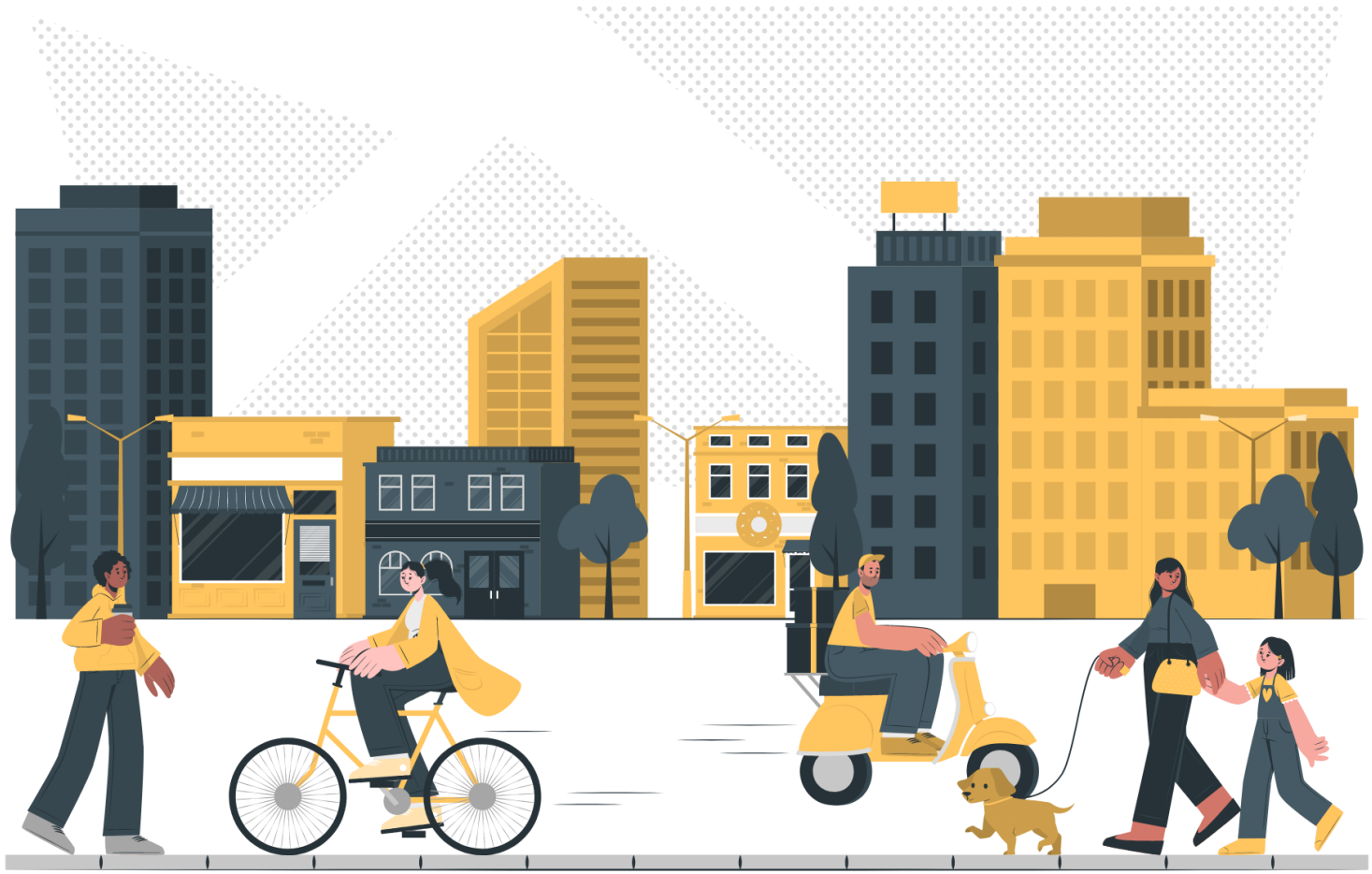


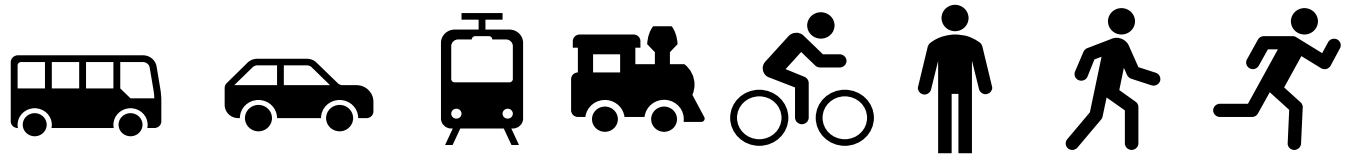
Università della Svizzera italiana

Enhancing XGBoost with Heuristic Smoothing for Transportation Mode and Activity Recognition

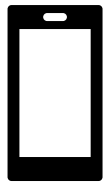
Lidia Alecci, Leonardo Alchieri, Nouran Abdalazim, Pietro Barbiero, Silvia Santini, Martin Gjoreski

Sussex-Huawei Locomotion (SHL) Challenge 2023
