

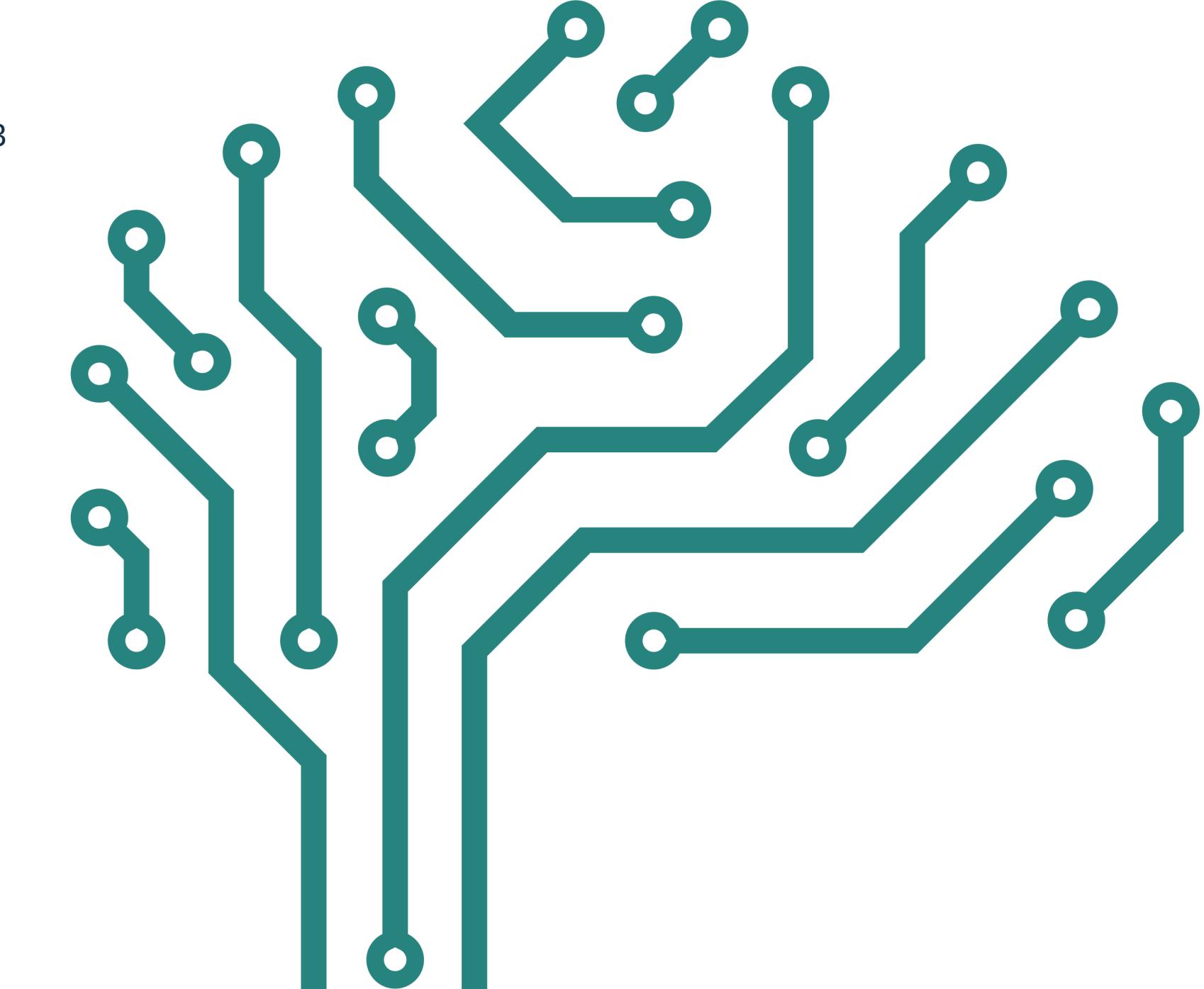
# SimCLR Framework for HAR

EXPLORE MORE

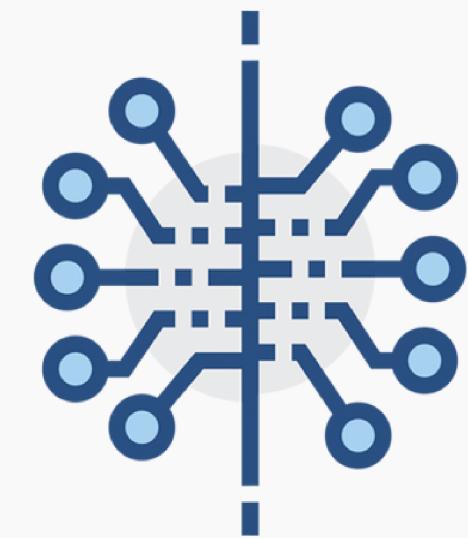


## Table Of Content

- Project Purpose
- Models structure
- Exploratory data analysis
- Model Training
- Types of dataset
- Performance Comparison
- Preprocessing: resampling and smoothing

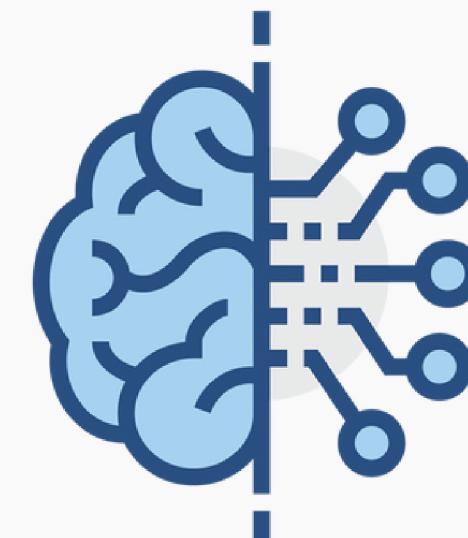


**Is it good to leverage unlabelled data when there is few labeled data?**

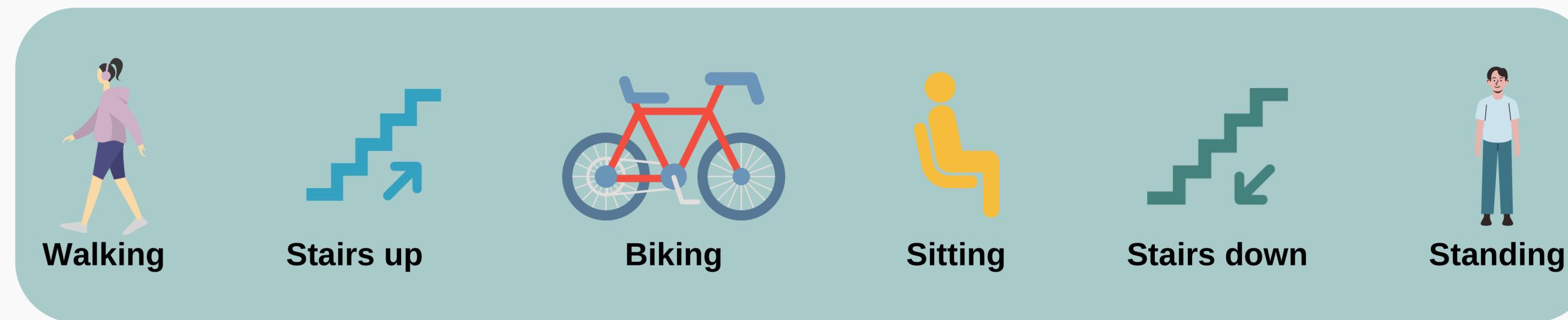


**Supervised  
Learning**

**VS**



**Self-Supervised  
Learning**



**Multiple devices recording during single experiment**

## → Single Device

- Used for **baseline model**
- Used for **downstream task**
- We will call it **Dataset A**

Most used smartphone

Most used smartwatch

## → Multiple Devices

- Used for **pretext task**
- We will call it **Dataset B**

Combination of devices recordings

**170 Minutes**

**1350 Minutes**

## PREPROCESSING: RESAMPLING AND SMOOTHING

### Temporal Alignment

Data clustered in 2 seconds windows

- Data from all the devices is grouped together if they belong to the same time interval

