

HAR

(Human Activity Recognition)



(C01 – C09)

Dataset Description

- Number of attributes:
 - 3 (X, Y, Z) - Accelerometer
 - 3 (X, Y, Z) – Gyroscope
- Missing attribute values:
 - None
- Sampling Rate :
 - 100 Hz

Sample_Train.csv

3000 Rows ++

Acc_X	Acc_Y	Acc_Z	Gyro_X	Gyro_Y	Gyro_Z
0.58507	0.958694	-0.91408	0.508786	0.003394	0.024296
0.554663	1.122004	-0.84205	0.543408	0.002409	0.034015
0.527227	1.208196	-0.75273	0.531366	-0.00668	0.071343
...
...
...

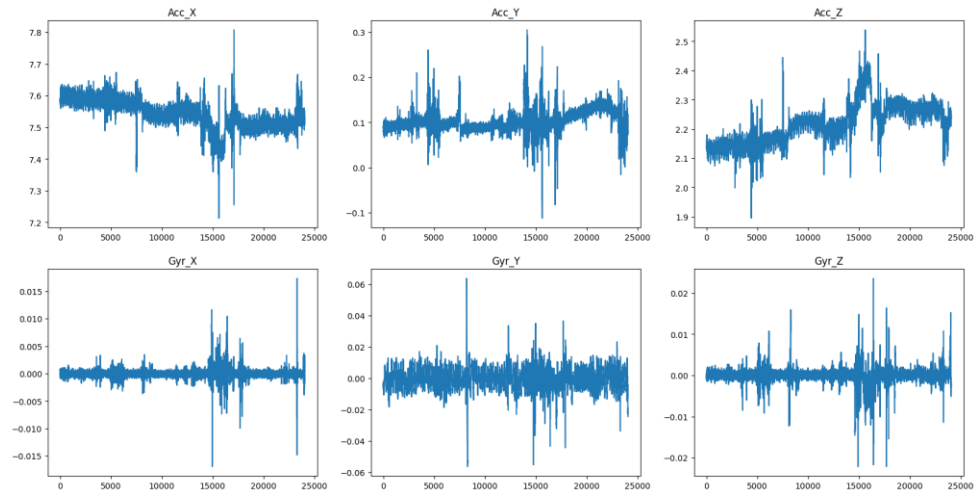
Sample_Test.csv

500 Rows

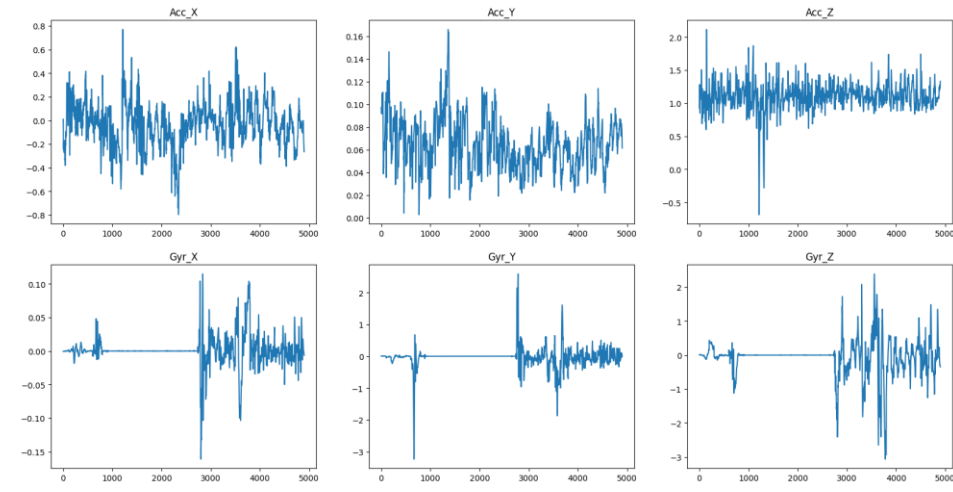
Acc_X	Acc_Y	Acc_Z	Gyro_X	Gyro_Y	Gyro_Z
0.58507	0.958694	-0.91408	0.508786	0.003394	0.024296
0.554663	1.122004	-0.84205	0.543408	0.002409	0.034015
0.527227	1.208196	-0.75273	0.531366	-0.00668	0.071343
...
...
...

