IEEE Biometrics Council | IEEE | IEEE International Joint Conference on Biometrics | SIAT

A Key Feature Screening Method for Human Activity Recognition Based on Multi-head Attention Mechanism

Hao Wang, Fangyu Liu, Xiang Li, Ye Li, Fangmin Sun* (fm.sun@siat.ac.cn)

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China

Problem & Motivation

- HAR features are high-dimensional and redundant; edge deployment constrained.
- Feature-level interpretability is often missing in existing pipelines.
- Need to know which features matter while keeping accuracy.

We aim to deliver a compact, high-accuracy, and feature-level interpretable pipeline.

Dataset & Features

KU-HAR dataset[1]:

- 90 subjects, 18 activities;
- Waist 6-axes IMU (accelerometer + gyroscope);
- 20,750 samples (non-overlapping 3-second windows).

Feature:

We use TSFEL[2] to extract a total of 156 features from each of the 6 IMU channels, results in a total of **936 features** per sample (156×6=936).

Proposed Method

Overall Workflow

Input & Feature

- Use KU-HAR dataset and extract TSFEL features
- 156 features per axis; 936 features per sliding window.

Channel-wise Projection

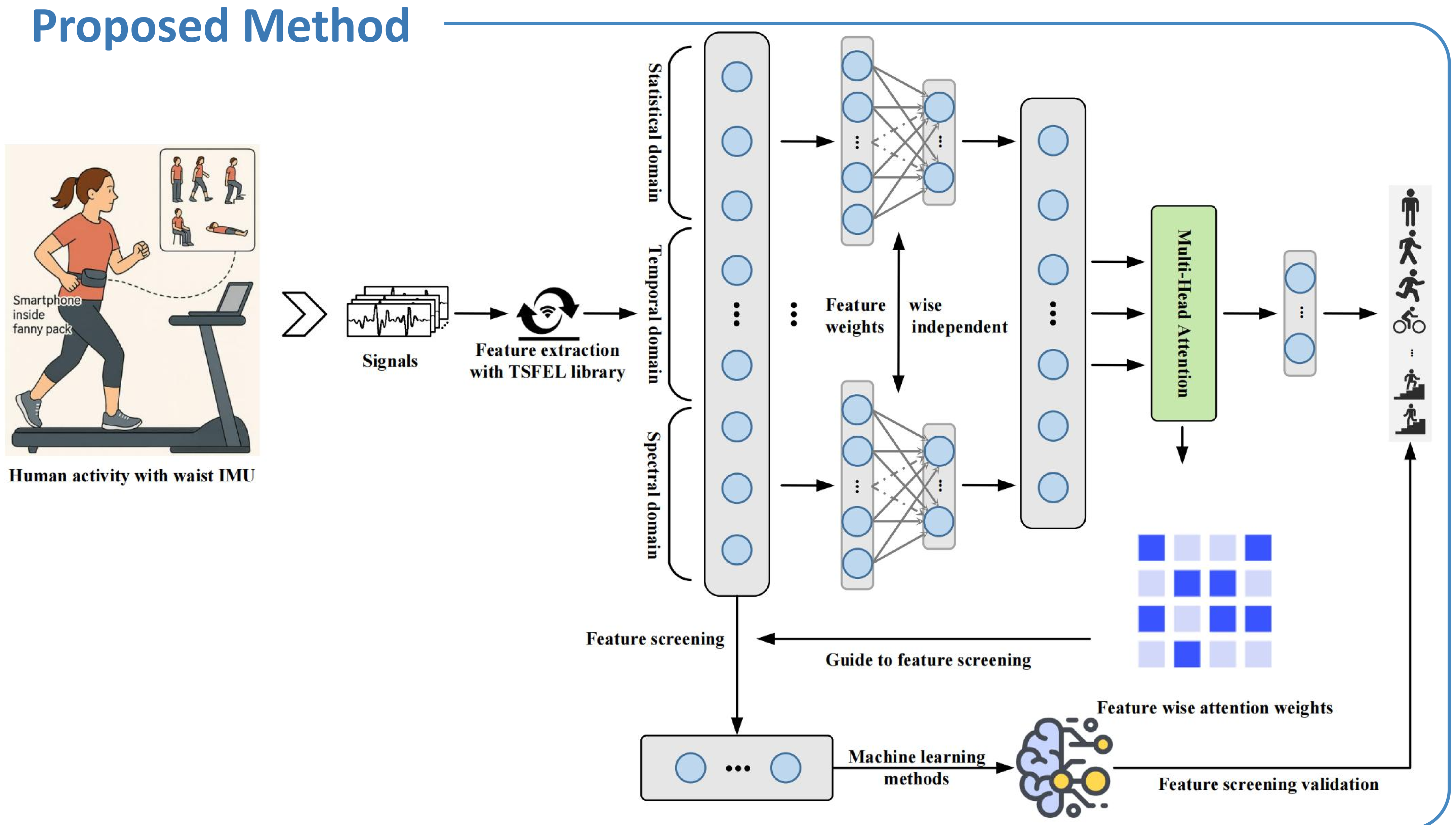
- Apply weight-independent linear layers to extract features from each feature and generate feature embeddings.

