Predictive Analytics for Human Activity Recognition (HAR) with OxWearables Supervised Self Learning (SSL) Models

Abideep Singh Kondal - 40241215

1 Paper Analysis

Introduction

Human Activity Recognition (HAR) has traditionally relied on manual feature engineering, often constrained by the availability of small labeled datasets. The labeling process is labor-intensive, particularly for HAR. However, collecting large-scale unlabeled data is feasible, as evidenced by datasets like the UK-Biobank (UKB). This study leverages self-supervised learning (SSL) to process real-world data, prioritizing the temporal dependencies of human motion through three tasks: Arrow of Time (AoT), Permutation, and Time Warping (TW).

The research highlights the generalizability of SSL-trained models across diverse datasets and scenarios. The proposed models outperformed traditional methods, offering improvements in environments with small labeled datasets while ensuring robust activity recognition.

Data Collection

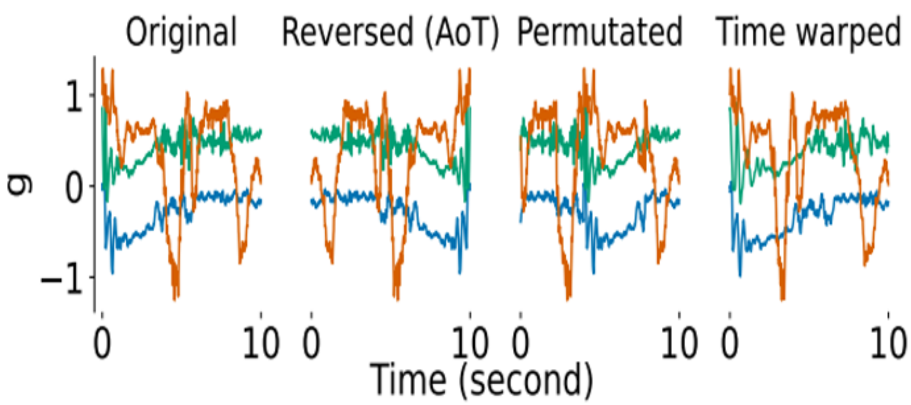
The UK-Biobank dataset is the cornerstone of this study, comprising over 700,000 person-days of accelerometer data collected at 100 Hz. Wrist-worn devices captured natural, unscripted human activities in free-living conditions, making this dataset uniquely suited for real-world applications.

- **Windowing**: Data was segmented into 10-second windows, allowing activity classification per segment.
- **Resampling**: Data was resampled to 30 Hz for consistency, capturing critical frequencies of human activity while preserving signal fidelity.
- **Evaluation**: The model was assessed on seven external datasets, varying in size (600 to 600,000 samples), activity classes (4 to 18), devices, and collection protocols (free-living, scripted, and lab-based settings).

Model Training and Architecture

The study employed a ResNet-V2 architecture with 18 layers and 1D convolutions tailored for time-series data. The feature extractor, shared across all self-supervised tasks (AoT, Permutation, TW), generated a 1024-dimensional feature vector. Additional fully connected (FC) layers enhanced downstream HAR tasks.

- **Pretext Tasks**:
  - **Arrow of Time (AoT)**: Predict whether the signal is normal or reversed.
  - **Permutation**: Identify if the signal chunks are shuffled.
  - **Time Warping (TW)**: Detect whether the signal has been stretched or compressed.



- **SSL Optimization**: The Adam optimizer with a scaled learning rate and a burn-in period of 5 epochs was used to handle large batch sizes.

- **Weighted Sampling**: Focused on dynamic, informative data segments to address the prevalence of low-activity periods in real-world data.

Evaluation Techniques

The study used two cross-validation approaches:

1. **Held-One-Subject-Out**: For datasets with fewer than 10 subjects.

2. **Five-Fold Subject-Wise**: For datasets with 10 or more subjects.

