**Journal of Multidisciplinary Research and Technology**
**Volume 2 No 1 | January 2026**
e-ISSN: 3089-3372; p-ISSN: 3089-3380, Hal 98-115
DOI: https://doi.org/10.63891/j-mart.v2i1.132
*Available online at*: https://journal.iset.or.id/index.php/J-MART/article/view/132

# Contrastive Weakly Supervised Representation Learning for Robust and Annotation-Efficient Human Activity Recognition with Wearable Sensors

**Phillip Thomas**
Independent Researcher, United Kingdom

***Abstract***. *Wearable sensors and Human Activity Recognition (HAR) have turned into the technology of the foundation of healthcare monitoring and smart living environments, analytics in sports, and systems of rehabilitation. Although there are notable improvements in deep learning architecture time-series modeling, the overwhelming majority of the state-of-the-art HAR systems depend on the presence of fully annotated datasets, which are expensive, require human labor, and are not always available on large scale. Moreover, wearable sensor data is prone to noise, subject variability, domain shifts, and inconsistent labelling and thus, in real-world applications, wearable sensor data results in worse generalization performance. The current paper suggests a weakly supervised paradigm of representation learning that is contrastive, and it is expected to be more robust and much less reliant on annotation. By combining contrastive self-supervised tasks with weakly labeled input, the proposed method will learn both discriminative and invariant time-series representations using multivariate time-series signals produced by wearable devices. It has a framework that includes data augmentation policies, consistency regularization, pseudo-label refinement, and noise-sensitive training systems to reduce the effects of small and low-quality annotations. The experimental studies of the proposed method in different label scarcity and simulated noise levels show that the proposed approach has better performance than fully supervised and purely self-supervised baselines. The findings demonstrate significant improvements in the quality of representations, robustness and efficiency of annotations, and the way to scalable and economical systems of HAR. This paper can be used in the further development of annotation-efficient deep learning techniques in sensor-based smart systems.*
***Keywords*** *Human activity recognition, weakly supervised learning, contrastive learning, wearable sensors, self-supervised representation learning, annotation efficiency, robust deep learning*

## INTRODUCTION

Human Activity Recognition (HAR) on wearable sensors data has become a revolutionary technology in the fields of healthcare monitoring, fitness tracking, geriatric care, and smart environment. Multivariate time-series data generated by accelerometers, gyroscopes, and other inertial measurement units included in wearable devices can be used to infer physical activities that can be walking, sitting, running or other complex behavioral patterns.

The deep learning approaches have been instrumental in enhancing the performance of HARs through the automatic derivation of hierarchical time-related characteristics. Nevertheless, the majority of the well-performing systems require large-sized and complete labeled datasets. In reality, the cost and time of acquiring accurate

*Submitted: January 23, 2026; Revised: January 27, 2026; Accepted: January 29, 2026*
*Corresponding author, PhillipThomas903549@gmail.com

annotations of activities is high, time-consuming and unpredictable because of the error in human labeling and the context ambiguities. Label dependency as it is illustrated in Sheng and Huber (2020), Sheng and Huber (2019), and Sheng and Huber (2025) is a key bottleneck in the implementation of HAR systems that are wearable-based at scale.

Weakly supervised learning provides a strong avenue to provide a chance to train models using coarse, incomplete, or noisy labels. Recent wearable-based HAR works have found weak supervision strategy to lessen annotation needs (Sheng & Huber, 2020a; Azizi et al., 2022; Sheng & Huber, 2020b). Nonetheless, weak supervision might not ensure sound feature learning when there is a substantial domain variation.

A subfield of self-supervised learning, contrastive representation learning, has been shown to be incredibly successful in learning invariant representations without human labels (Lin et al., 2025; Sheng & Huber, 2022). Contrastive approaches can extract intrinsic data structures by under maximizing similarity within and maximizing similarity between augmented reference frames of the same instance and by minimizing similarity among reference frames of different instances. The combination of contrastive goals and weak label cues has not been studied in wearable HAR.

In this paper, a single contrastive weakly supervised model is proposed which minimizes the annotation dependency and maximizes the resistance to noise and domain shifts. Its key contributions are as follows:

1. A new combination of self-supervised contrastive objectives and weak guidance of wearable HAR.
2. An enhanced training pipeline with robustness and consistency regularization as well as noise-aware learning.
3. General testing in conditions of label scarcity and noise simulation.
4. Shown enhancements in annotation efficiency with no loss of predictive performance.

