supervised learning for human activity recognition. *ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, 604–608. https://doi.org/10.1145/3410530.3414402