Self-supervised learning for human activity recognition using 700,000 person-days of wearable data

Authors: Hang Yuan[1,2,3,5], Shing Chan[1,2,5], Andrew P. Creagh[2,4], Catherine Tong[3], Aidan Acquah[2,4], David A. Clifton[4] & Aiden Doherty[1,2]

Presented by:

Karib shams

2025-1-96-014

Presented to:

Dr. Raihan Ul Islam

Associate Professor
Department of Computer Science & Engineering

Introduction & Motivation

Accurate physical activity monitoring is essential to understand the impact on physical health and well-being

Wearable sensors enable

Current Challenges in HAR

The UK Biobank accelerometer dataset, containing over 700,000 person-days of data from over 100,000 participants, was used.

The study uses self-supervised learning to leverage the unlabelled data for improving HAR models.

Self-supervised learning applied to large-scale data for better generalization.

Unlike NLP and computer vision, HAR hasn't benefited from self-supervised learning (SSL) at scale

Wearable Sensors Enable

- Fitness & wellness tracking
- Remote patient monitoring
- Early disease detection
- Large-scale population health studies
- Personalized medicine

Current Challenges in HAR

- Lack of large, labelled datasets limits progress.
- Most HAR datasets are:
  - Small
  - Collected in artificial/lab settings
  - Scripted and supervised
- Deep learning methods underperform due to data scarcity.

Background & Related Work

Small datasets often limited previous HAR models and often worked only in controlled environments

Many existing methods for HAR relied on traditional models and small labeled datasets, which didn't scale well.

Traditional HAR approaches use:

- **Shallow machine learning models** with handcrafted features (Random Forests).
- **Supervised deep learning** (DeepConvLSTM).

Most are trained on **small, scripted datasets** in controlled environments

Self-Supervised Learning in Other Fields

SSL has revolutionized:

- **Computer vision** (MoCo, SimCLR)
- **NLP** (BERT, GPT)

Applied with success using **large unlabelled data**.

Limitations of Prior Work

Small datasets lead to overfitting and poor generalisation.

Deep learning models did not significantly outperform simpler statistical models.

Pre-training often used same data for pre-training and fine-tuning, limiting transferability.

- HAR research has not fully explored SSL at scale.
- Existing HAR SSL studies:
- Use small datasets (e.g., n = 100)
- Lack evaluation across diverse populations, devices, and tasks
- Do not address domain shift or task shift

A major gap is the lack of large-scale, diverse, and real-world data for training, leading to models that don’t generalize well.

The paper introduces a self-supervised learning (SSL) approach on large, unlabelled datasets for improved generalization.

Core Methodology

Model Architecture:

Backbone: **ResNet-V2** (18 layers, 1D convolutions, ~10M parameters)

Output vector size: **1024**

Downstream FC layer: size **512**

To improve learning:

- **Sample windows based on standard deviation**
- Emphasize **high-movement segments**

Result: Better convergence, especially in AoT & Permutation tasks

Input: **Tri-axial accelerometer data** (x, y, z axes)

Data resampled to **30 Hz**, split into **10-second windows**

Weighted Sampling Strategy:

Motion data has long **low-movement periods**

Self-Supervised Learning Tasks:

- **Arrow of Time (AoT)**: detect if signal is reversed
- **Permutation**: detect if signal chunks are shuffled
- **Time Warping (TW)**: detect if signal segments are temporally stretched/compressed

Core Methodology

Implementation Details:

Pre-training Dataset UK Biobank
- ~100,000 participants
- 700,000 person-days (free-living, unlabelled)

Batch size **6000 windows** per iteration (1500 × 4 subjects)

Optimizer **Adam** with learning rate **1e-3**

Hardware **4× Tesla V100 GPUs (32GB)**
- **~420 GPU hours** for training

Augmentation **Random axis rotation and swapping**

Multi-Task Loss Function:

Each pretext task has **cross-entropy loss**

Total loss is the **equal-weighted sum** of all task losses

Early stopping: **Patience = 5 epochs**

Fine-tuning Strategies:

Two approaches

Fine-tune **all layers**

Freeze feature extractor, fine-tune **only FC layers**

Compared against:

Random forest baseline

Deep models trained from scratch

Experiments & Results

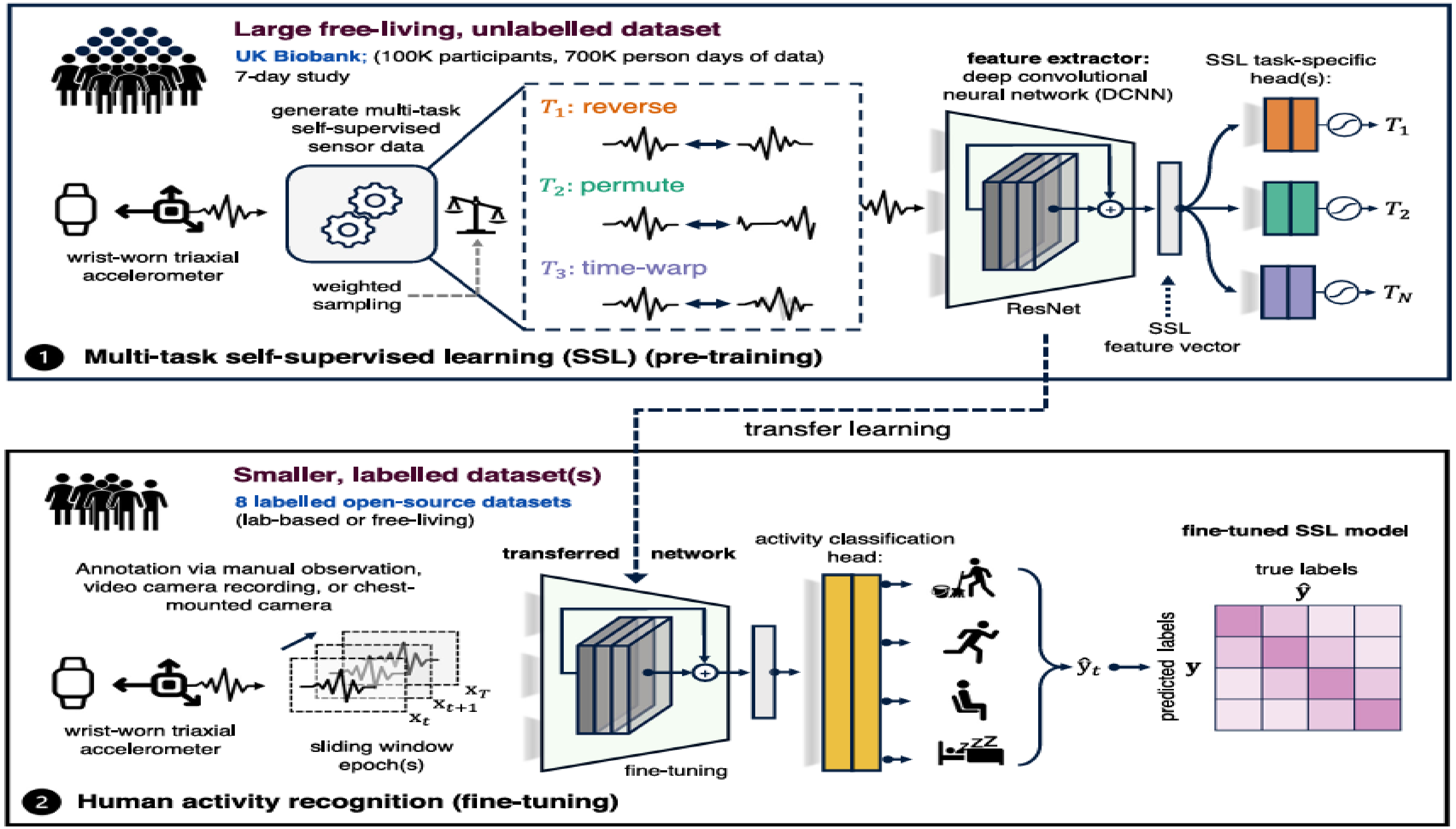


Fig. 1 Overview of the proposed self-supervised learning pipeline

Dataset	#Subjects	#Samples	#Classes	Environment	References
UK Biobank	~100K	6 B	Unlabelled	Free-living	55
Capture-24	152	573K	4	Free-living	8
Rowlands	55	36K	13	Lab	56
WISDM	46	28K	18	Semi free-living	57
MJFF-LR	28	12K	12	Lab	58
REALWORLD	14	12K	8	Lab	59
Opportunity	4	3.9K	4	Semi free-living	60
PAMAP2	8	2.9K	8	Lab	61
ADL	7	0.6K	5	Lab	62