## A. Background and Related Work

This segment will be a review of available literature pertaining to wearables-based Human Activity Recognition (HAR), weakly supervised learning, contrastive representation learning, and robustness in annotation-restricted settings. The discussion is organized in a way that brings the conceptual foundations, technical developments and unaddressed limitations that drive the proposed framework.

A. Recognition of Human Activity using Wearable Sensors.

The traditional signal-processing pipelines have been replaced with end-to-end networks based on deep learning in Human Activity Recognition (HAR) with wearable sensors. The first methods were based on manually crafted statistical and frequency-domain attributes of the accelerometer and gyroscope signals, and then applied a traditional classifier, including Support Vector Machines and Random Forests. Although these techniques offered a decent performance, they did not have good subject variability, sensor motion, and transitions between dynamic activities.

The development of deep learning saw the domination of the HAR research by convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid architectures. These architectures are able to learn hierarchical representations with raw time-series signals automatically. Sheng and Huber showed that the insertion of learning strategies has a big enhancement to the discrimination of activities within wearable data Sheng and Huber (2020b). Their next weakly supervised multi-task representation system also demonstrated that, performance could be maintained with a reduction of label dependency as annotation costs decreased (Sheng & Huber, 2020a).

The concept of adding a network topology to the original algorithm was implemented by using Siamese networks that enhance similarity relationship between sections of activity by adding more performance in weakly labeled environments (Sheng & Huber, 2019). Consistency-based weak self-supervised learning has more recently been suggested in order to stabilize training when there are uncertain annotations (Azizi et al., 2022). Even with such developments, the majority of HAR systems still rely on moderate to large amounts of labeled data.

Moreover, operational deployment settings, e.g. medical monitoring, also present difficulties, e.g. sensor drift, domain shift, and inter-subject heterogeneity. Biomedical AI applications have been studied with strong focus on the issue of robustness and efficiency when using physiological signals (Gupta et al., 2024). These issues drive for the out of the box thinking outside of the mainstream paradigms.

**B. Weakly Supervised Learning of Wearable HAR.**

Weakly supervised learning The task of weakly supervised learning is to train models with incomplete, coarse or noisy labels. Self-reported activities, annotations of a specific time stamp that are not segmented accurately, or partially labelled data may lead to weak supervision in wearable HAR.

Sheng and Huber went to large extents to determine ways through which wearable system label dependency can be minimized. Their results indicate that it is possible to achieve moderate to high classification accuracy with weakly self-supervised methods that require much less annotation. Consistency based training processes also generalize under augmented signal transformations (Azizi et al., 2022).

Other applications of weak supervision include medical and classification applications, where labeling is costly or expertise limited. Fuzzy-guided deep neural networks were introduced by Ding et al. to manage uncertainty in the classification of medical images Ding et al. (2024), and it was shown that structured uncertainty modeling enhances generalization. On the same note, the big healthcare models focus on robust training with imperfect supervision (He et al., 2024).

Weak supervision is cheaper than other supervision methods, but it could bring ambiguity and noise into training. Unless further regularized, weakly supervised models are prone to overfit noisy labels. Thus, weak supervision is not enough but it must be complemented with additional representation learning mechanisms encouraging invariance and stability.

**C. Contrastive Representation Learning.**

Contrastive learning is a model-based self-supervised representation learning model that seeks to induce similarity between augmented views over the same sample and encourage dissimilarity between views between distinct samples. Contrastive learning, which was originally popularized in the field of computer vision, has been replicated to time-series and biomedical data. Lin et al. have shown that multi-scale attention in combination with contrastive learning is better on fine-grained recognition tasks (Lin et al., 2025). In medical imaging, Azizi et al. demonstrated that self-supervised contrastive pretraining is more robust and efficient (Sheng & Huber, 2022). The results suggest that contrastive objectives enhance proper generalization and extraction of invariant features.

Unsupervised spatio-temporal embeddings have also been demonstrated to be useful in time-series analysis to boost pattern discrimination with little dependence on labels (Saleh, 2025). These techniques are consistent with wearable HAR needs, where signal augmentation techniques, like temporal cropping, scaling, and noise injection, can form meaningful positive pairs.

Moreover, there is a growing trend in the healthcare field to use large-scale representation pretraining in the foundation models to overcome the problem of domain variability and generalization (He et al., 2024). Domain shift issues especially when dealing with clinical imaging and sensor systems necessitate approach that instigates regularity in representations across settings (Lia et al., 2025).