EDA

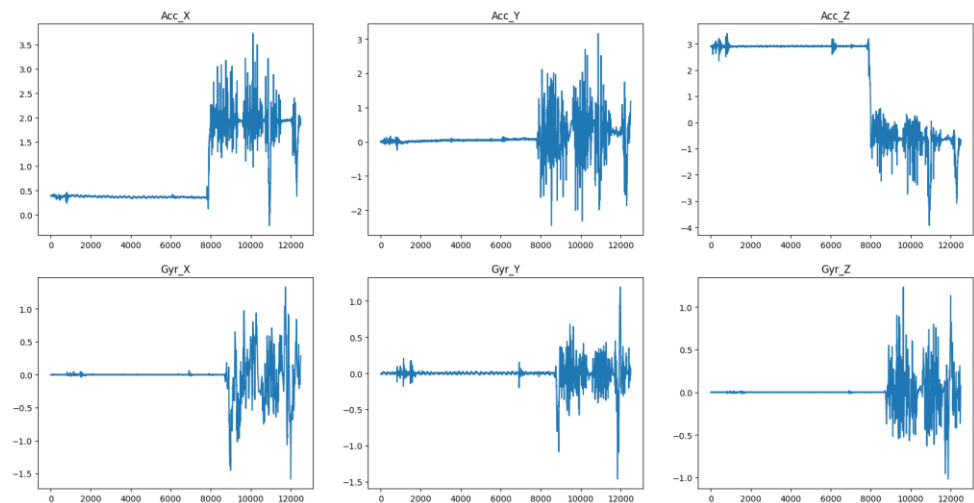
C01



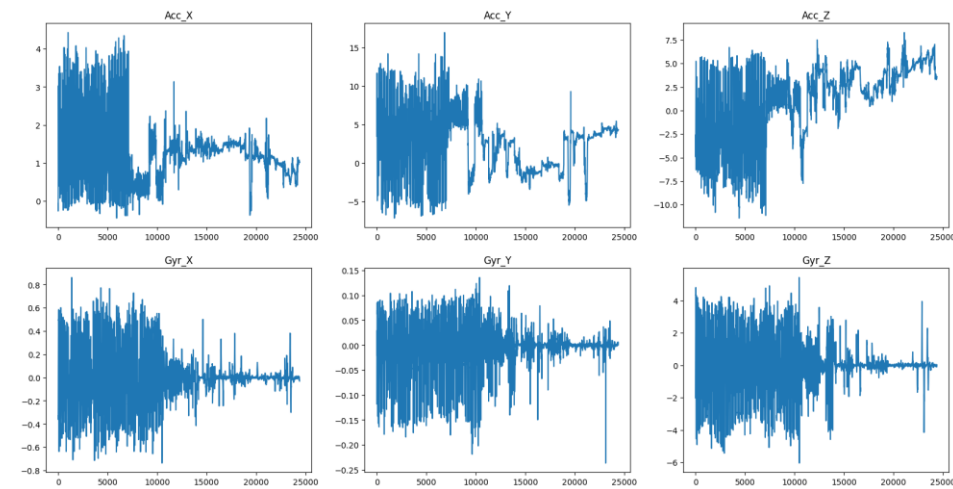
C02



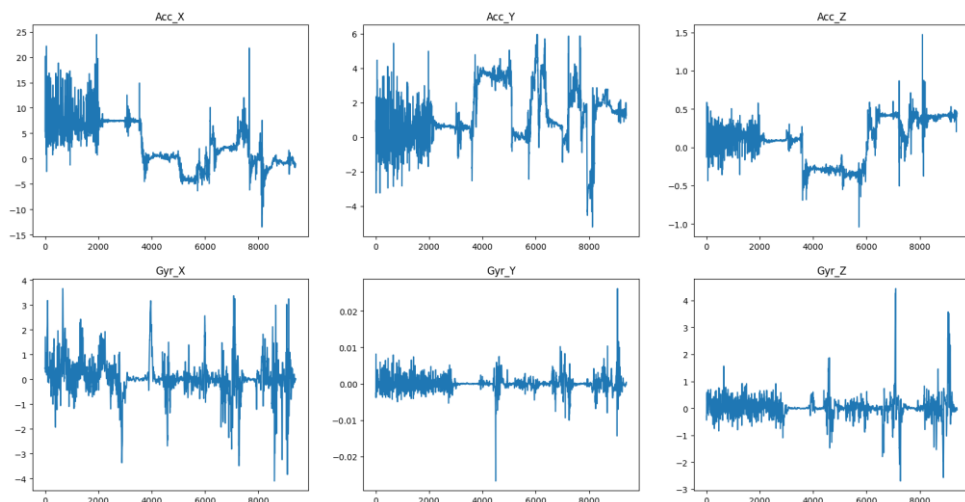
C03



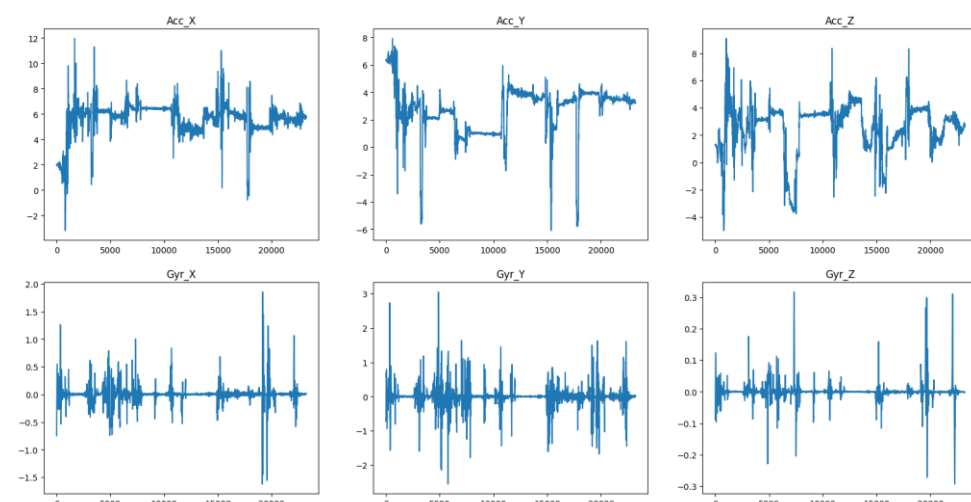
C04



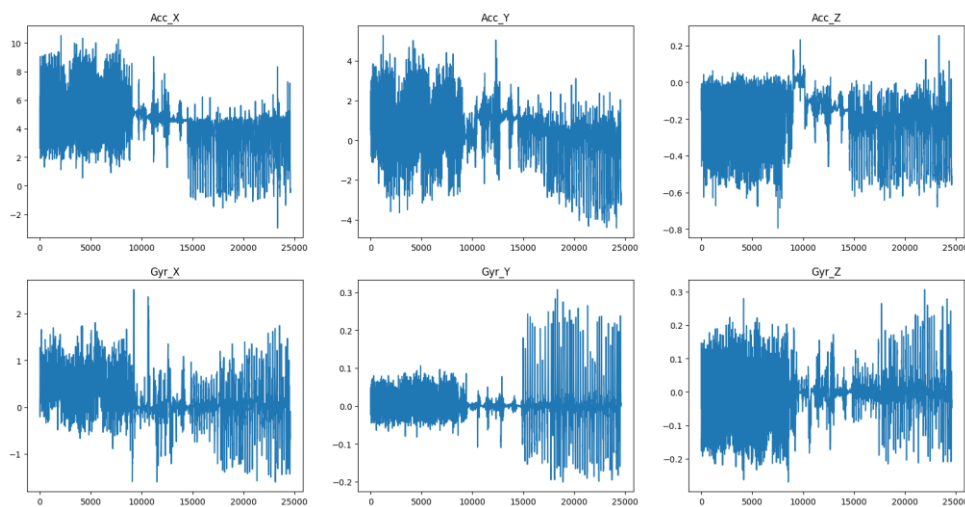
C05



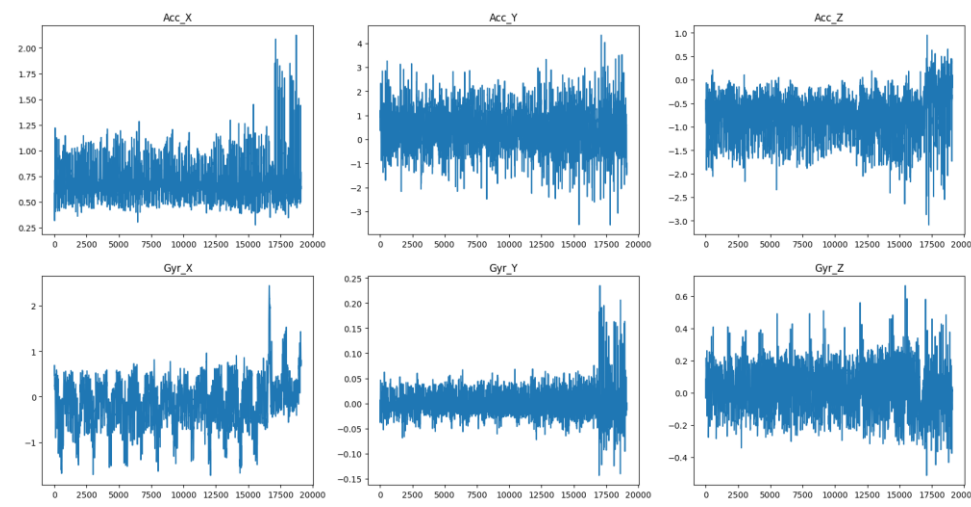
C06



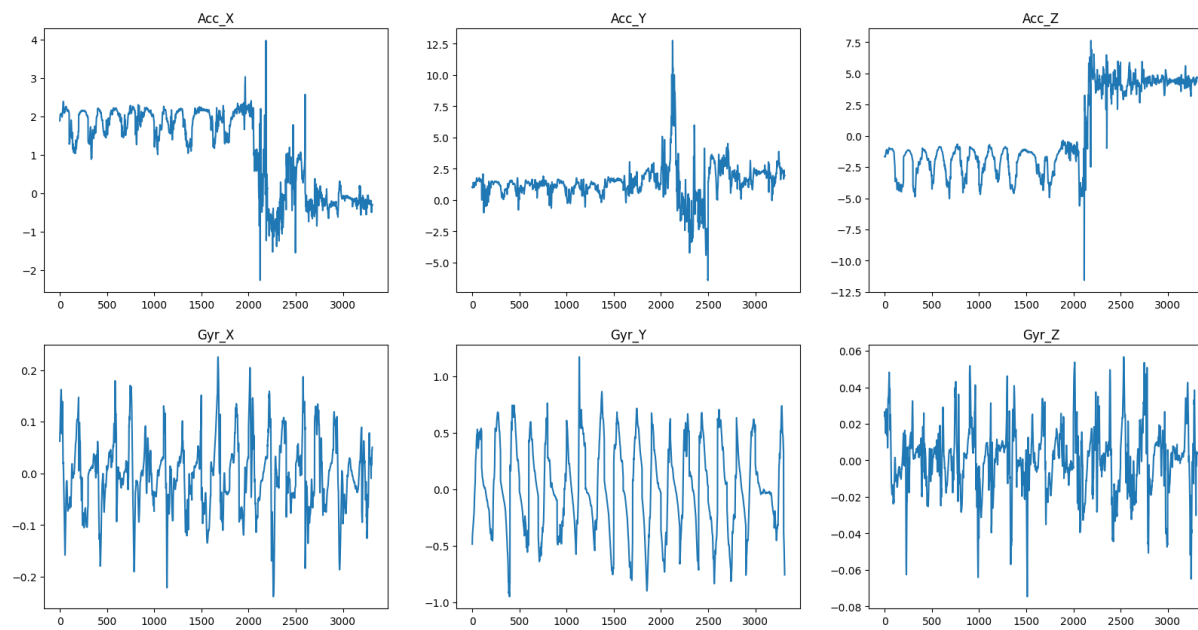
C07

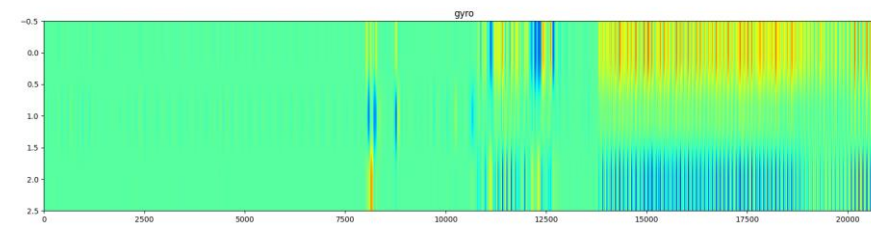
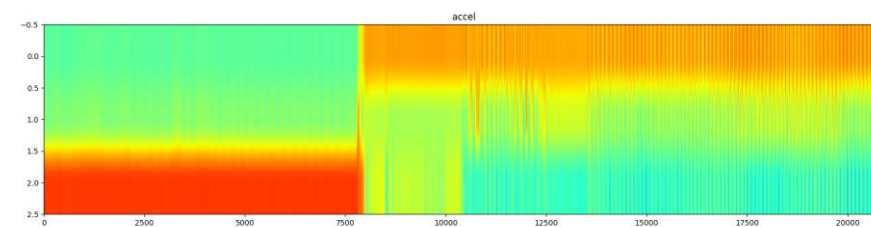