- **Benchmark Datasets**
- Evaluated on **8 diverse labelled HAR datasets**:

Experiments & Results

Experiments & Results

Evaluation Metrics: → **F1 Score** (subject-wise) → **Cohen's Kappa (κ)** → Cross-validation: **5-fold subject-wise** (≥10 subjects) **Leave-one-subject-out** (<10 subjects)

Baselines Compared: → **Random Forest (RF)** with time-series features → **ResNet** trained from scratch → **SSL pre-trained ResNet**:

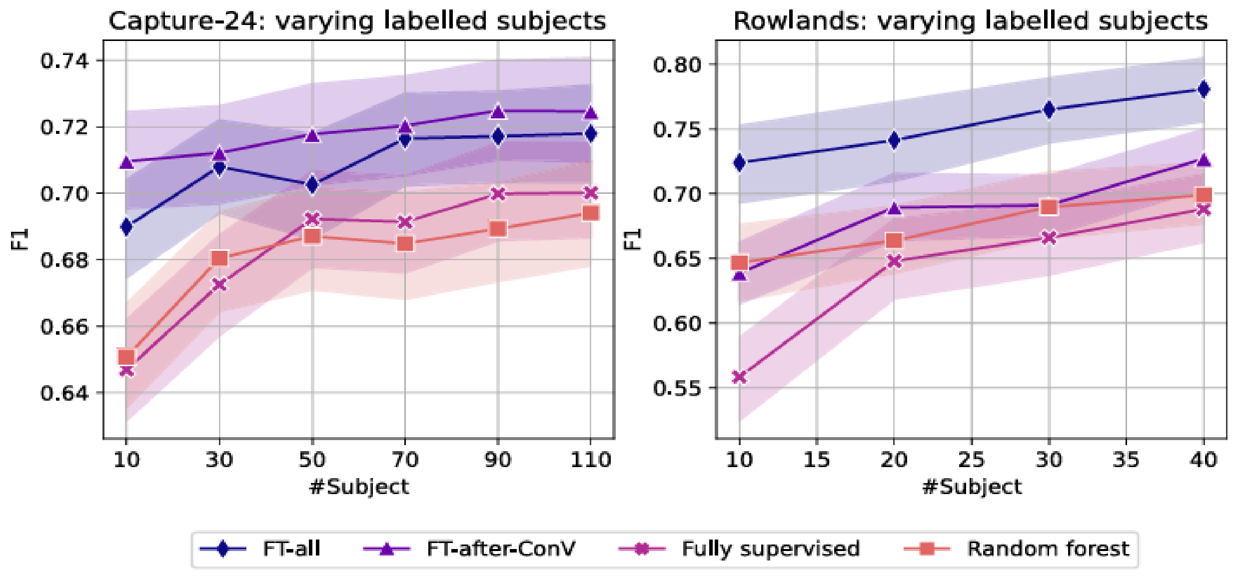
- Fine-tuned all layers
- Fine-tuned FC layers only

Key Results: → **SSL pre-trained models outperformed** both baselines on all 8 datasets → Median **F1 improvement: 24.4%** → Largest F1 gain:

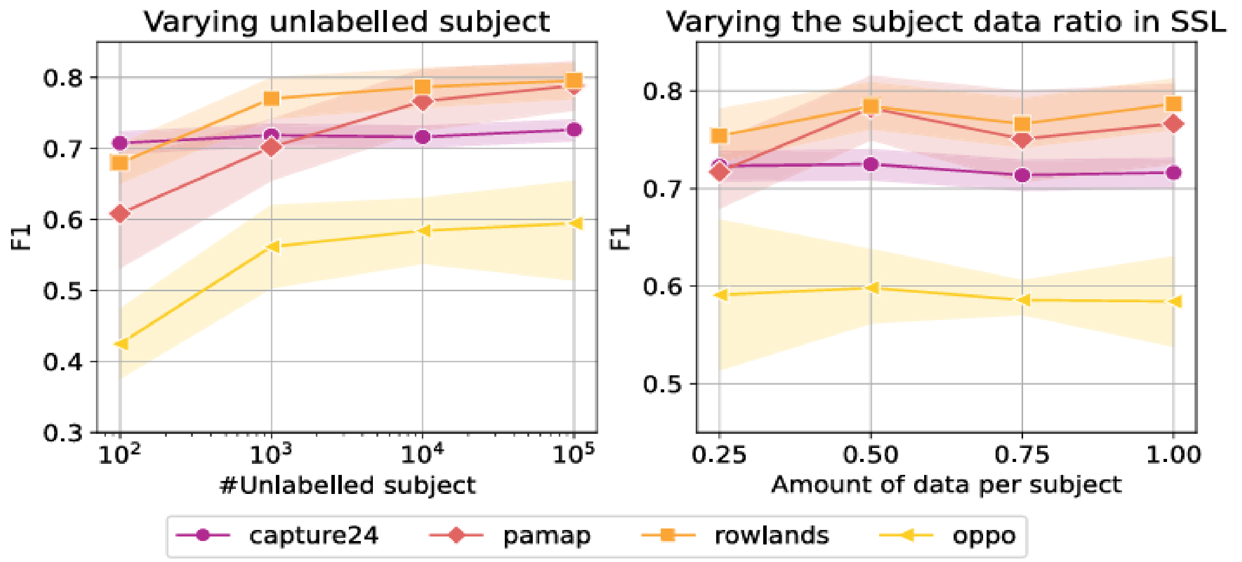
- **+130.9%** on MJFF-LR (clinical)
- **+100%** on ADL

Smaller datasets benefited most from SSL

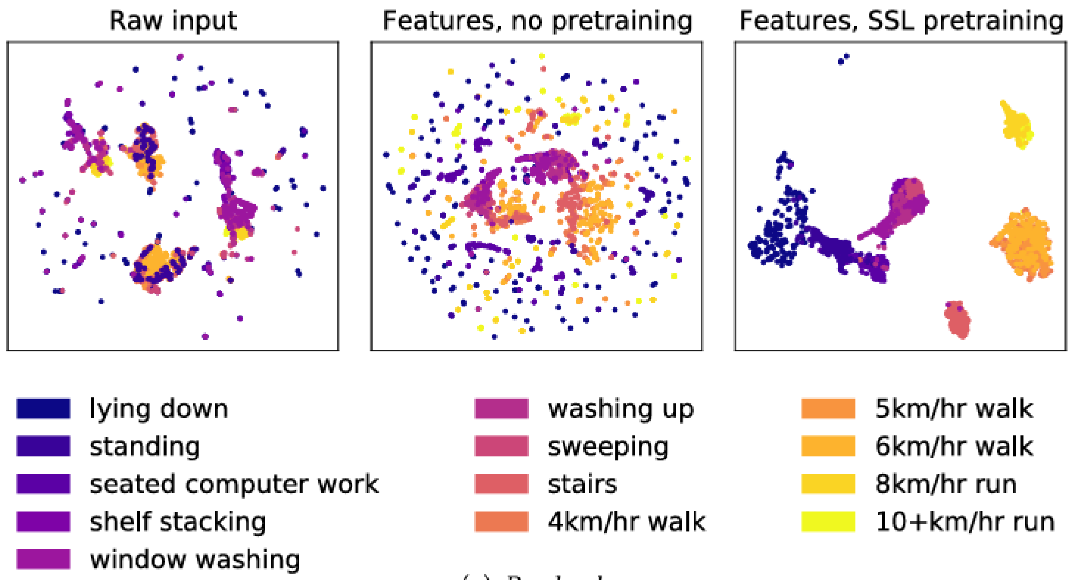
The association between data volume and model performance