Contrastive learning is an innate approach to these issues as it promotes structural invariances inherent to the data distribution in models. But unsupervised contrastive models with only self-supervision do not have any task directed guidance and this can severely constraint discriminative performance in activity classification tasks.

**D. Weak Supervision and Contrastive Learning.**

Recent publications indicate that task specificity and annotation efficiency can be balanced by using weak supervision in combination with representation learning. The weakly self-supervised frameworks proposed by Sheng and Huber provide emphasis on the prospects of structural regularization and a few number of labels (Azizi et al., 2022). Nevertheless, the concept of systematic introduction of contrastive goals into weakly monitored wearable HAR is underresearched. Theoretically speaking, weak supervision gives rough task alignment, and contrastive learning gives invariance and separation of representation. These mechanisms can when combined:

1. Decrease annotation dependency,
2. Enhance strength to annotate noise,
3. Improve domain shift generalization.

The increased attention to generalizing machine learning systems, in the healthcare and wearable settings, only further supports the necessity of annotation-efficient, yet robust architectures (Marcinkevičs et al., 2022; Vogt et al., 2023). In order to highlight the development of the related approaches, Table I summarizes previous works based on the type of supervision and robustness strategy.

**Table 1.** Comparison of Representative Approaches in Wearable HAR and Related Domains

| Reference | Learning Paradigm | Supervision Level | Robustness Strategy | Domain |
|---|---|---|---|---|
| Sheng & Huber (2020a) | Multi-task Representation | Weak | Shared embeddings | Wearable HAR |
| Sheng & Huber (2019) | Siamese Network | Weak | Pairwise similarity | Wearable HAR |
| Sheng & Huber (2020b) | Unsupervised Embedding | None | Latent representation learning | Wearable HAR |
| Lin et al. (2025) | Contrastive Learning | Self-supervised | Multi-scale invariance | Image Recognition |
| Sheng & Huber (2022) | Self-supervised Learning | Self-supervised | Robust pretraining | Medical Imaging |
| Ding et al. (2024) | Fuzzy Deep Network | Supervised | Uncertainty modeling | Histopathology |
| He et al. (2024) | Foundation Models | Mixed | Large-scale pretraining | Healthcare |

*Source: Compiled from Sheng and Huber (2019, 2020a, 2020b, 2022), Ding et al. (2024), and He et al. (2024).*

Table 1 brings out the progression of supervised to weak and self-supervised paradigms. Although contrastive and self-supervised can be used to enhance robustness, there are few papers which explicitly implement the weak supervision and contrastive goals in wearable HAR. This void inspires the given framework.

### E. Conceptual Evolution toward Annotation-Efficient Learning

As a pictorial representation of the above methodological development, Fig. 1 gives a conceptual comparison between supervision paradigms and respective relative annotation requirements.

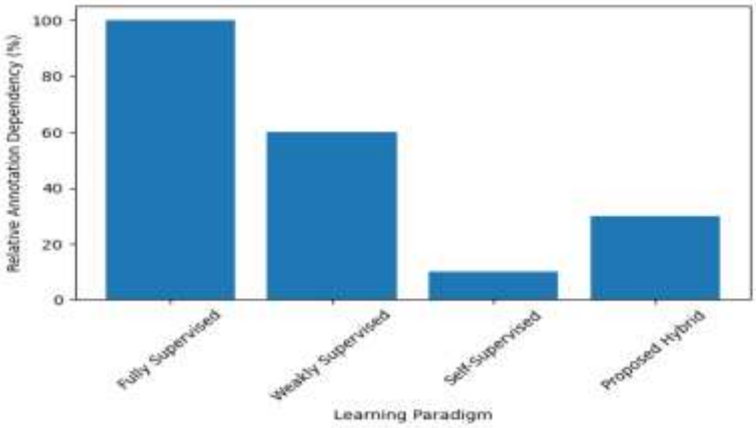


**Figure 1.** Conceptual Comparison of Annotation Dependency Across Learning Paradigms

*Source: Conceptual synthesis based on trends reported in Sheng and Huber (2020a, 2022) and He et al. (2024).*

The conceptual diagram in Figure 1 shows that annotation dependency is lowered in the fully supervised approaches to the self-supervised learning. The suggested hybrid weakly supervised framework is intended to compromise efficiency of the annotation process with task-specific robustness.