Human Activity Sensing Corpus and Applications (HASCA) Workshop @ UbiComp/ISWC2023, Cancun, Mexico

Dataset and Challenge Objective

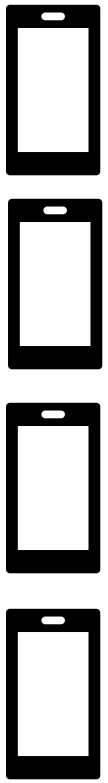
- 8 activities: 

Set	User	Body position	Number of days
Training	U1	Hand, hips, torso, bag	59
Validation	U2 and U3	Hand, hips, torso, bag	4
Test	U2 and U3	Unknown	28



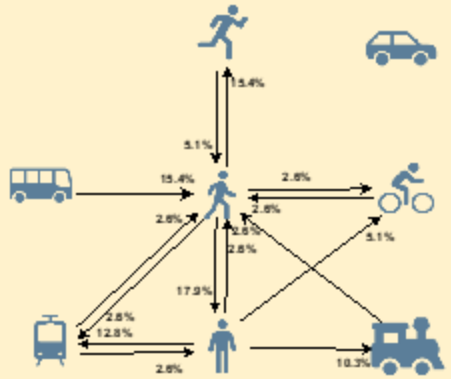
- Accelerometer (ACC)
- Magnetometer (MAG)
- Gyroscope (GYR)
- GPS and location