In both cases, the data was split as follows:

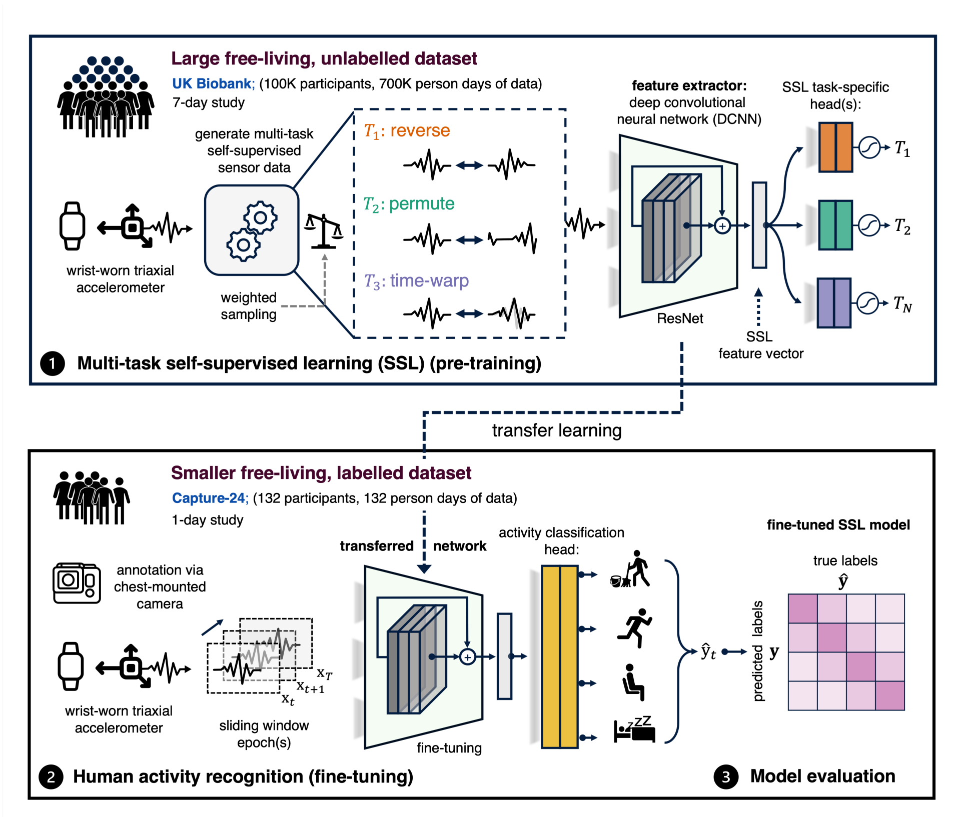
- **Training**: 70%

- **Validation**: 10%

- **Testing**: 20%

Results and Insights

The SSL models demonstrated significant improvements, particularly in small datasets, by leveraging multi-task learning (MTL) and fine-tuning techniques.



- **Transfer Learning Success**: Pre-trained models consistently outperformed models trained from scratch. Fine-tuning all layers yielded the highest performance.
- **Impact on Dataset Size**: SSL pre-training improved small datasets like Opportunity (F1 increase of 55.4%) more significantly than larger datasets like Capture-24 (F1 increase of 2.5%).
- **Pretext Task Combinations**: Using all three SSL tasks (AoT, Permutation, TW) during pre-training ensured better generalization across datasets.

Paper Conclusions

Self-supervised learning (SSL) demonstrates significant potential in enhancing HAR models, particularly when labeled data is limited. The multi-task learning framework and pretext tasks allow models to generalize effectively across diverse scenarios, datasets, and devices. The study establishes SSL as a robust alternative to traditional supervised methods, paving the way for scalable, real-world HAR applications.