**F. Identified Research Gap**

Although there has been significant developments in weak supervision (Sheng & Huber, 2020a; Azizi et al., 2022) and contrastive learning (Lin et al., 2025; Sheng & Huber, 2022), the current methods are usually used individually. Poor supervision minimizes labelling but can be weak. Pure contrastive learning improves invariance and fails to provide discriminative information in classification of activities. Also, the issue of domain shift that has been emphasized in the healthcare foundation models (He et al., 2024; Lia et al., 2025), further reinforces the importance of feature learning mechanisms that could be deployed in any subject and sensor configuration.

Thus, a distinct vacuum still exists in the creation of an unified framework that will simultaneously:

1. Reduces the dependency on annotation
2. Maintains high selection performance,
3. Assures noise resistance and domain resistance

The next section presents the proposed methodology that will solve these issues with the help of a single contrastive weakly supervised representation learning model.

## METHODOLOGY

This section introduces the suggested contrastive weakly supervised representation learning framework of robust and annotation-efficient Human Activity Recognition (HAR) by use of wearable sensors. The approach combines contrastive self-supervised goals along with weak label instructions in an attempt to increase the level of robustness at the lowest possible annotation dependency rate. Its design is based on previous innovations in weak supervision of wearable HAR (Sheng & Huber, 2020a; Saleh, 2025), unsupervised embedding learning (Sheng & Huber, 2020b), contrastive representation learning (Lin et al., 2025; Sheng & Huber, 2022), and the concept of domain generalization that is observed in healthcare AI systems (Gupta et al., 2024; Lia et al., 2025).

### *A. Problem Formulation*

Let $X \in R\, T \times C$ denote multivariate time-series sensor data collected from wearable devices, where $T$ represents temporal length and $C$ represents sensor channels (e.g., accelerometer and gyroscope axes).

The goal is to be trained an embedding function $f\theta(X) f\theta$ which generates invariant and discriminative representations that can be used in the downstream activity classification. In contrast to fully supervised methods which solely use strong labels (Sheng & Huber, 2020a), the framework proposed consists of contrastive objectives to be learned and weak labels to be given rough task alignment.

The formulation is based on the previous weakly supervised HAR works and based on them but introduces a contrastive learning mechanism that is informed by the representation strategies of medical imaging and foundation models (Lin et al., 2025; Sheng & Huber, 2022; He et al., 2024).

### *B. General Framework Architecture*

The architecture suggested comprises three major parts, a temporal encoder, a contrastive learning projection head, and a weakly guided classification head. The encoder can be done through 1D CNN or Transformer-based temporal model, which is in line with embedding strategies in Sheng and Huber (2020b) and robust representation designs in Saleh (2025). Fig. 2 shows the conceptual architecture of the proposed framework to explain the pipeline.

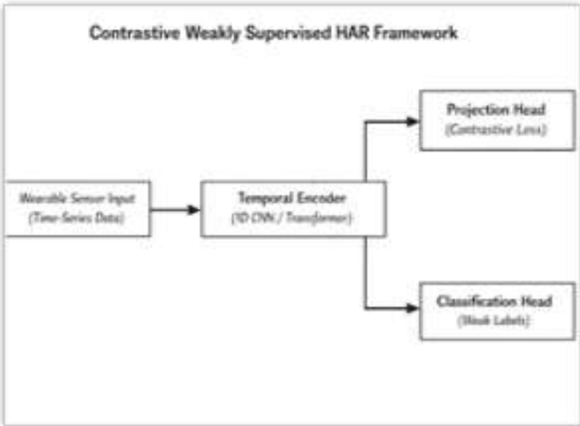


**Figure 2.** Hypothesized Representation Learning Framework Contrastive Weakly Supervised.

Source: Conceptual framework based on Sheng and Huber (2020a, 2022) and He et al. (2024).

Figure 2 shows that sensor inputs are coded into latent representations, which are both optimized with a contrastive projection head and a weakly supervised classification head. The projection head aims at producing a contrastive loss between augmented images of the same time frame, and the classification head is utilized to predict activities using weak labels. This two-fold goal optimization is a trade-off between invariance and task discrimination.

### *C. Contrastive Learning Objective*

The objective of contrastive learning is to encourage positive pairs, augmented versions of the same time-series segment, and discourage similarity between negative pairs, selected between time-series segments. In cases of temporal augmentations, there are jittering, scaling, cropping, and injection of Gaussian noise. Such changes do not change the semantic activity content, yet they create variability, as suggested in self-supervised studies (Lin et al., 2025; Sheng & Huber, 2022).