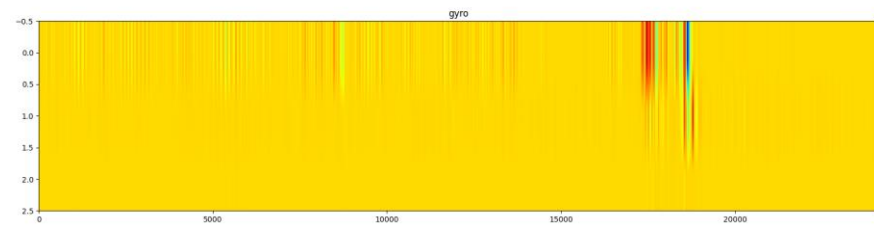
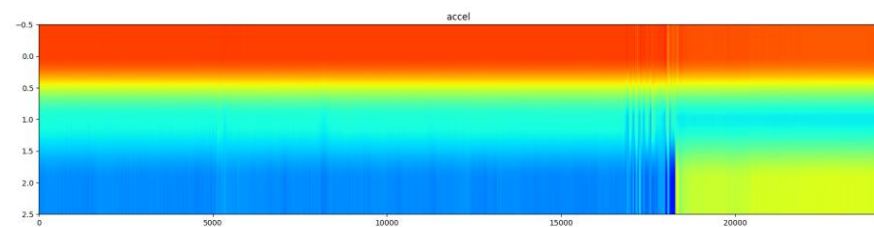
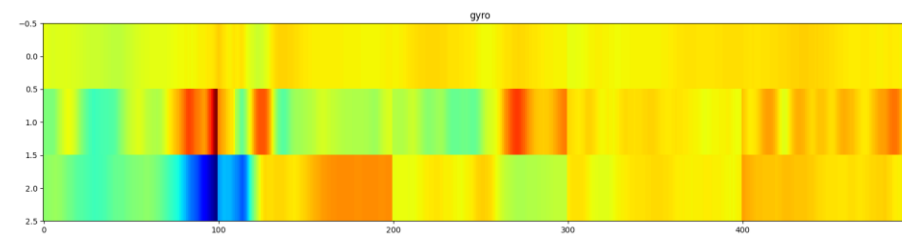
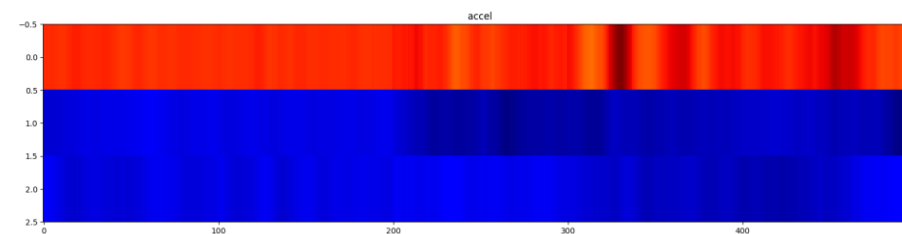
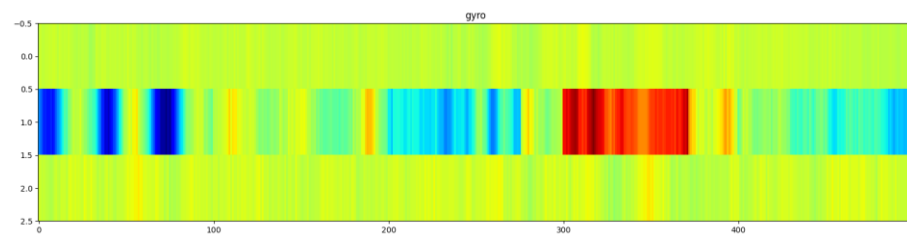
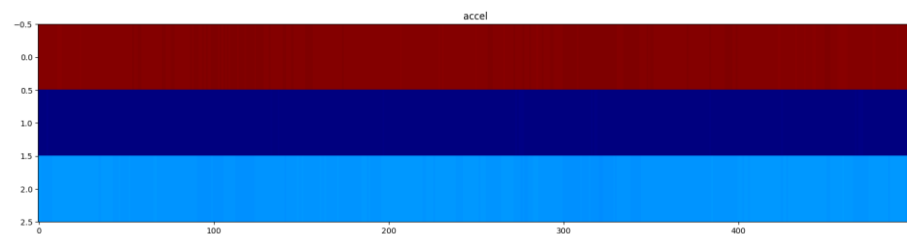


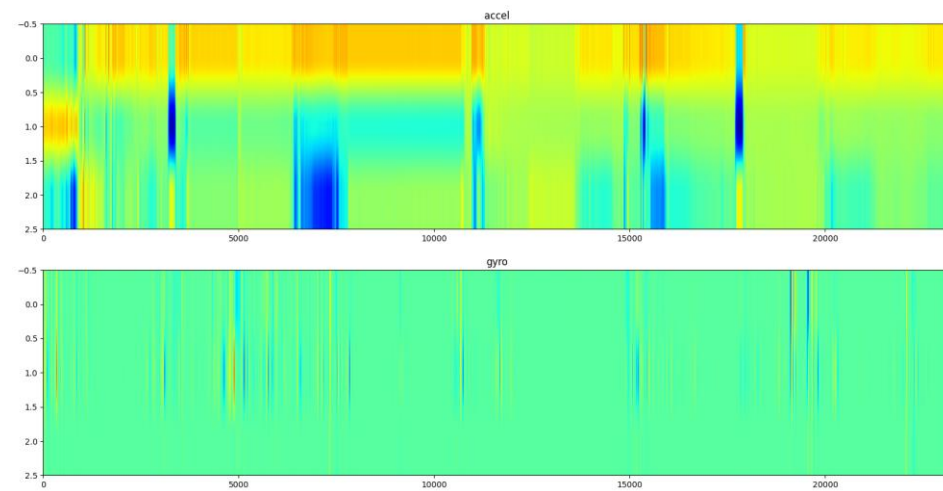
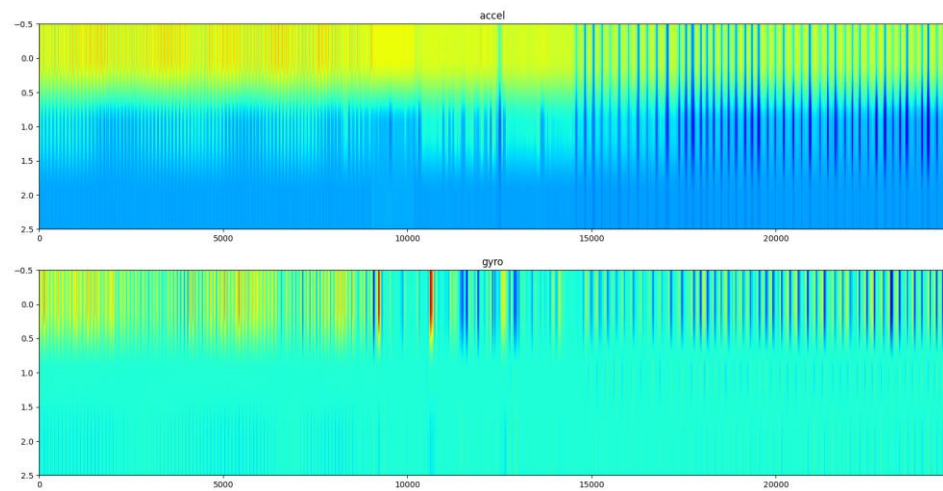
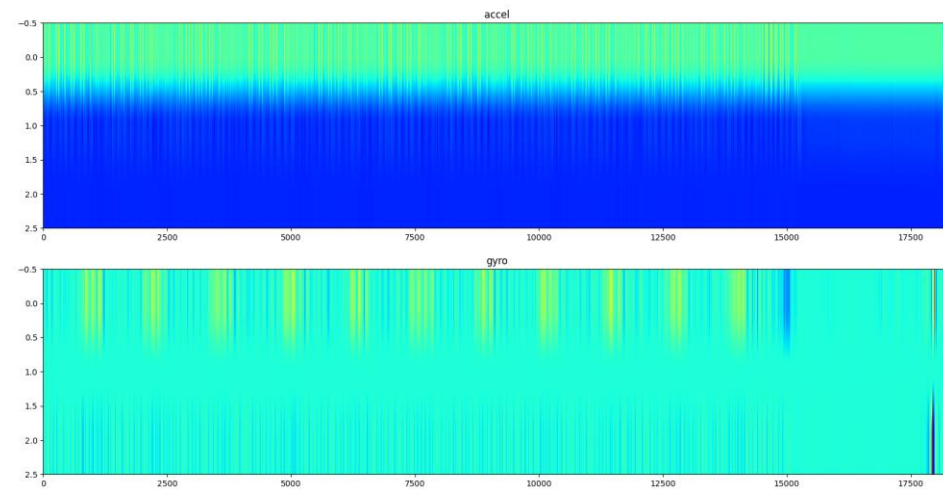
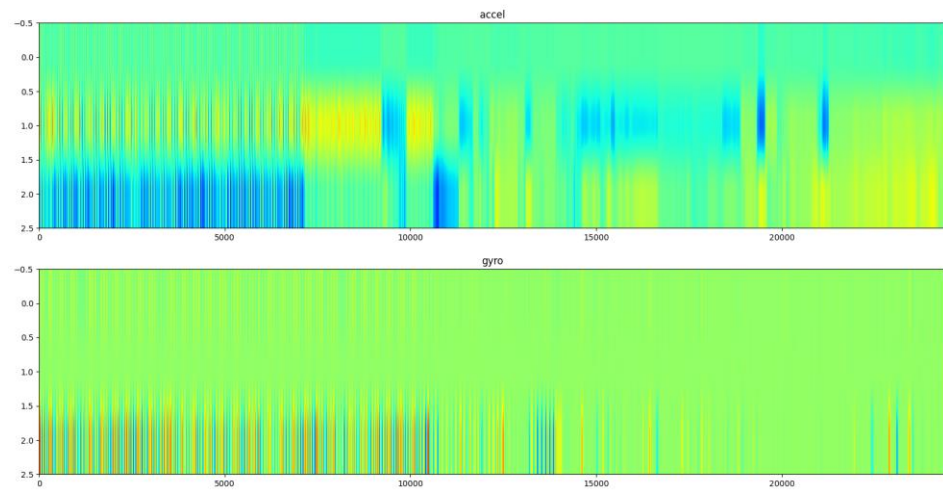
C08