2 Harnet10 Pre-trained SSL model's performance (My Analysis)

ADL Dataset Analysis

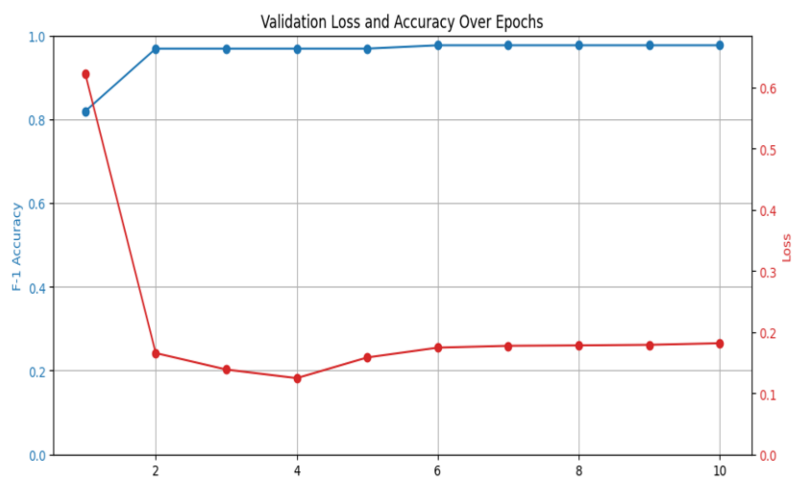
The ADL dataset was used to evaluate the fine-tuned HaRNet10 model's performance. This dataset was cleaned and resampled at 30 Hz to align with signal fidelity standards, featuring data from seven participants performing 12 distinct activities.

Model Training and Configuration:

- **Model:** Pre-trained HaRNet10 SSL model, chosen for its balance between computational efficiency and performance.
- **Classifier Adjustment:** Modified to support 12 activity classes specific to the ADL dataset.
- **Training Configuration:** Optimizer: Adam, Learning rate: 1×10^{-4}. Batch size: 32, Sliding window size: 300 samples.
- **Data Split:** 80% training, 20% validation.

Validation Results:

	Epoch	Validation Loss	Validation Accuracy	F-1 Score
0	1	0.622550	0.818898	0.791721
1	2	0.165657	0.968504	0.968170
2	3	0.138723	0.968504	0.969312
3	4	0.124685	0.968504	0.968599
4	5	0.158450	0.968504	0.968599
5	6	0.174524	0.976378	0.976343
6	7	0.177521	0.976378	0.976343
7	8	0.178184	0.976378	0.976343
8	9	0.179146	0.976378	0.976343
9	10	0.181700	0.976378	0.976343



Results and Observations:

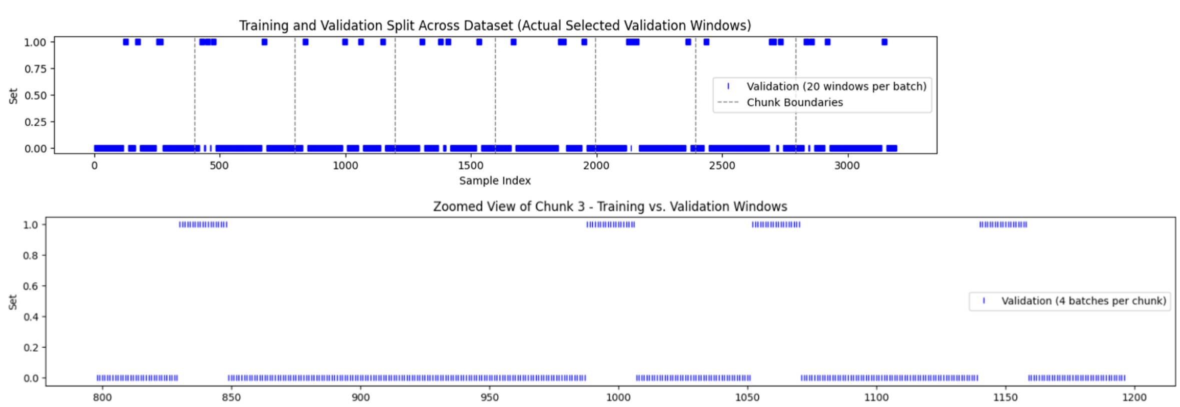
- **Validation Performance:** F1 score improved steadily from 0.79 (Epoch 1) to 0.97 (Epoch 10), showcasing the model's strong transfer learning capability.
- **Stability:** The model demonstrated no signs of overfitting, with consistent improvement across epochs.
- **Observation:** Weighted F1 scores highlighted robust activity recognition. Model performed extremely well.