Attention Weighting & Screening

- Use multi-head self-attention to score features; average scores across heads and folds, then select per-axis Top-10.

Classification

- Train machine learning classifier on the screened features to achieve high accuracy with low compute at inference.



Experimental Result

Top-10 important features.

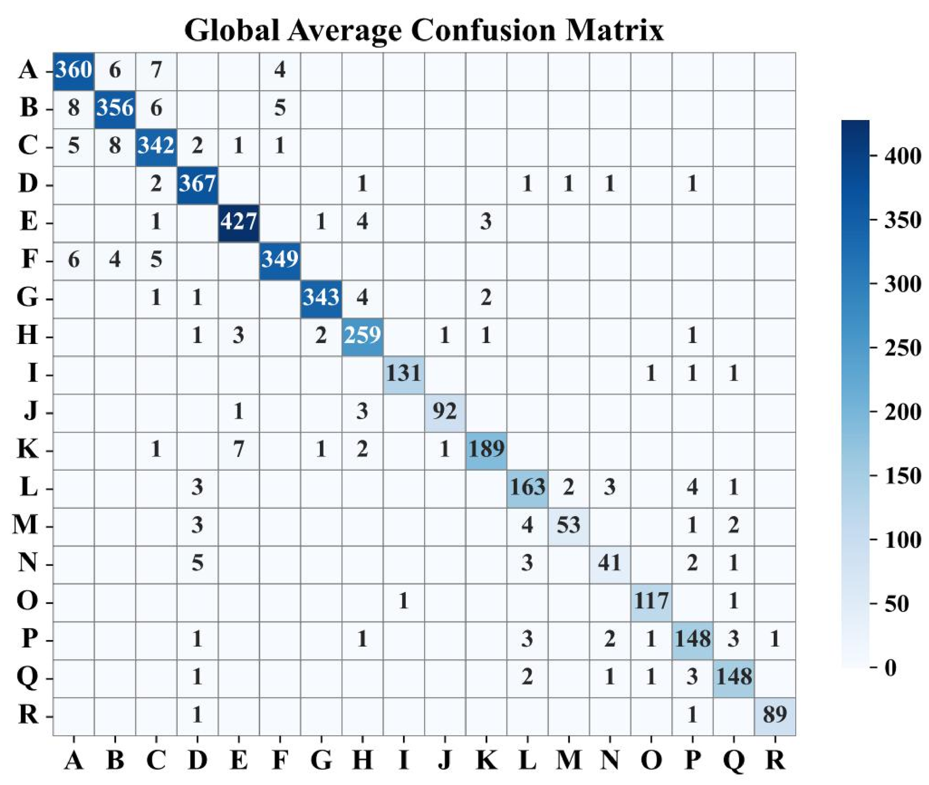
Feature name	Domain	Description	Formula
MFCC_9	Spectral	Mel-scale frequency cepstral coefficients	As in paper [3]
Spectral distance	Spectral	The signal spectral distance	$\sum_{i=0}^{N} lr_{fmag_i} - cumsum_{fmag_i}$
Positive turning points	Temporal	Number of positive turning points of the signal	$\sum_{i=0}^{N-2} \mathbf{1}(\frac{ds_i}{dt} > 0 \wedge \frac{ds_{i+1}}{dt} < 0)$
Maximum frequency	Spectral	Maximum frequency of the signal	$freq[min\{i \mid cumsum_{fmag_i} \geq 0.95 \cdot cumsum_{fmag_{max}}\}]$
ECDF Percentile Count_1	Statistical	The cumulative sum of samples that are less than the percentile	$\sum_{i=0}^{N} \mathbf{1}(ECDF\ values(s_i) < p)$
Signal distance	Temporal	Signal traveled distance	$\sum_{i=0}^{N-1} \sqrt{1 + \Delta s_i^2}$
Spectral positive turning points	Spectral	The number of positive turning points of the fft magnitude signal	$\sum_{i=0}^{N-2} \mathbf{1}(\frac{dfmag_i}{dfreq_i} > 0 \wedge \frac{dfmag_{i+1}}{dfreq_{i+1}} < 0)$
Negative turning points	Temporal	Number of negative turning points of the signal	$\sum_{i=0}^{N-2} \mathbf{1}(\frac{ds_i}{dt} < 0 \wedge \frac{ds_{i+1}}{dt} > 0)$
Power bandwidth	Spectral	Power spectrum density bandwidth of the signal	$\lvert max\{freq \mid C(freq) \leq 0.95 \cdot C(freq_{max})\} - min\{freq \mid C(freq) \geq 0.95 \cdot C(freq_{max})\}\rvert$
Zero crossing rate	Temporal	Zero-crossing rate of the signal	$\sum_{i=0}^{N-1} \mathbf{1}(sign(s_i) \neq sign(s_{i+1}))$

Comparison against representative baselines.