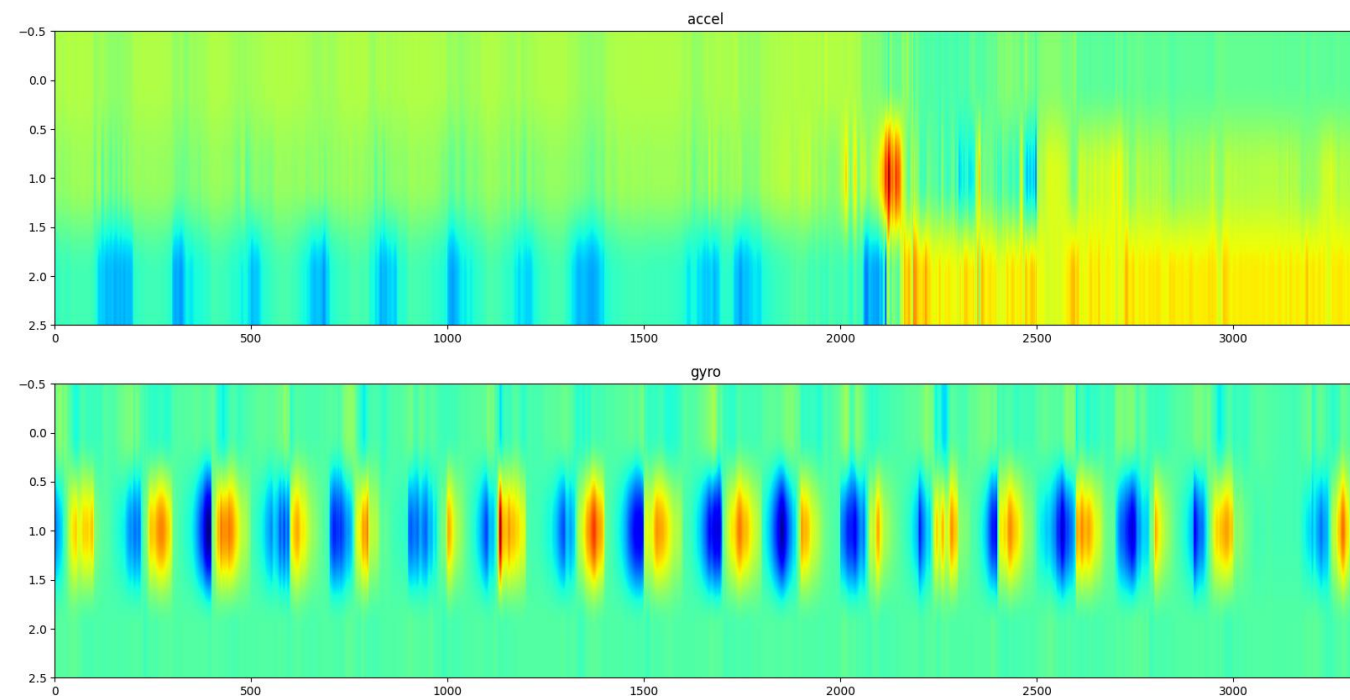


C09









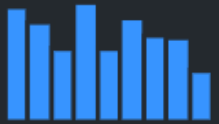






Pipeline #1

Lightgbm+caret experiment

Add Features

- Mean
- Variance
- Lagged+Downsample
- Multiple predictions per cell

Lagged Features

# index	# 0	# Acc_X1	# Acc_X2	# Acc_X3	# Acc_X4	# Acc_X5	# Acc_X6	...
	n 647203 (>99%) Distinct: 9 (<1%)	Missing: 1 (<1%) Distinct: 69204 (11%)	Missing: 2 (<1%) Distinct: 69204 (11%)	Missing: 3 (<1%) Distinct: 69204 (11%)	Missing: 4 (<1%) Distinct: 69204 (11%)	Missing: 5 (<1%) Distinct: 69204 (11%)	Missing: 6 (<1%) Distinct: 69204 (11%)	
								
	Min 0.0 Max 8.0	Min -13.4... Max 38.8...	Min -13.46... Max 38.8...	Min -13.4... Max 38.8...	Min -13.4... Max 38.8...	Min -13.4... Max 38.8...	Min -13.46... Max 38.8...	
0	0.0	Missing value	Missing value	Missing value	Missing value	Missing value	Missing value	
1	0.0	7.909174341745075	Missing value	Missing value	Missing value	Missing value	Missing value	
2	0.0	7.921172913243259	7.909174341745075	Missing value	Missing value	Missing value	Missing value	
3	0.0	7.93517194271138	7.921172913243259	7.909174341745075	Missing value	Missing value	Missing value	
4	0.0	7.959169085707749	7.93517194271138	7.921172913243259	7.909174341745075	Missing value	Missing value	
5	0.0	7.969168034500284	7.959169085707749	7.93517194271138	7.921172913243259	7.909174341745075	Missing value	
6	0.0	7.965167953824697	7.969168034500284	7.959169085707749	7.93517194271138	7.921172913243259	7.909174341745075	
7	0.0	7.961168708413398	7.965167953824697	7.969168034500284	7.959169085707749	7.93517194271138	7.921172913243259	
8	0.0	7.943170433533979	7.961168708413398	7.965167953824697	7.969168034500284	7.959169085707749	7.93517194271138	