HMP ADL Dataset Analysis

The HMP ADL dataset captures time-series data of daily activities for 16 participants, collected over 35 days using wrist-worn accelerometers.

Dataset and Training Highlights:

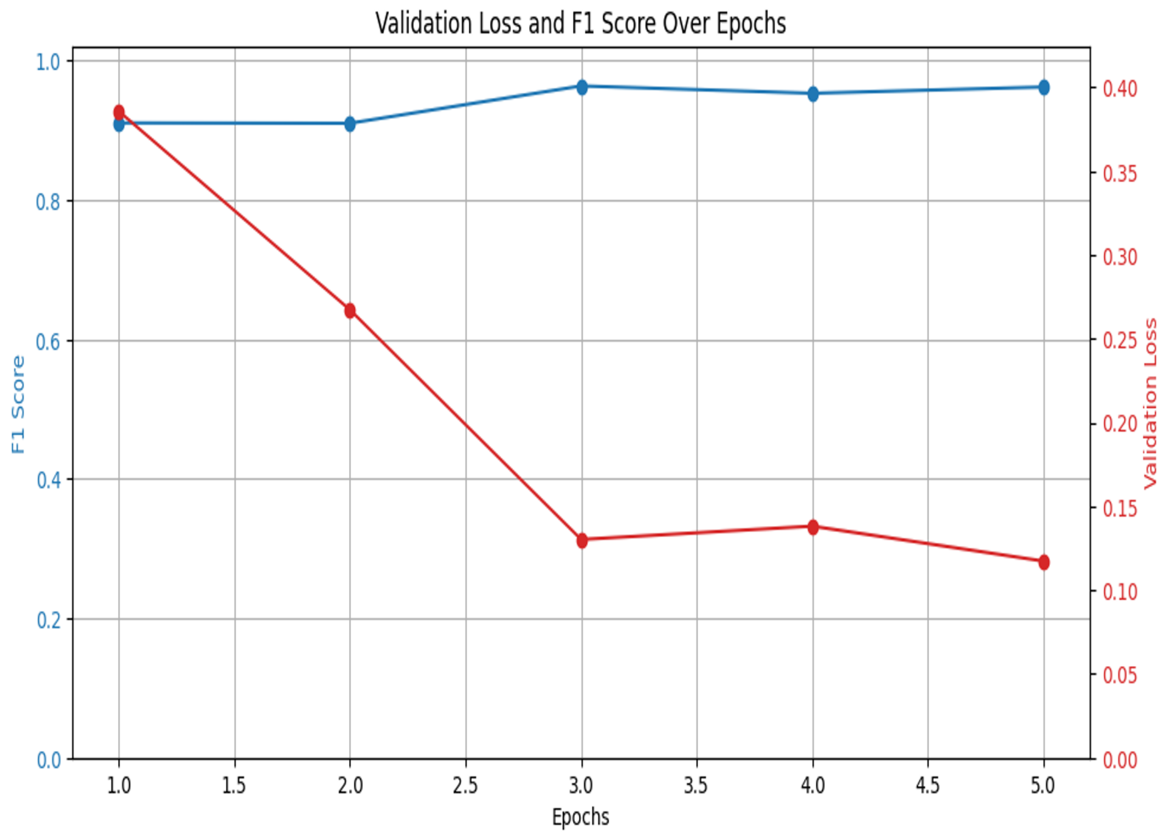
- **Data Split:** 80% training, 20% validation via stratified sampling.
- **Normalization:** X, Y, Z sensor data normalized for consistency.
- **Sliding Window:** Sequence length: 300 samples with a 50% overlap.
- **Chunking:**
  - **Why:** To ensure a non-overlapping, representative selection of validation data and prevent overfitting by maintaining strict separation between training and validation sets. This also ensures that the validation data is never seen by the training model, preventing data leakage.
  - **How:** The dataset was divided into 8 chunks, each containing approximately 12.5% of the data. Four validation batches (20 windows each) were randomly selected per chunk, while the remaining data was used for training.