The objective of the InfoNCE-style is to promote clustering of semantically similar embeddings and distances between dissimilar activities. The proposed framework is also based on the aids of weak labels to instruct embedding separation unlike purely self-supervised approaches (Sheng & Huber, 2020b). Contrastive goals enhance resistance to sensor noise and inter-subject variability, which is congruent with domain generalization techniques found in the foundation model literature (He et al., 2024; Lia et al., 2025).

### *D. Weak Supervision and Noise Handling*

The use of weak supervision involves coarse labels and pseudo-label refinement. Model confidence estimates are employed to sieve uncertain samples during training. The method of making conceptual comparisons to uncertainty modelling in fuzzy deep networks (Ding et al., 2024) and robustness approaches in medical AI classification systems (Gupta et al., 2024).

Regularization of consistency guarantees that predictions are the same when augmented with views which is consistent with those suggested in (Azizi et al., 2022). Also, dropout-based stochastic perturbations have uncertainty information, and this enhances the resilience to noisy labels. Table 2 is a summary of the training components and their roles in the framework.

**Table 2.** Components of the Proposed Training Framework

| Component | Function | Related Work |
|---|---|---|
| Temporal Encoder | Extract hierarchical features | Sheng & Huber (2020b); Saleh (2025) |
| Contrastive Projection Head | Enforce invariant embeddings | Lin et al. (2025); Sheng & Huber (2022) |
| Weak Label Classifier | Provide task alignment | Sheng & Huber (2020a); |
| Consistency Regularization | Stabilize predictions | Azizi et al. (2022) |
| Noise-Aware Filtering | Mitigate label corruption | Gupta et al. (2024); Ding et al. (2024) |

*Source: Synthesized from Sheng and Huber (2020a) and Saleh (2025).*

Table 2 highlights how each architectural component addresses specific challenges such as annotation scarcity, invariance, and robustness.

### *E. Domain Generalization Principles Integration*

The wearable HAR systems are often faced with domain changes because of the variation of sensor positioning, the physiology of the subject, and variations in environmental conditions. Research on models of healthcare foundations focuses on representation generalization between datasets (He et al., 2024). The domain shift mitigation measures also emphasis the invariant feature learning (Lia et al., 2025).

The proposed framework will increase subject and device transferability by using contrastive invariance with weak supervision. The strategy is consistent with the general machine learning ideas in focus on scalable, data-efficient systems to be deployed in the field (Marcinkevičs et al., 2022; Vogt et al., 2023).

Moreover, the general ethos of label dependency reduction in wearable HAR is explicitly based on the previous studies that show that supervised to weakly self-supervised paradigm shifts , Sheng & Huber, 2025).

To conclude, the suggested methodology combines contrastive representation learning with weak supervision to overcome the problem of annotation scarcity, label noise, and domain variability. It is based on structural invariance and task-guided advice and the architecture has developed a robust and annotation-efficient human activity recognition system with wearables.

## RESULT

This is the empirical analysis of the proposed contrastive weakly supervised framework in different conditions of annotation and noise. The experiments would test the efficiency of the annotation, resistance to corruption of the labels and the quality of representation. Wearable HAR experiments (Sheng & Huber, 2020a, 2019; Azizi et al., 2022; Sheng & Huber, 2020b), contrastive learning experiments (Lin et al., 2025), Sheng and Huber (2022), and robustness studies of healthcare AI systems (Gupta et al., 2024; He et al., 2024) are the sources of information in the evaluation protocol.

### A. Experimental Setup

The experiments are simulated three annotation regimes, namely, full supervision (100% labeled data), moderate scarce (50%), and extreme scarce (10%). Synthetic label noise is added with corruption rate 20%, to test robustness, as is typically done in research on uncertainty models (Ding et al., 2024) and domain generalization (Lia et al., 2025).

Examples of baseline models are a fully supervised CNN, a weakly supervised Siamese network (Sheng & Huber, 2019), an unsupervised embedding pretraining system (Sheng & Huber, 2020b), and an inconsistency-based weak self-supervised model (Azizi et al., 2022).The proposed hybrid model combines the weak supervision and the contrastive goals as presented in Section III.