submission.csv

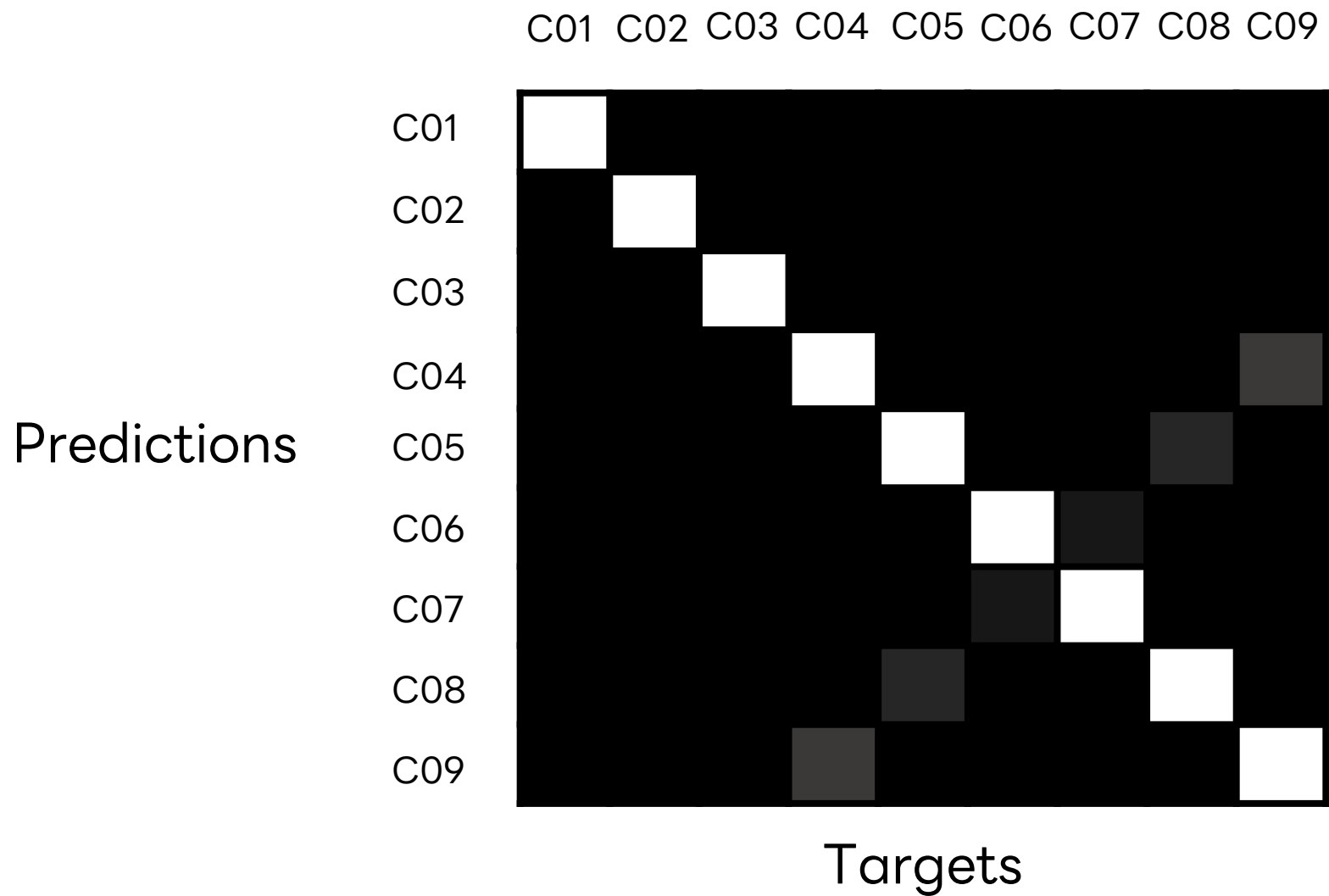
Complete · 402829-ss1ss · 20h ago · Light gbm + 10 lag + mean + var