Key Results:

- **Epoch-Wise Metrics:** Validation accuracy increased from 91.3% to 96.1%, with F1 scores improving from 91.0% to 96.4%.

```
Epoch 1/5, Loss: 1.1121, Validation Loss: 0.3858, Validation Accuracy: 0.9128, F1 Score: 0.9106
Epoch 2/5, Loss: 0.3260, Validation Loss: 0.2673, Validation Accuracy: 0.9030, F1 Score: 0.9101
Epoch 3/5, Loss: 0.1918, Validation Loss: 0.1303, Validation Accuracy: 0.9605, F1 Score: 0.9637
Epoch 4/5, Loss: 0.1106, Validation Loss: 0.1382, Validation Accuracy: 0.9474, F1 Score: 0.9531
Epoch 5/5, Loss: 0.0797, Validation Loss: 0.1173, Validation Accuracy: 0.9605, F1 Score: 0.9621
```



- **Observations:** Overall HaRNet10 model gave amazing downstream performance on this dataset. Sliding windows and chunking improved data representation and model generalization, enabling the model to adapt effectively to this dataset.

REALDISP Dataset Analysis

The REALDISP dataset evaluates activity recognition under three scenarios: ideal placement, self-placement, and mutual displacement. The analysis focused on left wrist accelerometer data.

Dataset Details:

- **Scenarios:** Ideal, self, and mutual displacement placements.

- **Signals Captured:** X, Y, Z accelerometer data.
- **Participants:** Data collected from 17 participants.

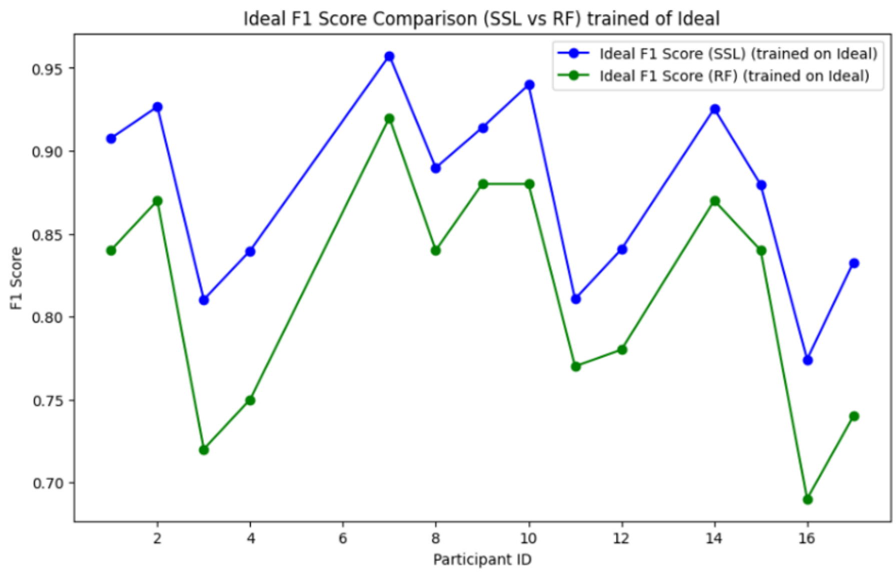
Model Training and Evaluation:

- **Pipeline:** Data is preprocessed and resampled to 30hz. Pre-trained HaRNet10 model is then fine-tuned on the dataset with Adam optimizer and CrossEntropy-Loss. As for RandomForest, relevant features are extracted.

- **Evaluation Methodology:** Left-out participant's data was used for testing. Performance was compared against a Random Forest (RF) model.