### Preprocessing

#### Resampling

- Estimate the **mode** of the sampling frequency
- All data is **resampled** at the same frequency

#### Removing outliers

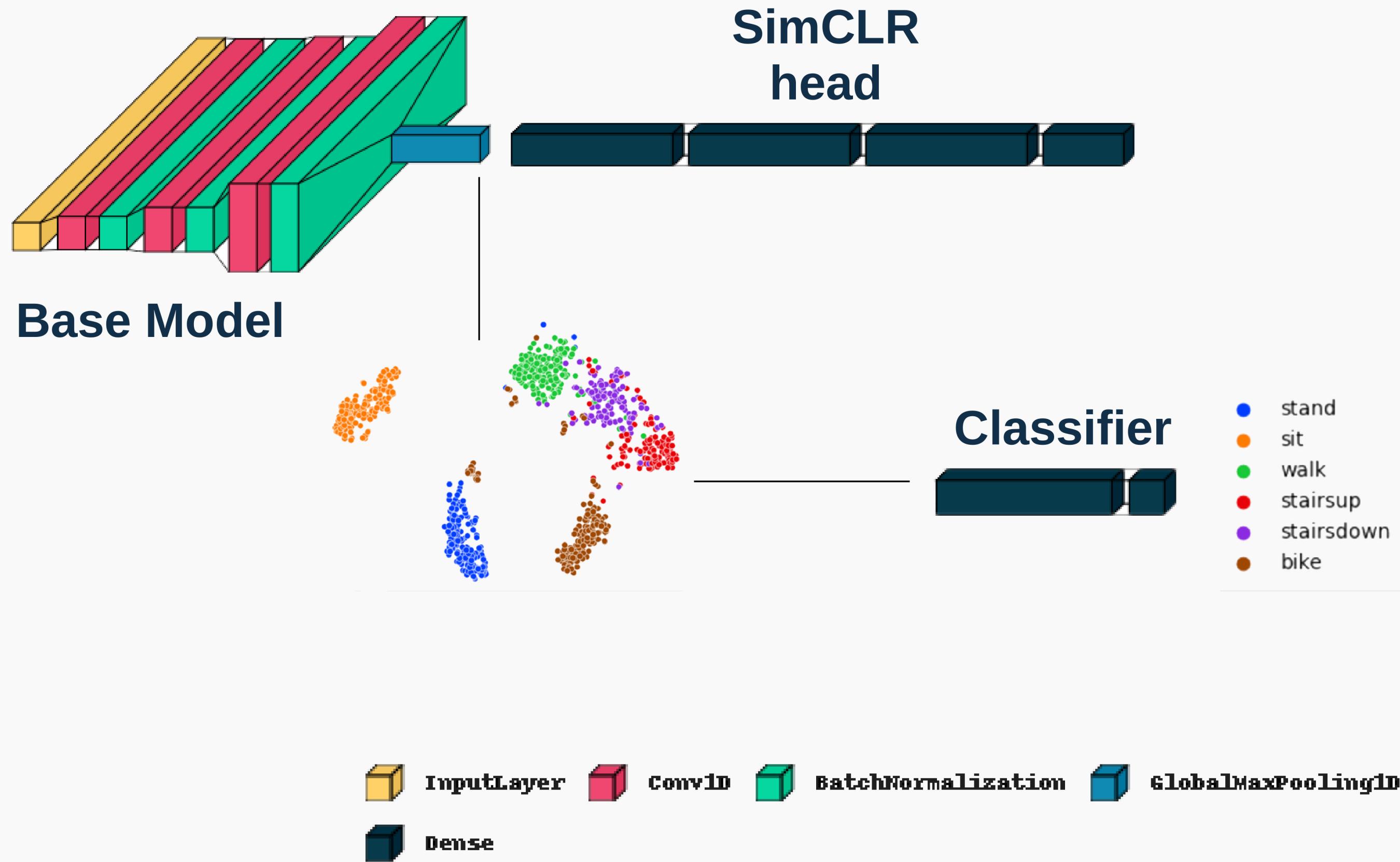
- Rolling window with **median** filter

Align data from different devices

- All data is gathered together in a single DataFrame that has temporally aligned data coming from all the available sensors