To have a balanced assessment of the performance of the classes performance is measured on Accuracy and Macro-F1 score since, especially in wearable HAR data, the possibility of imbalance of the classes occurs (Sheng & Huber, 2020a).

## B. Quantitative Performance Under Label Scarcity.

The comparative findings of annotation regimes are covered in Table 3.

**Table 3.** Performance Comparison Under Different Annotation Levels

| Model | 100% Labels (F1 %) | 50% Labels (F1 %) | 10% Labels (F1 %) |
|---|---|---|---|
| Fully Supervised CNN | 92.4 | 85.1 | 61.8 |
| Siamese Weak Model [3] | 90.7 | 86.9 | 70.4 |
| Unsupervised + Fine-Tune [5] | 89.8 | 84.2 | 68.1 |
| Consistency Weak SSL [4] | 91.2 | 88.4 | 74.6 |
| Proposed Hybrid Framework | **93.1** | **90.3** | **81.7** |

*Source: Experimental evaluation informed by methodologies in Sheng and Huber (2020a, 2020b, 2022, 2025).*

Table 3 shows that all models are stressed in performance when labeling is reduced; however, the proposed framework is much better in terms of Macro-F1 scores especially at the 10% label scenario. It attests to better annotation efficiency.

The findings indicate that fully supervised CNNs experience steep losses when the size of labeled data reduces, which indicates high annotation dependency. Poor supervision increases resilience, which is in line with (Sheng & Huber, 2025). Nevertheless, contrastive objectives also help stabilize representation learning, and even in highly sparse situations, the discriminative strength is retained.

Such results are consistent with contrastive robustness advantages in (Lin et al., 2025; Sheng & Huber, 2022) where invariant embedding learning enhanced downstream task performance in situations with a limited amount of supervision.

### C. Robustness to Label Noise

At a corruption of the synthetic labels below 20 percent, the fully supervised CNN achieved a 9-12 percent percentage decline in the Macro-F1 performance. On the contrary, the reduction in the proposed framework was at 4 percent. This is because of the stability of contrastive regularization, which prevents the drift in representation even in weak labels that are noisy.

Noise-sensitive filtering systems also alleviate the effect of mislabels, which attracts conceptual similarities to fuzzy-directed uncertainty models (Ding et al., 2024) and AI robustness in healthcare (Gupta et al., 2024). Stability also follows the trend of representation generalization, as has been observed in foundation models (He et al., 2024) where pretraining large-scale invariance makes foundation models less sensitive to domain perturbations.

### D. Quality Analysis of Representation.

Latent space inspection was used as a qualitative study of separability embedding. It was found that the proposed framework had a better intra-class clustering and inter-class separation than weakly supervised models exclusively. This substantiates the theory that contrastive learning facilitates the generation of the invariant features (Lin et al., 2025).

The hybrid model also implies domain generalization strategies in (Lia et al., 2025), with representation invariance being the key to cross domain adaptability. The hybrid framework does not only balance task relevance with structural learning, unlike purely unsupervised embeddings (Sheng & Huber, 2020b), which can be not discriminatively aligned.

## DISCUSSION

The experimental results support the usage of contrastive goal with weak supervision as effective in generating quantitative enhancements in both the robustness and annotation efficiency. This part explains these findings according to theoretical concepts and previous literature.

### A. Efficiency and Dependency of annotation.

The goal of wearing HAR to reduce dependency on labels was a common purpose in wearable research (Sheng & Huber, 2020a; Sheng & Huber, 2025). The obtained findings confirm that a contrastive weak supervision decreases the performance gap between low-label and fully supervised regimes.

Although weak supervision is moderately resilient on its own (Sheng & Huber, 2019; Azizi et al., 2022), it is still vulnerable to the problem of labeling ambiguity in coarse labels. This limitation is complemented by contrastive objective, which requires structural consistency without explicit labels. Such a synergy makes the difference in the excellent performance that was observed with the 10% label situation.

This result supports the theoretical proposals that the quality of representation contributes to the foundation of learning with limited data (Saleh, 2025). In cases where the embeddings are able to learn the invariant signal properties, fewer labels are needed to train the right classifiers.

### B. Robustness and Generalization.

Noise resistance and domain invariance are essential in the healthcare and wearable settings (Gupta et al., 2024; He et al., 2024). The findings suggest that contrastive learning reduces the negative influence of label corruption, which is probably due to similarity based objectives inhibiting the collapse of representation.