Methodology



1. Segmentation
2. Orientation Unification
3. Feature Extraction
4. Data Imputation
5. Feature Fusion

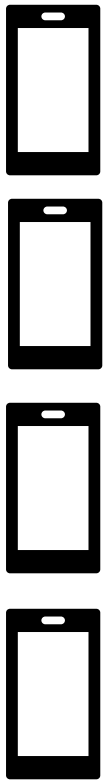
DecayXGBoost



Activity Recognition Model

1. Label Smoothing
2. Exponential Decay

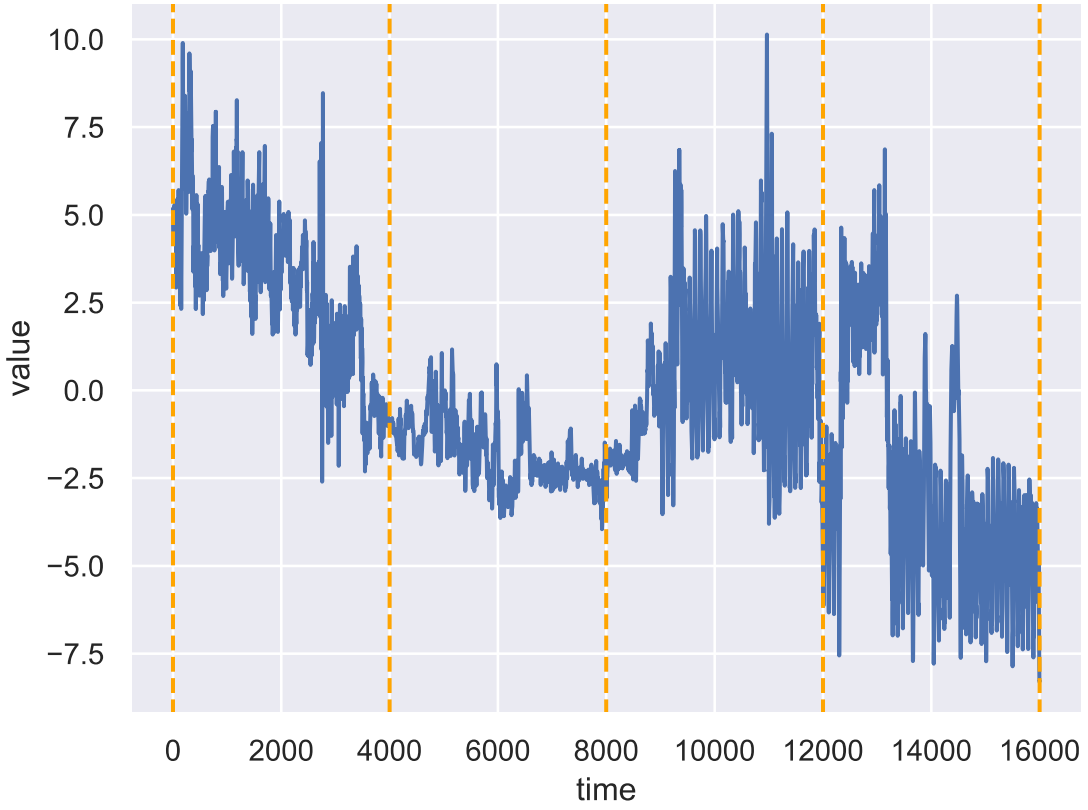
Methodology



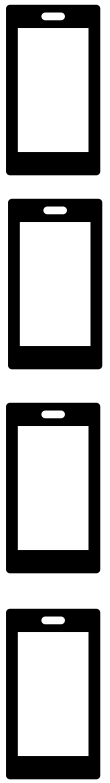
1. Segmentation
2. Orientation Unification
3. Feature Extraction
4. Data Imputation
5. Feature Fusion

1. Segmentation

40-second non-overlapping window



Methodology

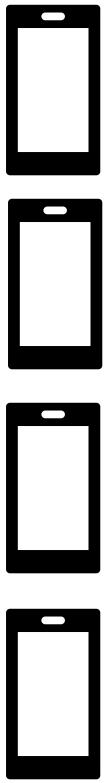


1. Segmentation
2. Orientation Unification
3. Feature Extraction
4. Data Imputation
5. Feature Fusion

2 . Orientation Unification

Madgwick's technique

Methodology



1. Segmentation
2. Orientation Unification
3. Feature Extraction
4. Data Imputation
5. Feature Fusion

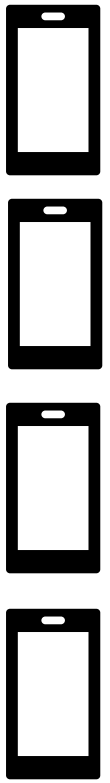
3. Feature Extraction

Features:

- Time-domain (T)
- Frequency-domain (F)
  Location-specific (L)