Year	Method	Data	Human activity recognition							
			KU-HAR Dataset: 20,750 samples from 90 subjects (75 Male / 15 Female)							
			ACC	PRE	REC	F1	MCC	AUC	FLOPs	Params
2021	DenseNet-GRU	Waist IMU	0.89±0.01	0.89±0.01	0.89±0.01	0.89±0.01	0.88±0.01	0.97±0.00	54.53M	1.31M
2022	CNN	Waist IMU	0.83±0.02	0.84±0.01	0.83±0.02	0.82±0.02	0.82±0.02	0.98±0.00	3.28M	1.19M
2022	ResRNN	Waist IMU	0.76±0.01	0.76±0.06	0.76±0.01	0.71±0.02	0.76±0.01	0.90±0.02	17.19M	1.29M
2023	ResNet-BiGRU-SE	Waist IMU	0.89±0.01	0.90±0.01	0.89±0.01	0.89±0.01	0.89±0.01	0.99±0.00	0.08G	4.06M
2024	CNN-LSTM	Waist IMU	0.80±0.01	0.82±0.02	0.80±0.01	0.80±0.01	0.79±0.01	0.97±0.00	7.05M	1.85M
2024	Multi-STMT	Waist IMU	0.85±0.01	0.87±0.02	0.85±0.01	0.85±0.01	0.84±0.01	0.98±0.01	47.70M	5.35M
Ours	Linear+Attention	All Features	**0.93±0.01**	**0.93±0.01**	**0.93±0.01**	**0.93±0.01**	**0.93±0.01**	0.90±0.02	**1.17M**	**0.79M**
Ours	LR	Select Features	0.81±0.00	0.81±0.00	0.81±0.00	0.81±0.00	0.80±0.00	0.99±0.00	1.05K	**1.07K**
Ours	DT	Select Features	0.83±0.00	0.83±0.00	0.83±0.00	0.83±0.00	0.82±0.00	0.90±0.00	3.00K	5.99K
Ours	KNN	Select Features	0.78±0.00	0.78±0.00	0.78±0.00	0.77±0.01	0.76±0.01	0.96±0.00	**0.59K**	0.95M
Ours	RF	Select Features	0.93±0.00	0.93±0.00	0.93±0.00	0.93±0.00	0.92±0.00	1.00±0.00	0.33M	0.66M
Ours	SVM	Select Features	0.85±0.01	0.85±0.01	0.85±0.01	0.85±0.01	0.84±0.01	0.99±0.00	0.56M	0.56M
Ours	GB	Select Features	0.93±0.00	0.93±0.00	0.93±0.00	0.93±0.00	0.92±0.00	1.00±0.00	1.45K	2.90K
Ours	LightGBM	Select Features	**0.96±0.00**	**0.96±0.00**	**0.96±0.00**	**0.96±0.00**	**0.95±0.00**	**1.00±0.00**	0.26M	0.51M

Tablenotes: FLOPs is the computational cost per sample inference

Confusion matrix and class-wise performace.



Class	PRE	REC	F1
Stand	0.952	0.954	0.953
Sit	0.955	0.950	0.952
Talk-sit	0.937	0.952	0.944
Talk-stand	0.955	0.982	0.969
Stand-sit	0.973	0.981	0.977
Lay	0.970	0.962	0.966
Lay-stand	0.986	0.973	0.979
Pick	0.947	0.971	0.959
Jump	0.991	0.980	0.986
Push-up	0.977	0.958	0.967
Sit-up	0.973	0.940	0.956
Walk	0.924	0.926	0.925
Walk-backward	0.947	0.842	0.891
Walk-circle	0.862	0.792	0.824
Run	0.973	0.982	0.977
Stair-up	0.921	0.929	0.924
Stair-down	0.940	0.945	0.943
Table-tennis	0.980	0.972	0.976
Avg=‘Weighted’	0.958	0.957	0.957

Conclusion

- We propose an attention-guided feature screening framework for wearable HAR.
- Our method combines independent channel-wise linear transformations with attention-guided feature selection, producing a compact and highly informative feature set that enhances both classification performance and interpretability.
- Coupled with a lightweight LightGBM classifier, the screened features reach 96.0% accuracy on KU-HAR while drastically reducing compute and memory compared with deep baselines.

Reference

[1] Ku-har: An open dataset for heterogeneous human activity recognition, Pattern Recognition Letters, vol. 146, pp. 46–54, 2021.

[2] Tsfel: Time series feature extraction library, SoftwareX, vol. 11, p. 100456, 2020.

[3] Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, IEEE transactions on acoustics, speech, and signal processing, vol. 28, no. 4, pp. 357–366, 1980.