0.93018

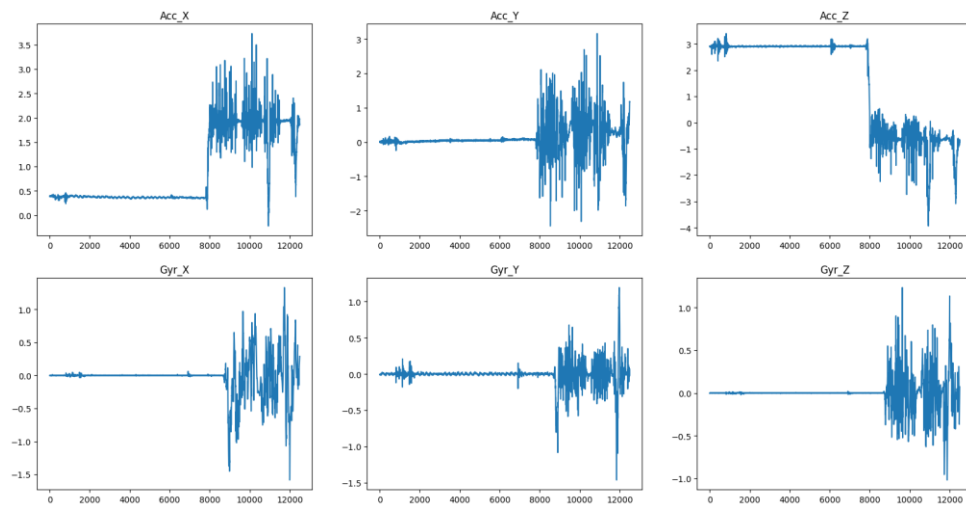
0.92402



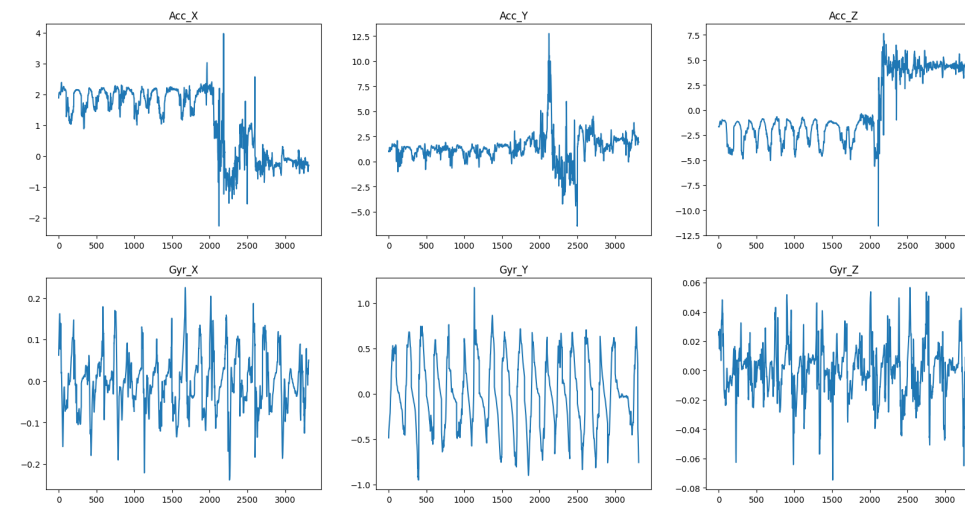
Confusion Matrix



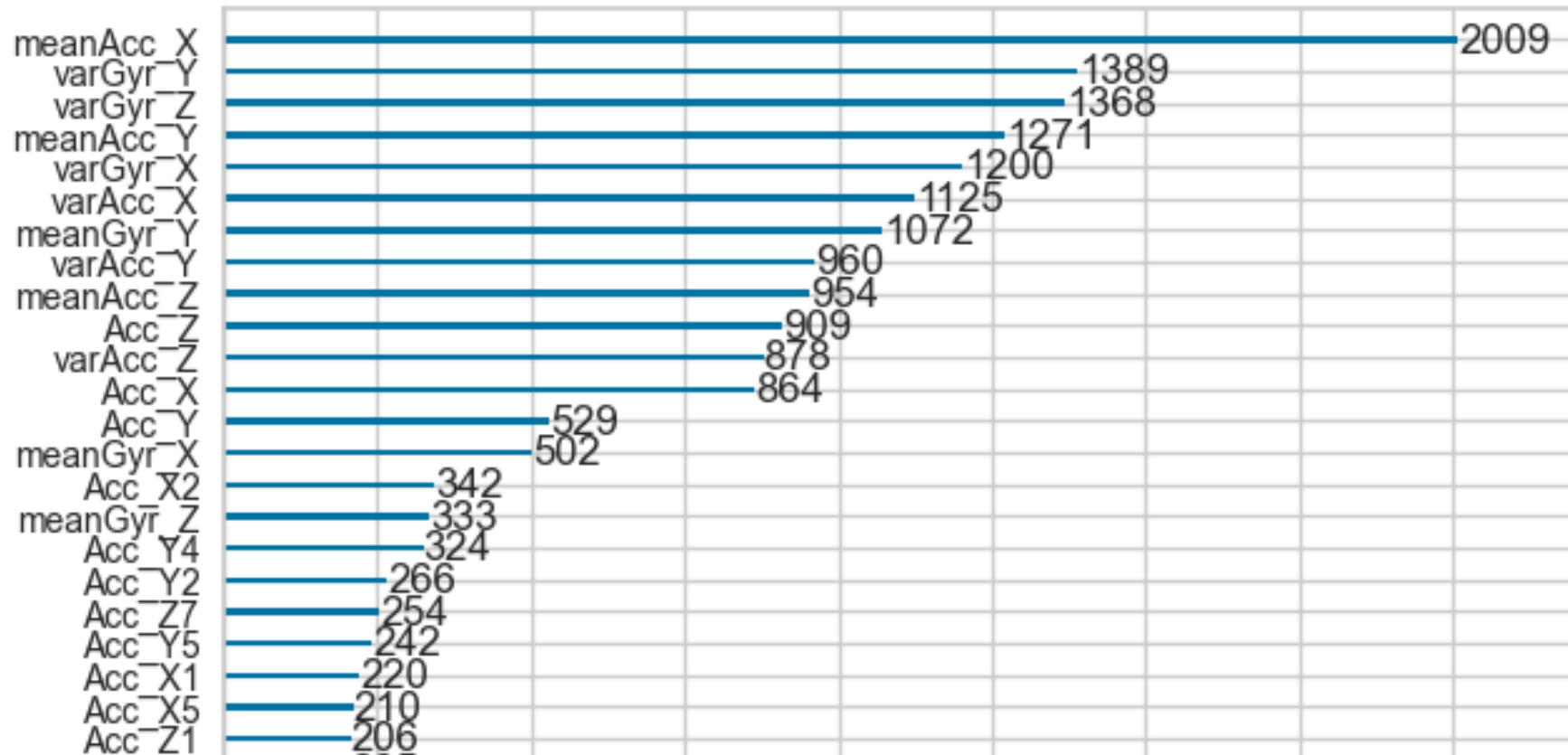
C03



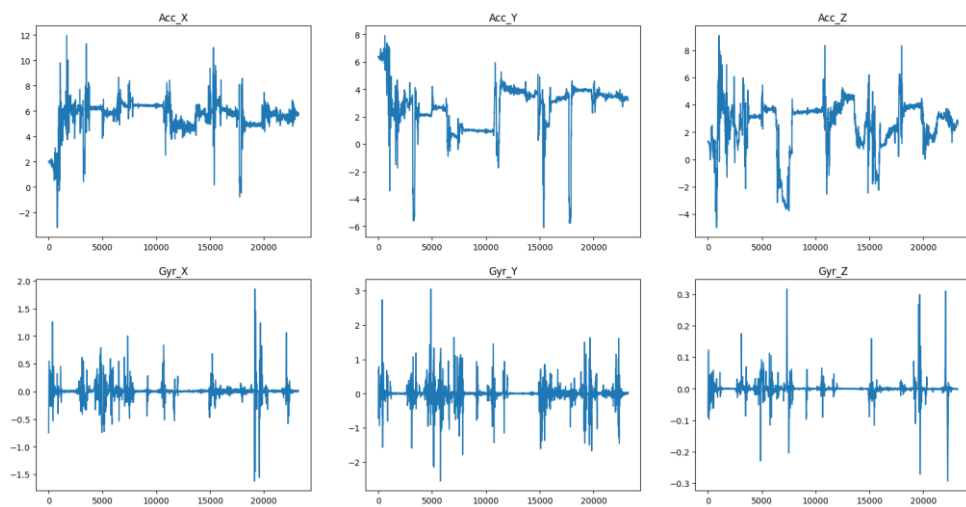
C09



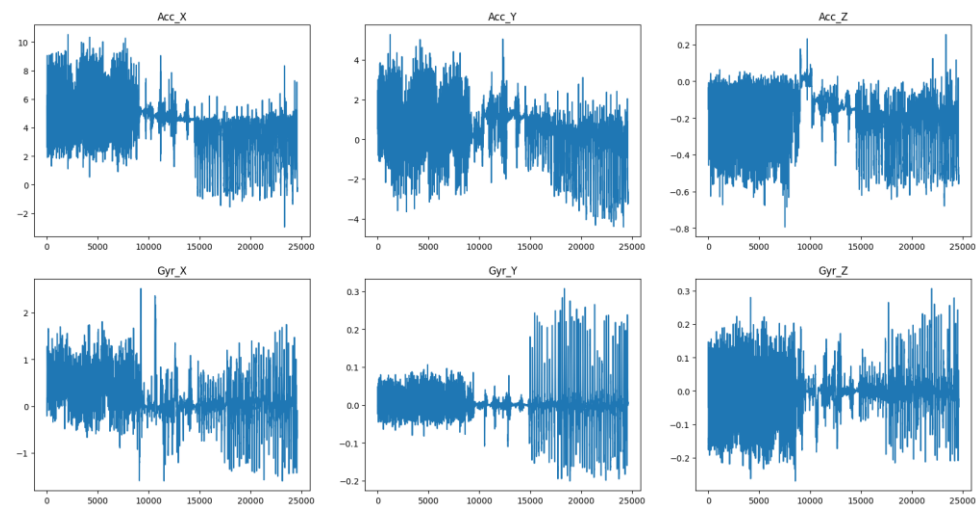
Feature Importance



C06



C07

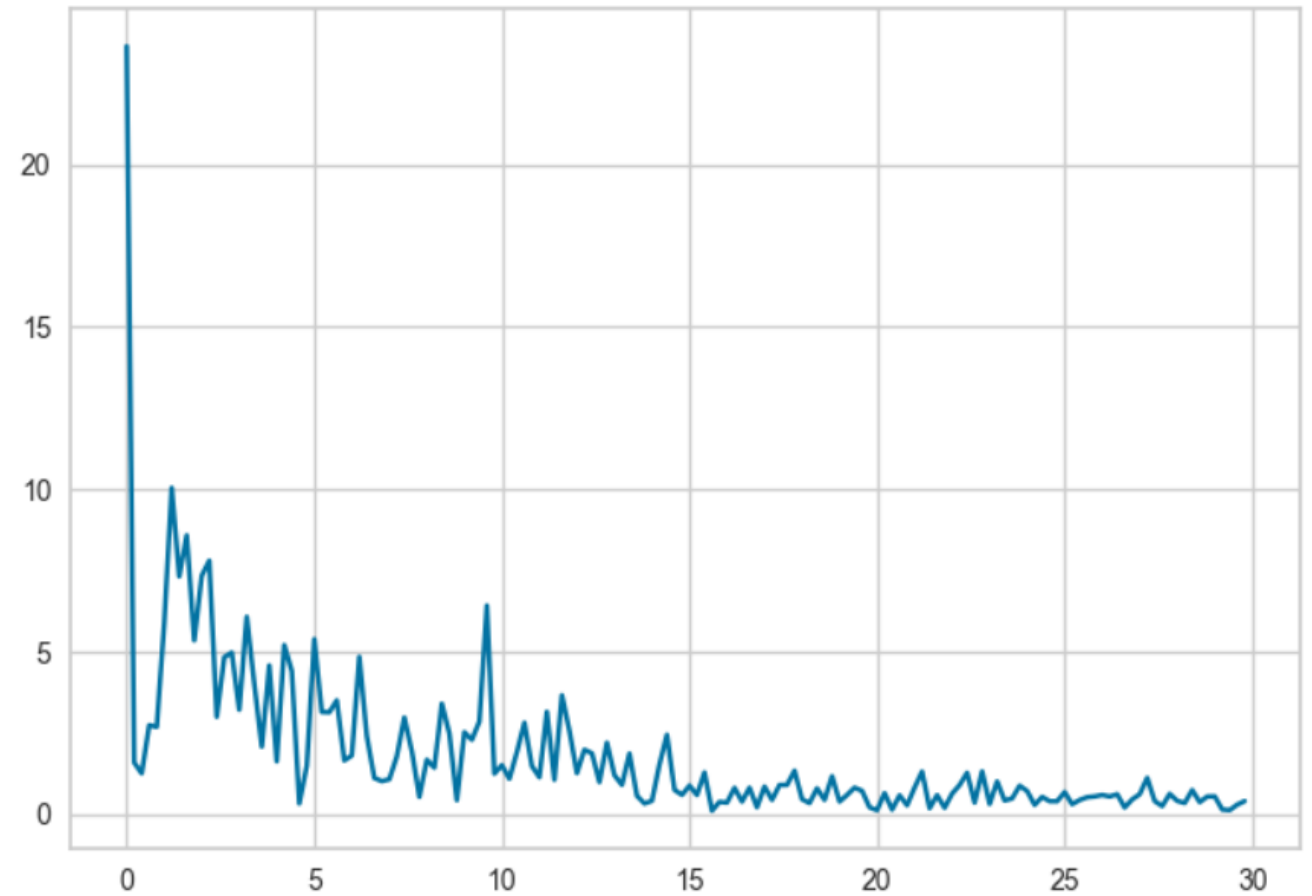


Pipeline#2

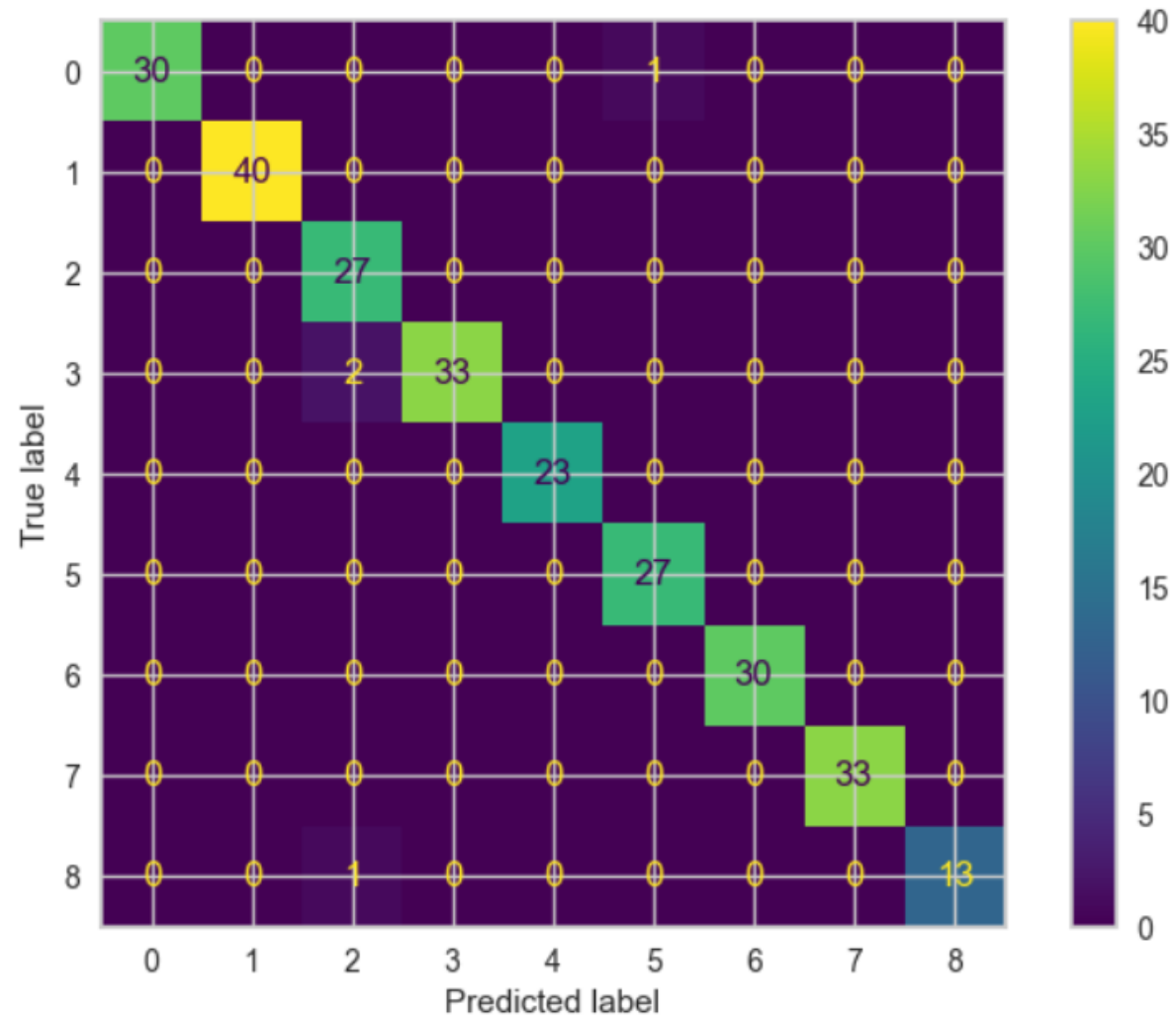
LightGBM / Catboost

Add Features

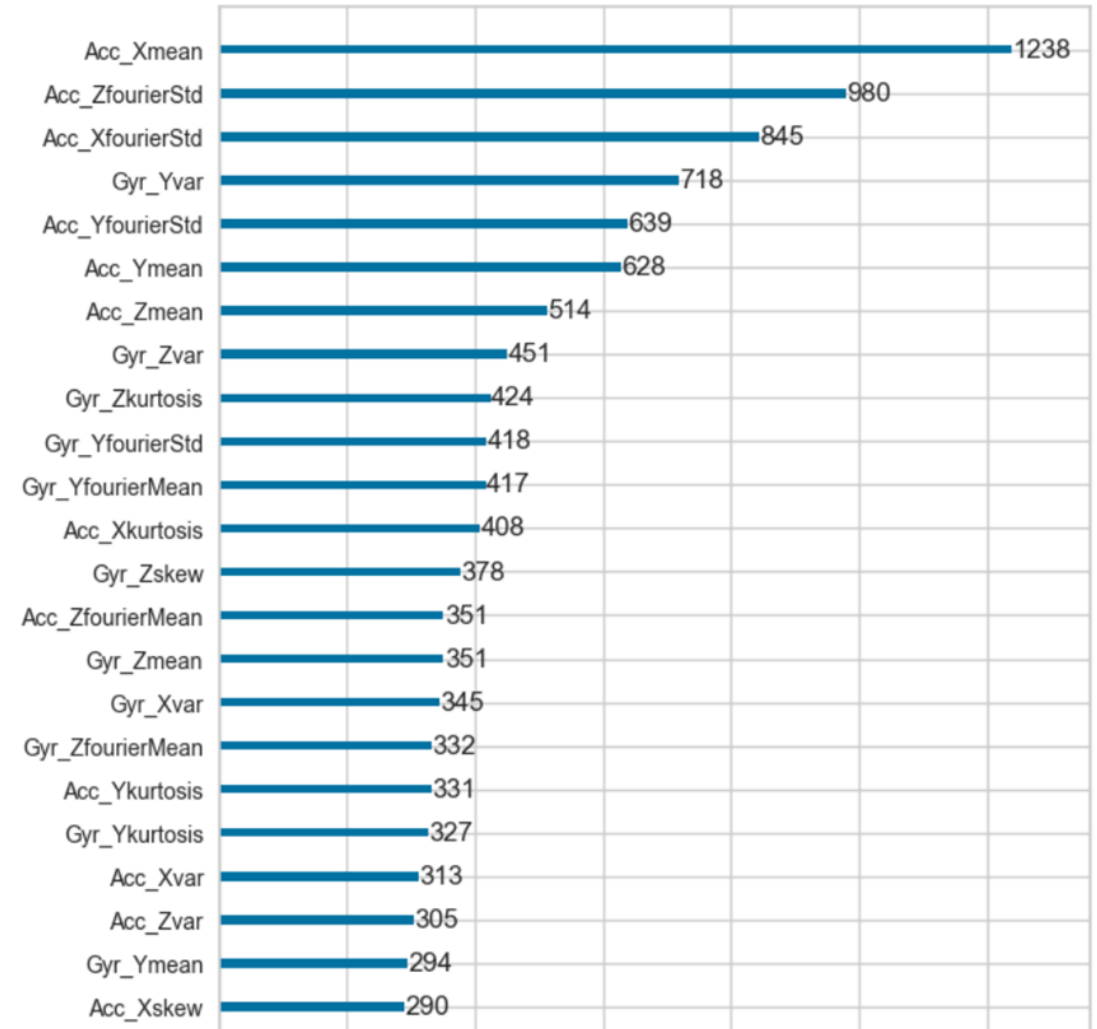
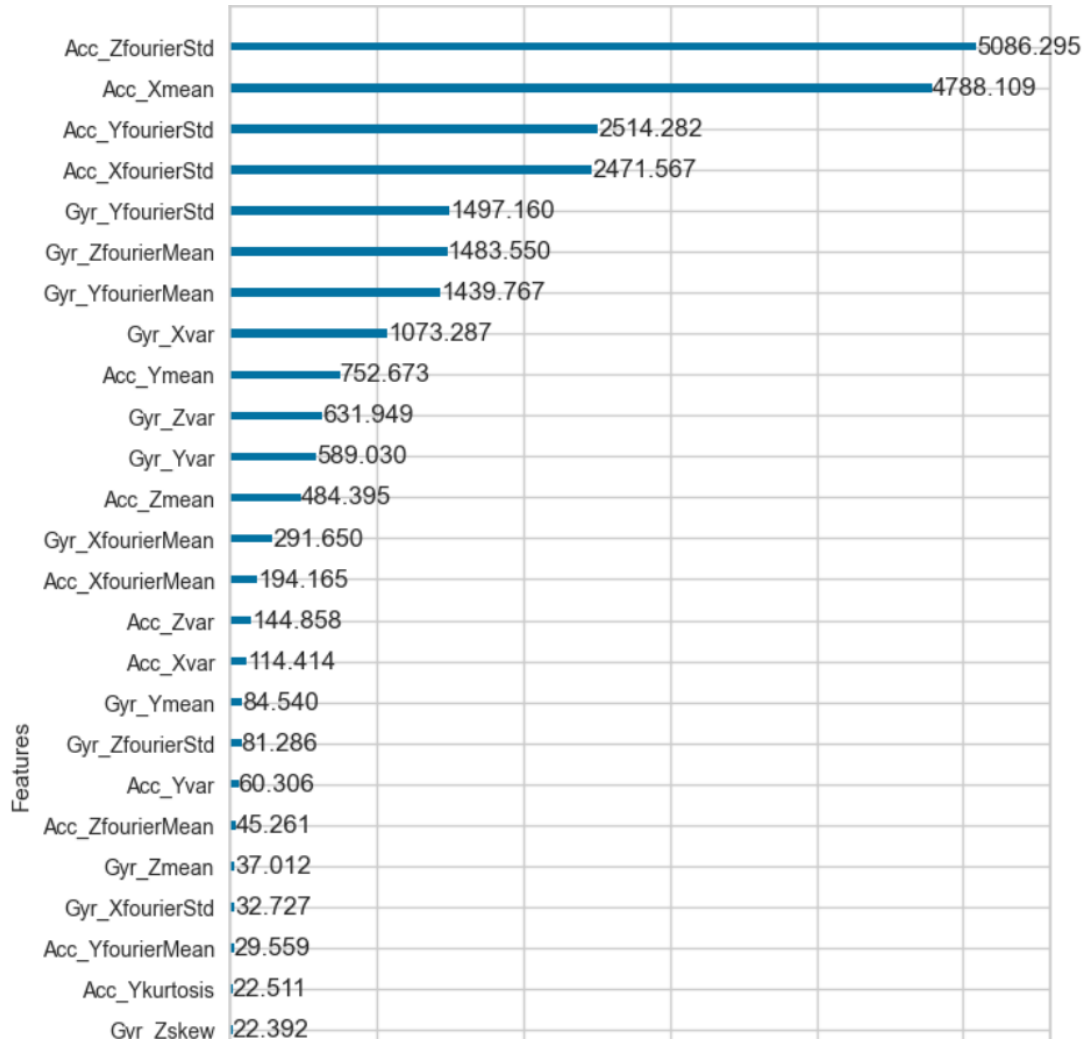
- Fourier Transform
- Kurtosis/Skew
- Stratified 5-Fold







Confusion Matrix



Feature Importance(gain/split)

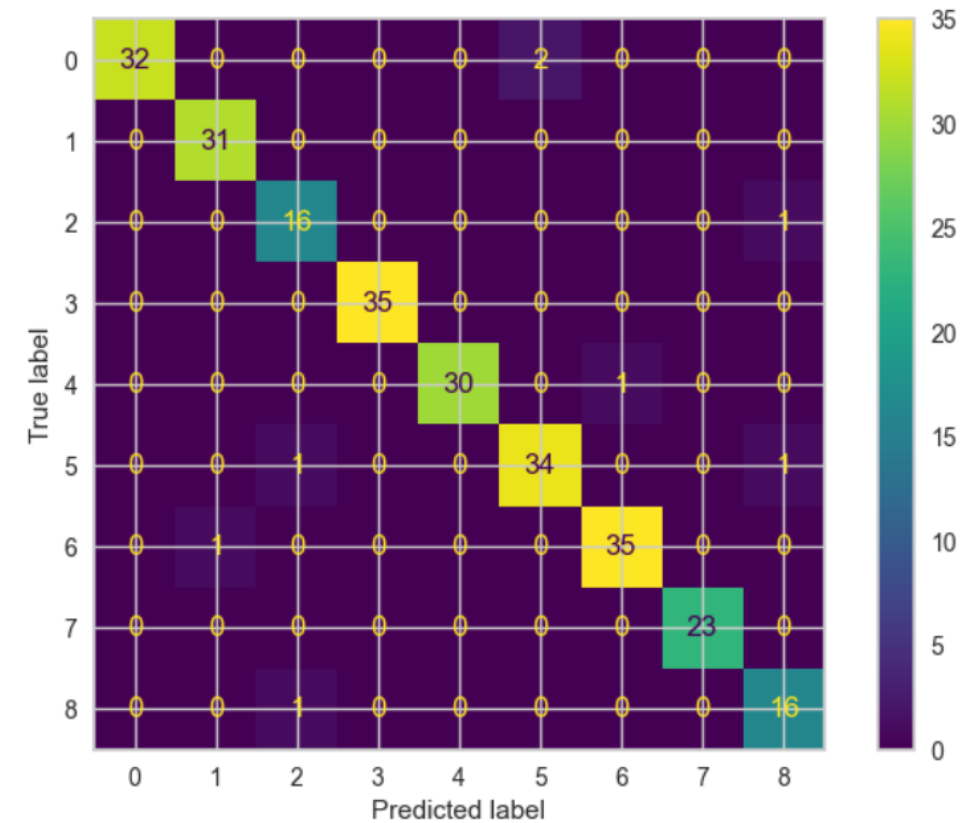


Results: inference time 1+0.8 second

Submission and Description		Private Score ⓘ	Public Score ⓘ	Selected
	submission_lgbm_5fold.csv Complete · 400256-สมบูรณ์ · 5h ago	0.97125	0.99178	
	submission_catboost_5fold.csv Complete · 400256-สมบูรณ์ · 5h ago · Before	0.98562	0.98767	

Same confusion matrix?