Name	Formula	Type
Mean	$\overline{q}$	T
Std	$\sigma(q)$	T
Min	$\min(q)$	T
Max	$\max(q)$	T
Median	$median(q)$	T
25th Percentile	$p_{25}(q)$	T
75th Percentile	$p_{75}(q)$	T
Diff min max	$\max(q) - \min(q)$	T
Cross mean	$\sum_{t=1}^{n} \mathbb{I}_{(q_t \geq \overline{q}) \neq (q_{t-1} \geq \overline{q})}$	T
Std $f'(q)$	$\sigma(\overline{\nabla_t(q)})$	T
Mean $f'(q)$	$\overline{\nabla_t(q)}$	T
Mean $f''(q)$	$\overline{\nabla_t^2(q)}$	T
Std $f''(q)$	$\sigma(\nabla_t^2(q))$	T
Max abs	$\max_{t \in \{0,\ldots,n\}} \lvert q_t \rvert$	T
Variance	$\sigma^2(q)$	T
Energy	$\sum_{t=0}^{n} {q_t}^2$	T
Entropy	$H(\lvert u \rvert)$	F
Spectral energy	$\sum \lvert u \rvert^2$	F
1st peak value	$\max u$	F
1st peak	$\arg\max u$	F
2nd peak value	$\max_2 u$	F
2nd peak	$\arg\max_2 u$	F
3rd peak value	$\max_3 u$	F
3rd peak	$\arg\max_3 u$	F
Mean d	$\overline{d(w_t, w_{t-1}),\ t \in \{0,\ldots,n\}}$	L
Std d	$\overline{\sigma(d(w_t, w_{t-1})),\ t \in \{0,\ldots,n\}}$	L
Mean d with A	$\overline{\sqrt{d(w_t, w_{t-1})^2 + d_A(w_t, w_{t-1})^2}}$	L
Std d with A	$\sigma\left(\sqrt{d(w_t, w_{t-1})^2 + d_A(w_t, w_{t-1})^2}\right)$	L
Mean s	$\overline{s(w_t, w_{t-1})}$	L
Std s	$\sigma(s(w_t, w_{t-1}))$	L
Mean s with A	$\overline{s_A(w_t, w_{t-1})}$	L
Std s with A	$\sigma(s_A(w_t, w_{t-1}))$	L