Evaluation Scenarios and Key Results

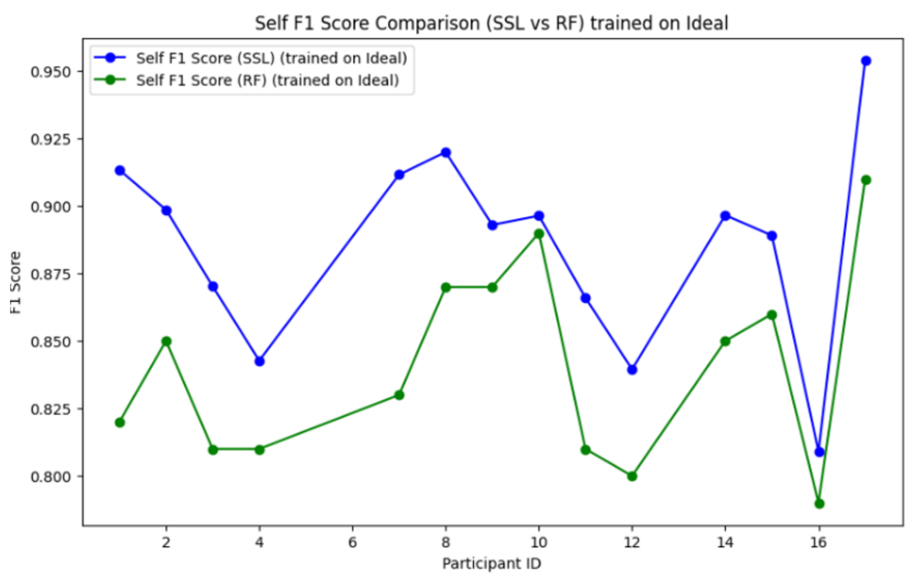
1. **Scenario 1: Ideal (train) vs. Ideal (test)**
   Key Findings:
   - (Average)SSL F1: 0.8749, (Average)RF F1: 0.8136.



- Observation: SSL demonstrated superior performance under controlled conditions.

2. **Scenario 2: Ideal (train) vs. Self (test)**
   Key Findings:

   - (Average)SSL F1: 0.8857, (Average)RF F1: 0.8407.
   - Observation: SSL models outperformed RF, showcasing generalization to real-world, user-positioned scenarios.



3. **Scenario 3: Self (train) vs. Self (test)**
   Key Findings:

   - (Average)SSL F1: 0.8905, (Average)RF F1: 0.8321.
   - Observation: SSL demonstrated robust performance in real-world self-placement scenarios, adapting better than RF.

Conclusion

Across all scenarios, SSL models consistently demonstrated superior performance compared to RF models, showcasing their robustness and adaptability in activity recognition tasks:

- **Scenario 1 (Train on Ideal, Test on Ideal):** SSL models excelled in controlled conditions, achieving an average F1 score of 0.8749.
- **Scenario 2 (Train on Ideal, Test on Self):** SSL maintained its edge, scoring an average F1 of 0.8857, demonstrating better generalization to real-world variability.
- **Scenario 3 (Train on Self, Test on Self):** SSL outperformed RF with an F1 of 0.8905, confirming its effectiveness in handling real-world data variability.

The use of pre-trained SSL models enables robust feature extraction, allowing for better activity recognition even in challenging real-world conditions. Fine-tuning and multi-task learning strategies have proven to be critical in achieving these results, demonstrating the potential of SSL in advancing human activity recognition systems.

Acknowledgments

The OxWearables pre-trained models and their associated self-supervised learning tasks, as well as the ADL dataset referenced, are the intellectual property of OxWearables. The results and evaluations presented in this report are based on their publicly available models and datasets provided on their Website. All rights and credits for the original research, datasets, and pre-trained models belong to OxWearables.

This work is a derivative analysis and evaluation conducted independently to explore and demonstrate the applicability and effectiveness of their models in different datasets and scenarios.

Links and References

Below are the relevant links and references used in this document:

- OxWearables Website

- ADL Dataset - GitHub Repository
- ADL Dataset - Analysis Google Colab Notebook link

- HMP ADL Dataset - UCI Repository
- HMP ADL Dataset - Analysis Google Colab Notebook link

- REALDISP Dataset - UCI Repository
- REALDISP Dataset - Ananlysis Google Colab Notebook link (SSL)
- REALDISP Dataset - Ananlysis Google Colab Notebook link (RF)