✓	submissionCleanedData.csv Complete · 402829-ssrass · 11h ago · amogus	0.86447	0.88706
✓	submissionFixed.csv Complete · 402829-ssrass · 17h ago · fixed rolling window	0.93634	0.94866
✓	submission.csv Complete · 402829-ssrass · 18h ago · Light gbm + 10 lag + mean	0.93018	0.92402




Future plans

- Data oversampling
- Training on extra datasets

Human Activity Recognition with Smartphones

Recordings of 30 study participants performing activities of daily living



[Data Card](#) [Code \(401\)](#) [Discussion \(13\)](#) [Suggestions \(0\)](#)

About Dataset

The Human Activity Recognition database was built from the recordings of 30 study participants performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors. *The objective is to classify activities into one of the six activities performed.*

Description of experiment

The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The experiments have been video-recorded to label the data manually. The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data.

The sensor signals (accelerometer and gyroscope) were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window). The sensor acceleration signal, which has gravitational and body motion components, was separated using a Butterworth low-pass filter into body acceleration and gravity. The gravitational force is assumed to have only low frequency components, therefore a filter with 0.3 Hz cutoff frequency was used. From each window, a vector of features was obtained by calculating variables from the time and frequency domain.

Attribute information

For each record in the dataset the following is provided:

Usability

7.06

License

[CC0: Public Domain](#)

Expected update frequency

Not specified

Tags

[Arts and Entertainment](#)[Earth and Nature](#)

Future plans

- More visualization (Sankey-diagram)
- More Feature engineering(entropy, PSD, energy, jerk, Freq. binning)
- Optuna implementation
- Experiment with filters
- Experiment with ViTs

Q & A