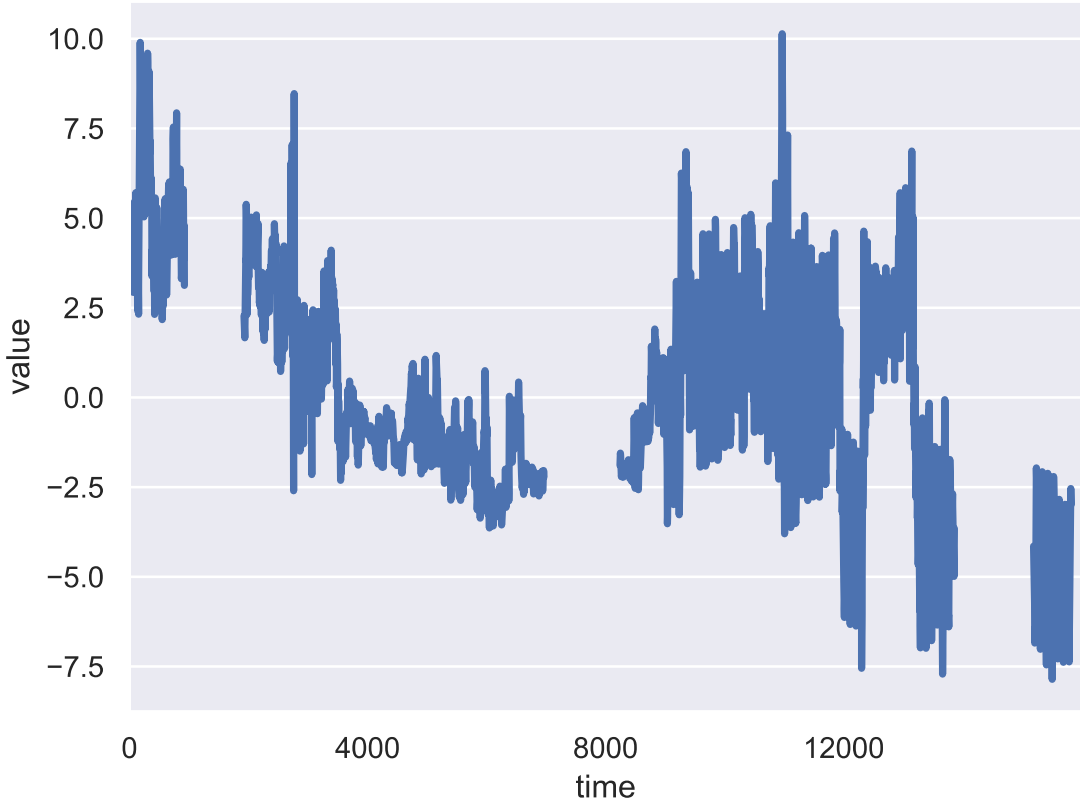
Methodology



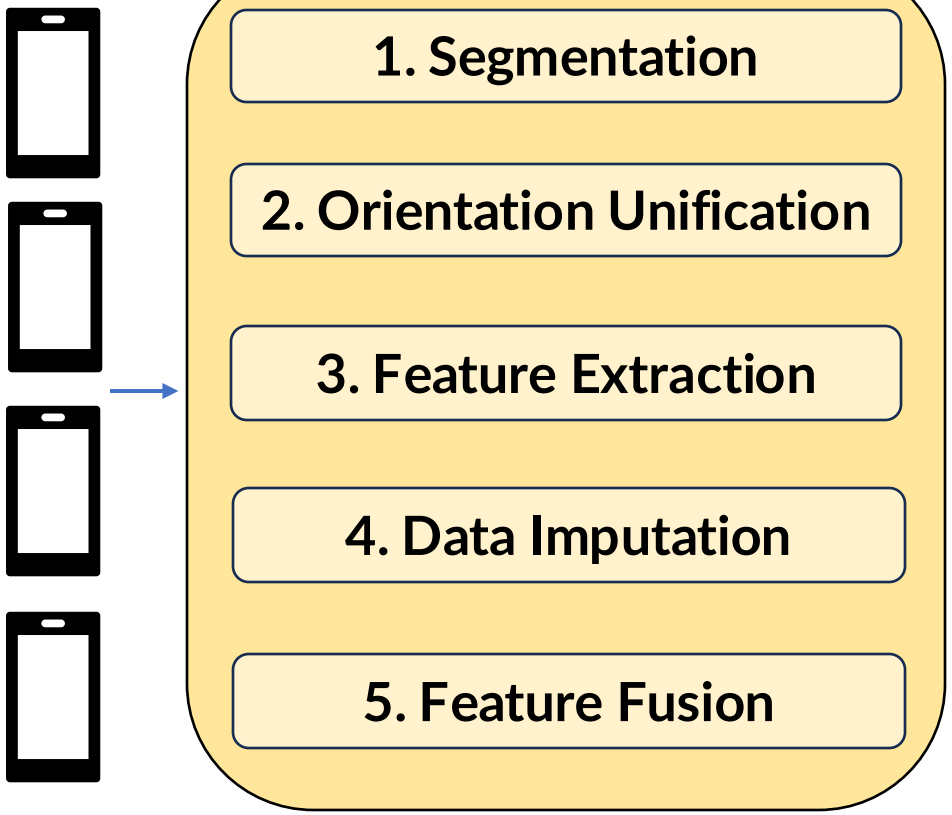
1. Segmentation
2. Orientation Unification
3. Feature Extraction
4. Data Imputation
5. Feature Fusion