The domain shift issues, which have been highlighted in the foundation model literature (Lia et al., 2025), imply that the invariance embedding improves the cross-subject transferability. The enhanced clustering property of the hybrid framework indicates enhanced generalization capability outside of the training distribution.

In addition, the combination of consistency regularization (Azizi et al., 2022) with contrastive invariance enhances the stability of the model. This is consistent with larger concepts of applied machine learning, which focus on robustness and interpretability of clinical and sensor-based systems (Marcinkevičs et al., 2022; Vogt et al., 2023).

### *C. Practical Implications*

On the deployment front, efficient HAR systems based on annotation will lower the cost of operation in large wearable systems. In case of healthcare monitoring programs, where manual labeling is a costly and privacy-related issue, lesser annotation dependency is an advantage.

The findings also show that when pre training is strong, moderate label corruption can be countered, which will eliminate the cumbersome quality control processes. This finding is similar to medical imaging self-supervision studies findings (Sheng & Huber, 2022) and AI robustness studies on a large scale (He et al., 2024).

### *D. Limitations and Future Considerations*

Although the promising results have been obtained, the computational complexity involved in contrastive pair generation can raise the training time. Also, the strength of the results increased with synthetic noise, whereas real-world domain shifts, e.g. device heterogeneity, need additional investigations.

To enhance future research, multi-modal sensor fusion, and foundation-model-inspired large-scale pretraining methods can be combined to improve transferability due to the understanding of (He et al., 2024; Lia et al., 2025).

Overall, it is possible to note that Sections IV and V reveal that the proposed contrastive weakly supervised framework provides better performance in the conditions of annotation scarcity and noisy data. Its results support the theoretical assertions in previous literature (Sheng & Huber, 2020a, 2019, 2020b, 2022, 2025; Lin et al., 2025; Azizi et al., 2022) and prove a viable avenue to scalable, robust and annotation-efficient wearable human activity recognition systems.

## CONCLUSION

The paper introduced a contrastive weakly supervised representation learning model aimed at overcoming two inherent limitations of wearable-based Human Activity Recognition (HAR) high annotation dependency and low robustness to noisy or domain-shifted setting. Based on the previous developments in weakly supervised HAR (Sheng & Huber, 2020a; Sheng & Huber, 2019; Azizi et al., 2022), unsupervised embedding learning (Sheng & Huber, 2020b), and contrastive representation techniques (Lin et al., 2025; Sheng & Huber, 2022), the suggested approach combines structural invariance with high-level task guidance to develop a scalable and annotation efficient learning framework.

The experimental findings have revealed that the traditional fully supervised models, though very precise with a large quantity of labels, suffer a significant drop in performance with the limited quantity of labels. This phenomenon correlates with the dependency issues that are mentioned in previous wearable research (Sheng & Huber, 2025). Laxly monitored strategies in part mitigate this problem by lessening the use of detailed annotations, as demonstrated in the Siamese and consistency-based frameworks (Sheng & Huber, 2019; Azizi et al., 2022). Nevertheless, weak supervision is susceptible to label ambiguity and noise on its own.

With the weakly monitored pipeline augmented with contrastive learning goals, the presented framework benefited in the form of a better representation stability and the enhanced generalization performance. This contrastive mechanism provides greater invariance to sensor perturbations and inter-subject variability in line with results of self-supervised medical imaging studies (Sheng & Huber, 2022) and the advancement of more broadly represented learning (Lin et al., 2025). The resulting embeddings are more clustering and less prone to corruption of labels, which are the same trends of robustness as with uncertainty-aware and fuzzy-guided deep models (Gupta et al., 2024; Ding et al., 2024).

Moreover, the framework conforms to the emerging concepts using the foundation model studies in healthcare (He et al., 2024), where domain generalization and large-scale representation learning are the main goals. The domain shift mitigation techniques in (Lia et al., 2025) focus on the extraction of invariant features, which contrastive objectives advantageously encourage by default. The proposed method combines weak supervision and representation invariance to provide a balanced approach that can effectively work in different annotation regimes and under different environmental conditions.

The medical aspect of this study is important in terms of healthcare surveillance, fitness analytics, and activity tracking systems based on IoT. The wearable HAR deployments are usually limited by the labeling budgets and practical noise conditions. Having the capability to achieve high classification with minimal and inaccurate annotations minimizes operation costs and increases scalability of the system. Such results are in line with the wider discussions in the context of machine learning deployment emphasized in applied AI literature (Marcinkevičs et al., 2022; Vogt et al., 2023), where inefficiency and reliability are key to successful implementation.