### Segmentation

Window segmentation with no overlap



MODEL TRAINING

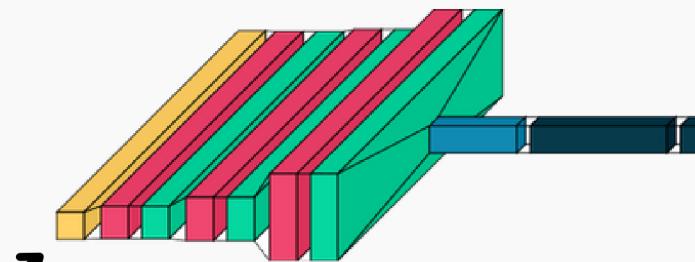
## Leave One Subject Out

CV

Dataset A



Baseline



1

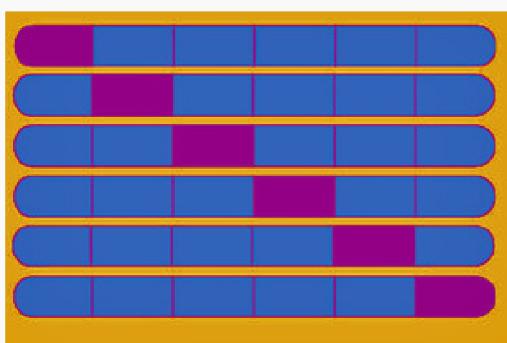
3  
10%  
25%  
50%

Downstream

Metrics Aggregation



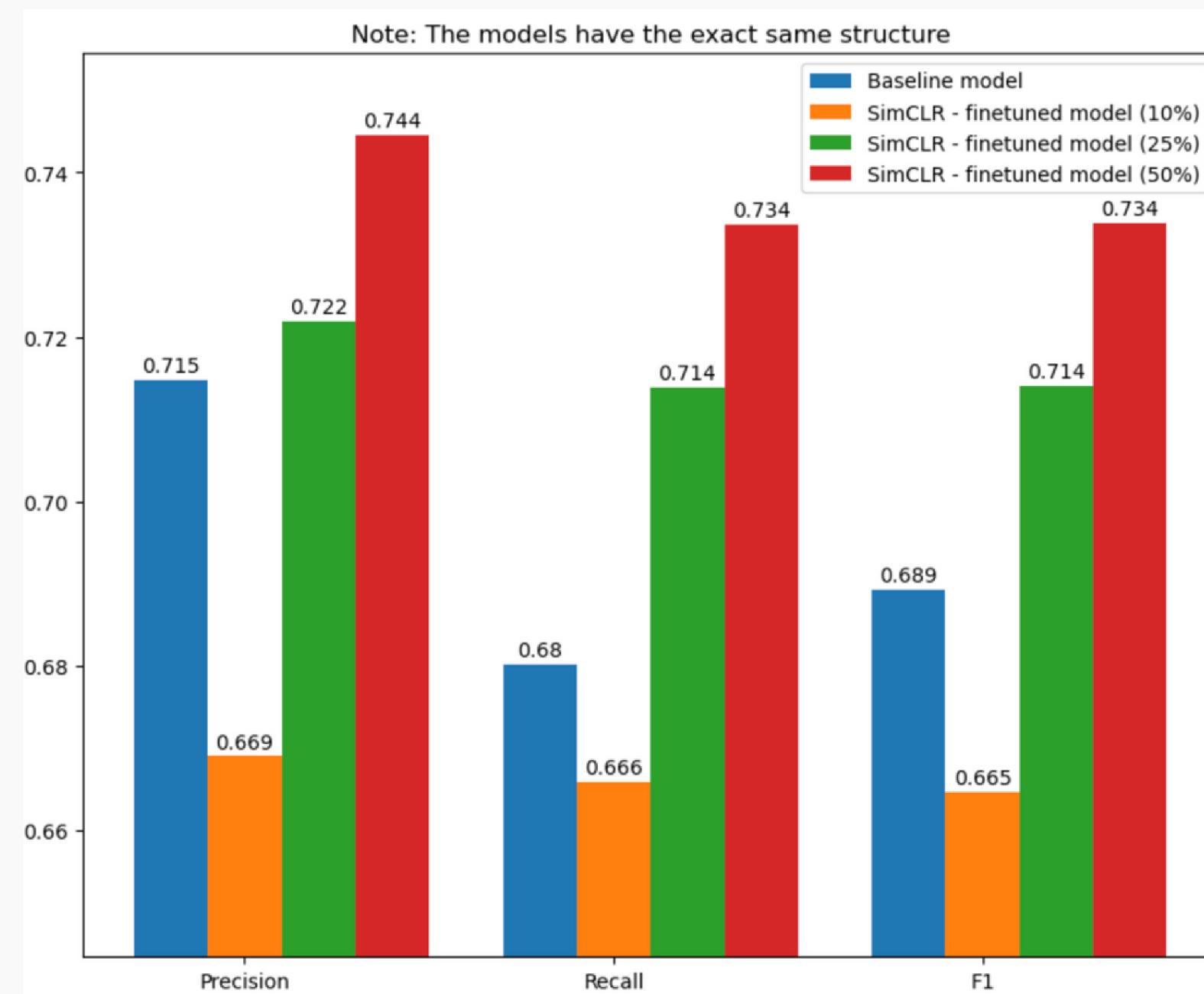
Dataset B



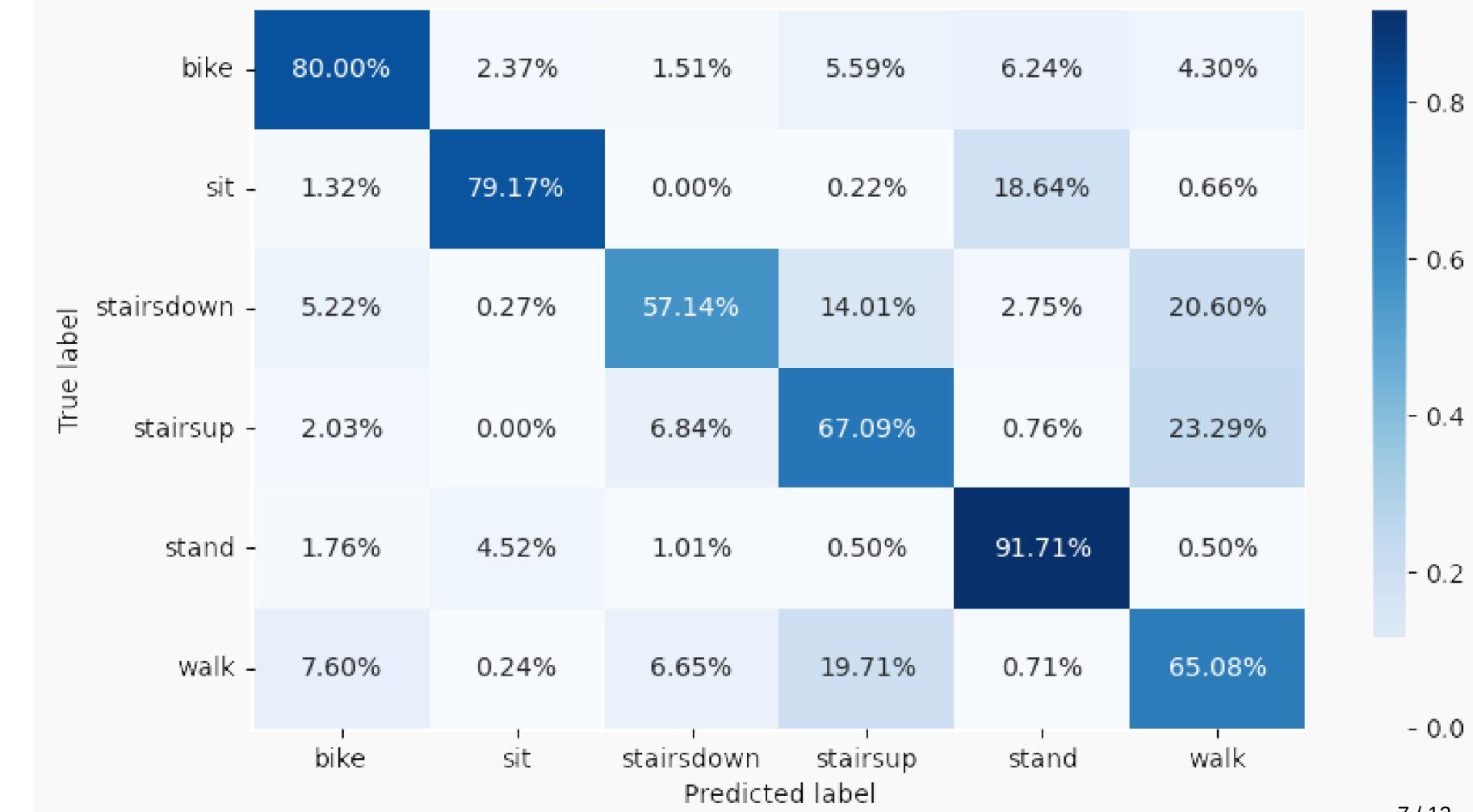
Pretext