4. Data Imputation

Missing Not At Random (MNAR) → impute data with impossible values

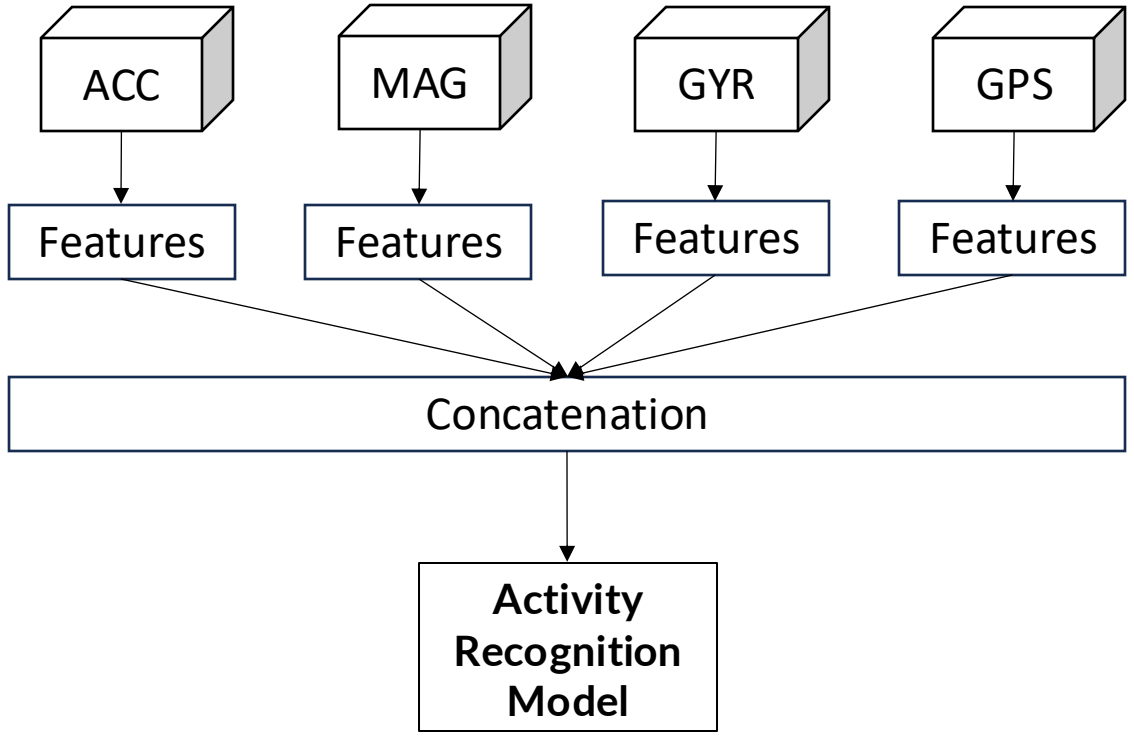


Methodology



5. Feature Fusion

Intermediate feature fusion