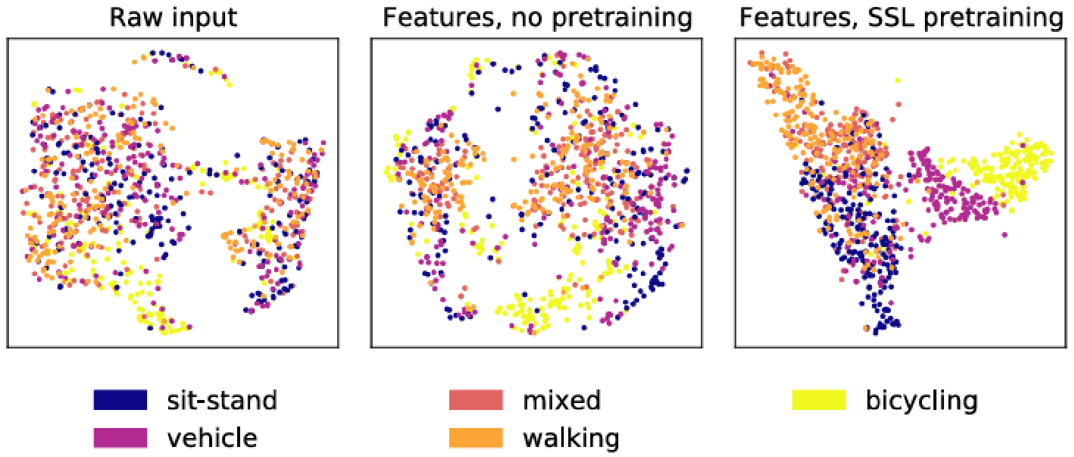
(a) How different models perform in the downstream tasks when we change the number of labelled subjects.



Cluster analysis on raw inputs, untrained features, and self-supervised-pre-trained features.



(a) *Rowlands*



(b) *Capture-24*

Critical Analysis

Strengths

•**Largest pre-training HAR dataset to date:**
"700,000 person-days of free-living data" (UK Biobank)

•**Superior generalization:**
"Pre-trained model generalizes across external datasets, tasks, devices, health statuses, and populations."

•**Reduces dependence on labelled data:**
"High performance with limited labelled data—important for clinical studies."

•**Open-source release:**
"Pre-trained models and code made available for the digital health research community."

Weaknesses & Limitations

•**Population bias:**
"Pre-training data (UK Biobank) mostly consists of Caucasians from the UK."

•**Limited sensor diversity:**
"Lack of raw accelerometer datasets from different global regions."

•**Only three SSL tasks used:**
"Other transformations (e.g., contrastive learning, autoencoders) not fully explored due to lower performance."

Critical Analysis

Reproducibility

•**High reproducibility:**
"All datasets, code, and models publicly available."
GitHub: OxWearables/ssl-wearables

Ethical Implications

•**Informed consent**:
"Most datasets include participant consent and ethics approval."

•**Data governance gap**:
"Many open HAR datasets lacked licensing or consent info due to older data collection standards."

Real-World Impact

•"Our pre-trained model can serve as a foundational HAR model that removes the need to pre-train on unseen datasets."
•Already used in:
- "Clinical populations with motor impairment"
- "Epidemiological research on sleep and mortality"

Future Directions

Incorporate multi-modal data (ECG) to improve model robustness.

Investigate more diverse datasets to reduce bias towards certain populations.

Explore newer self-supervised techniques (e.g., contrastive learning) for HAR.

Expand demographic diversity:
"Include raw accelerometer datasets from different global regions to reduce bias."

Explore additional sensor modalities:
"Incorporate electrocardiogram (ECG) and other wearable time-series data."

Evaluate newer SSL methods:
"Assess performance of recent self-supervised techniques beyond multi-task learning, such as masked reconstruction and contrastive learning."

THANK YOU ☺