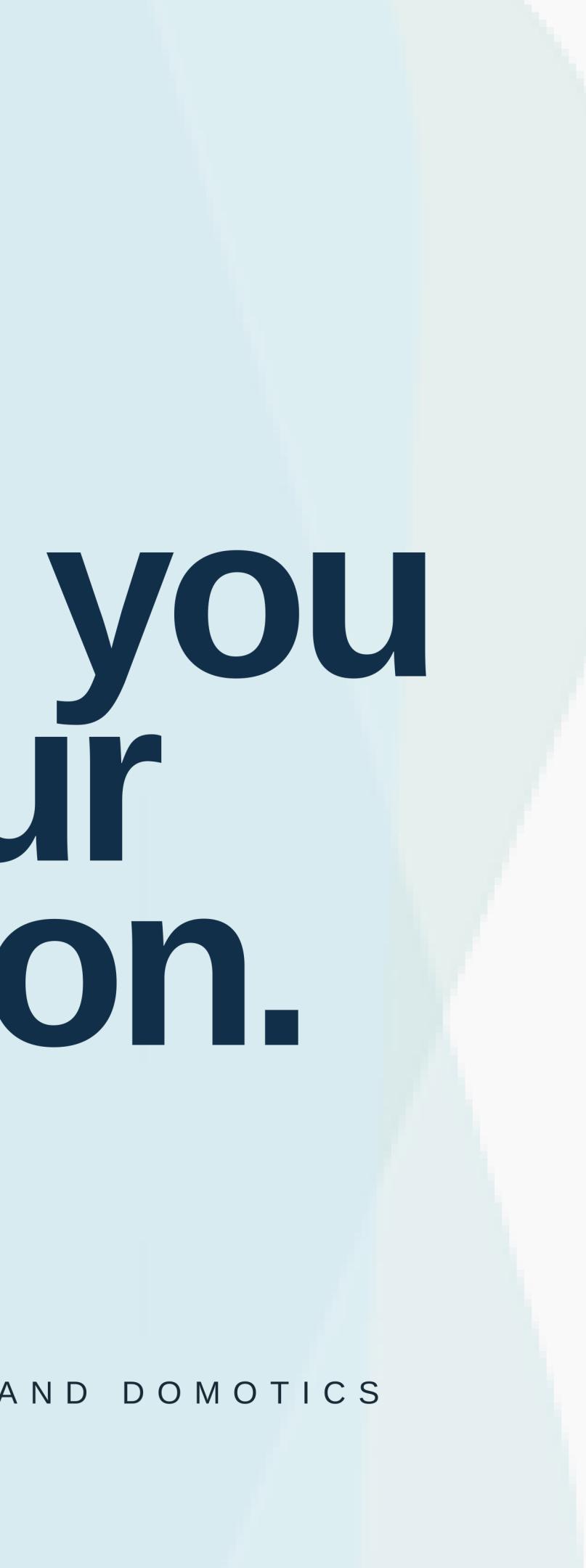
2

## Models metrics averaged by subject



## Best model (tuned with 50% of train data) normalized confusion matrix





**Thank you  
for your  
attention.**