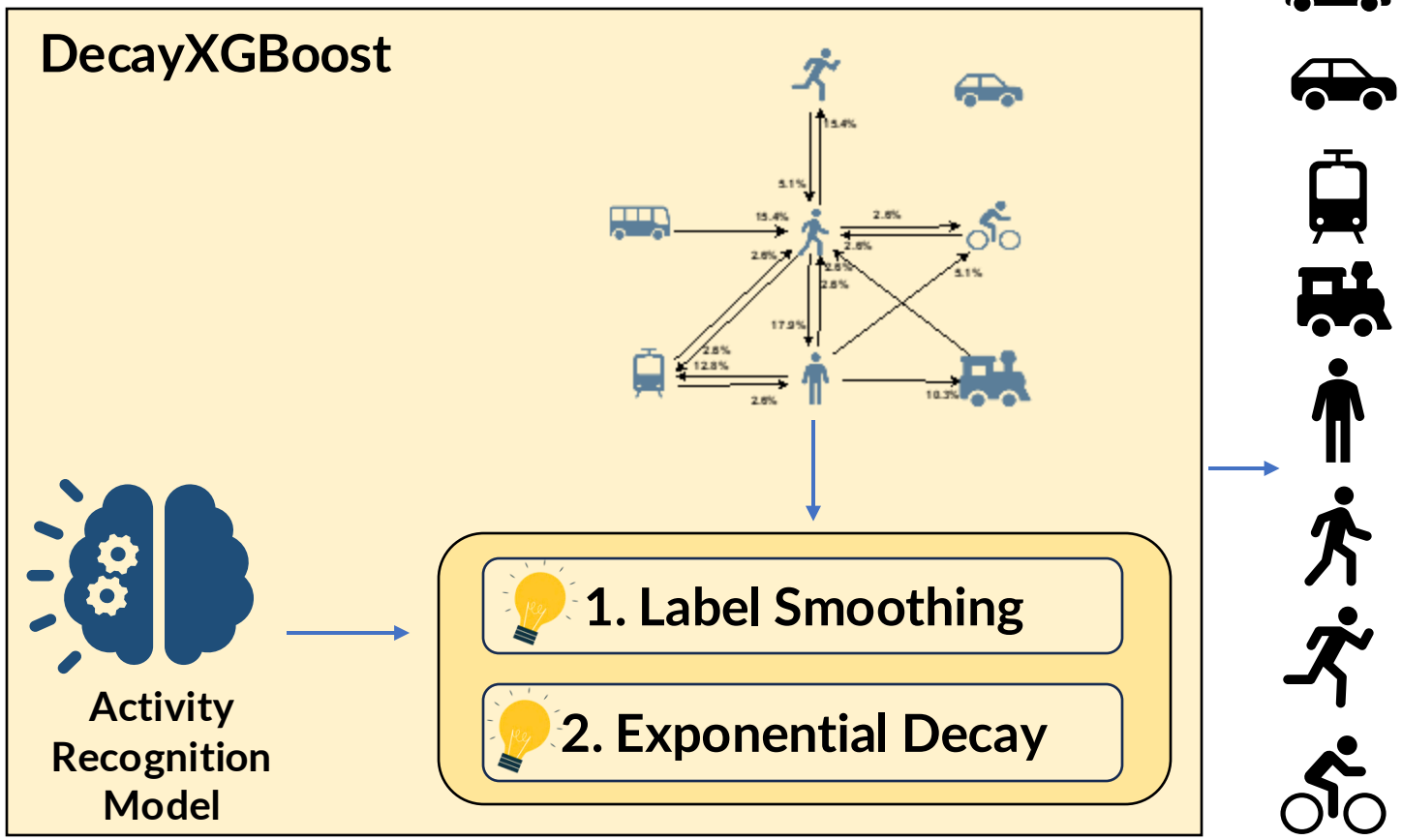


Methodology

1. Label Smoothing

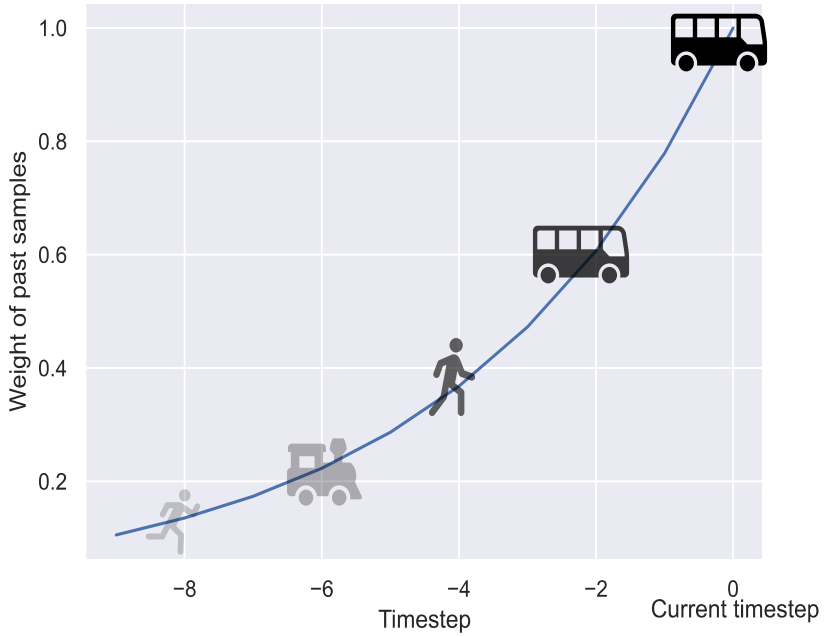
- Correct spurious predictions
- Changing vehicle required