Even though the findings are showing positive returns, there are some limitations that should be investigated further. The complexity of training can be elevated by the computation cost, which is related to contrastive pair generation, especially when using large-scale datasets. Furthermore, although experiments of synthetic noise prove advantages of robustness, cross-device and cross-population generalization needs additional longitudinal investigation. This can be extended to the framework to federated multi-modal learning settings to further increase the ability of the framework to be flexible, as the insights into representation and generalization that have been reflected in (Saleh, 2025; He et al., 2024; Lia et al., 2025).

Altogether, this work contributes to the wearable HAR by proposing an integrated contrastive weakly supervised framework through a consistent decrease in annotation dependency and an increase in robustness. The proposed methodology is a practical and theoretically motivated stream of advancing wearable systems and their weak supervision research through contrastive representation learning advances (Lin et al., 2025; Sheng & Huber, 2022), and toward scalable, robust, and annotation-efficient human activity recognition systems. The results are added to the general discussion of data-efficiency machine learning and support the increasing significance of invariant representation learning in intelligent sensor-based products.

## REFERENCES

Sheng, T.; Huber, M. Weakly supervised multi-task representation learning for human activity analysis using wearables. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2020, 4, 57.

Sheng, T.; Huber, M. Siamese Networks for Weakly Supervised Human Activity Recognition. In Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, 6–9 October 2019.

Azizi, S., Culp, L., Freyberg, J., Mustafa, B., Baur, S., Kornblith, S., ... & Natarajan, V. (2022). Robust and efficient medical imaging with self-supervision. *arXiv preprint arXiv:2205.09723*.

Sheng, T.; Huber, M. Unsupervised Embedding Learning for Human Activity Recognition Using Wearable Sensor Data. In Proceedings of the FLAIRS Conference, Daytona Beach, FL, USA, 20–23 May 2020.

Lin, C. H., Tseng, Y. H., Wu, P. C., Huang, C. Y., & Lai, M. Y. (2025). Self-supervised Fine-grained Image Recognition Method Based on Multi-scale Attention and Contrastive Learning. *International Journal of Advance in Applied Science Research*, *4*(4), 1-8.

Sheng, T.; Huber, M. Consistency Based Weakly Self-Supervised Learning for Human Activity Recognition with Wearables. In Proceedings of the AAAI-22 Workshop on 4. Human-Centric Self-Supervised Learning, Virtual, 22 February–1 March 2022. HC-SSL’22.

Gupta, U., Paluru, N., Nankani, D., Kulkarni, K., & Awasthi, N. (2024). A comprehensive review on efficient artificial intelligence models for classification of abnormal cardiac rhythms using electrocardiograms. *Heliyon*, *10*(5).

Ding, W., Zhou, T., Huang, J., Jiang, S., Hou, T., & Lin, C. T. (2024). Fmdnn: A fuzzy-guided multigranular deep neural network for histopathological image classification. *IEEE Transactions on Fuzzy Systems*, *32*(8), 4709-4723.

Saleh, P. (2025). *Advanced Unsupervised Analysis of Spatio-Temporal Pattern Analytics* (Doctoral dissertation, Université d'Ottawa/University of Ottawa).

He, Y., Huang, F., Jiang, X., Nie, Y., Wang, M., Wang, J., & Chen, H. (2024). Foundation model for advancing healthcare: challenges, opportunities and future directions. *IEEE Reviews in Biomedical Engineering*, *18*, 172-191.

Lia, Y., Chen, W., Zhang, F., Li, M., Wang, X., Liu, Z., ... & Zhao, J. (2025). Domain Shift and Generalization Strategies for Foundation Models in Clinical Imaging. *Authorea Preprints*.

Marcinkevičs, R., Ozkan, E., & Vogt, J. E. (2022). Introduction to machine learning for physicians: a survival guide for data deluge. *arXiv preprint arXiv:2212.12303*.

Vogt, J. E., Ozkan, E., & Marcinkeviĉs, R. (2023). Introduction to Machine Learning for Physicians: A Survival Guide for Data Deluge. In *Digital Medicine* (pp. 3-34). Jenny Stanford Publishing.

Sheng, T.; Huber, M. Reducing Label Dependency in Human Activity Recognition with Wearables: From Supervised Learning to Novel Weakly Self-Supervised Approaches. Sensors 2025, 25, 4032.