  non-vehicle activity

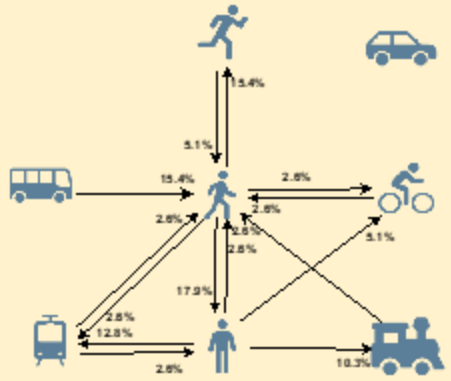


Methodology

2. Exponential Decay



DecayXGBoost

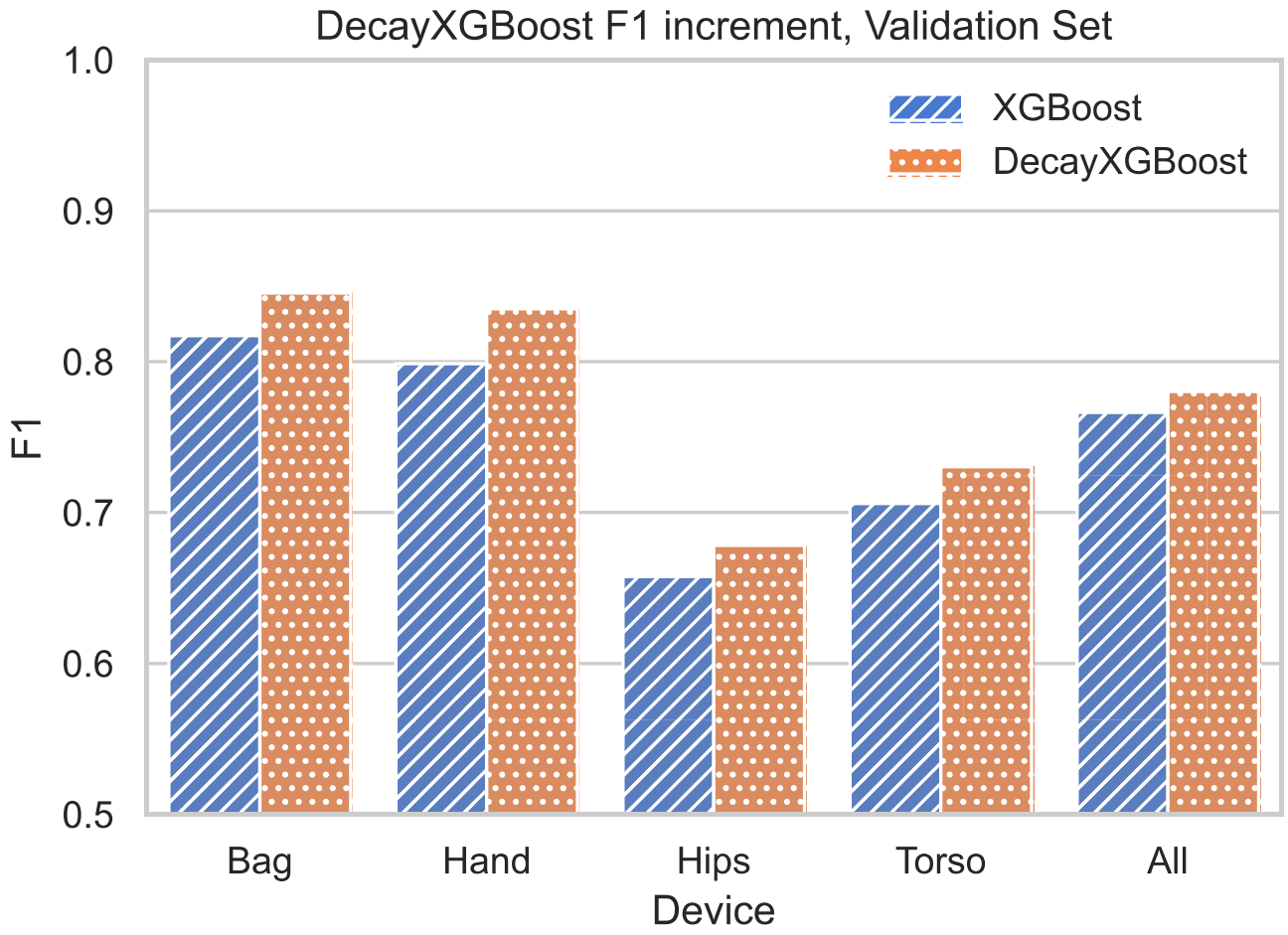


Activity Recognition Model

1. Label Smoothing

2. Exponential Decay

Results


DecayXGBoost F1 increment, Validation Set
XGBoost
DecayXGBoost
1.0
0.9
0.8
0.7
0.6
0.5
F1
Bag
Hand
Hips
Torso
All
Device


Thank you!