Research on Image Encoding Techinques for Multivariate Time Series Data from Human Activity Recognition

Yairi Laboratory

A1777489 Dong Hyeon Park

Table of Contents

Table of Contents

INTRODUCTION

- Human Activity Recognition (HAR)
- Application for Sensor Based HAR
- Objective

Human Activity Recognition (HAR)

HAR refers to sets of technology predicting human current actions

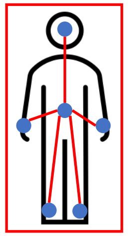
Vision Based

Pros

- Affordable and easy to collect
- Intensive history of studies

Cons

- Limited mobility only within camera coverage
- Computationally expensive



Sensor Based

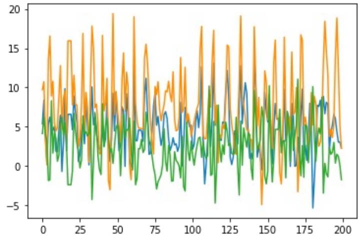
This Study

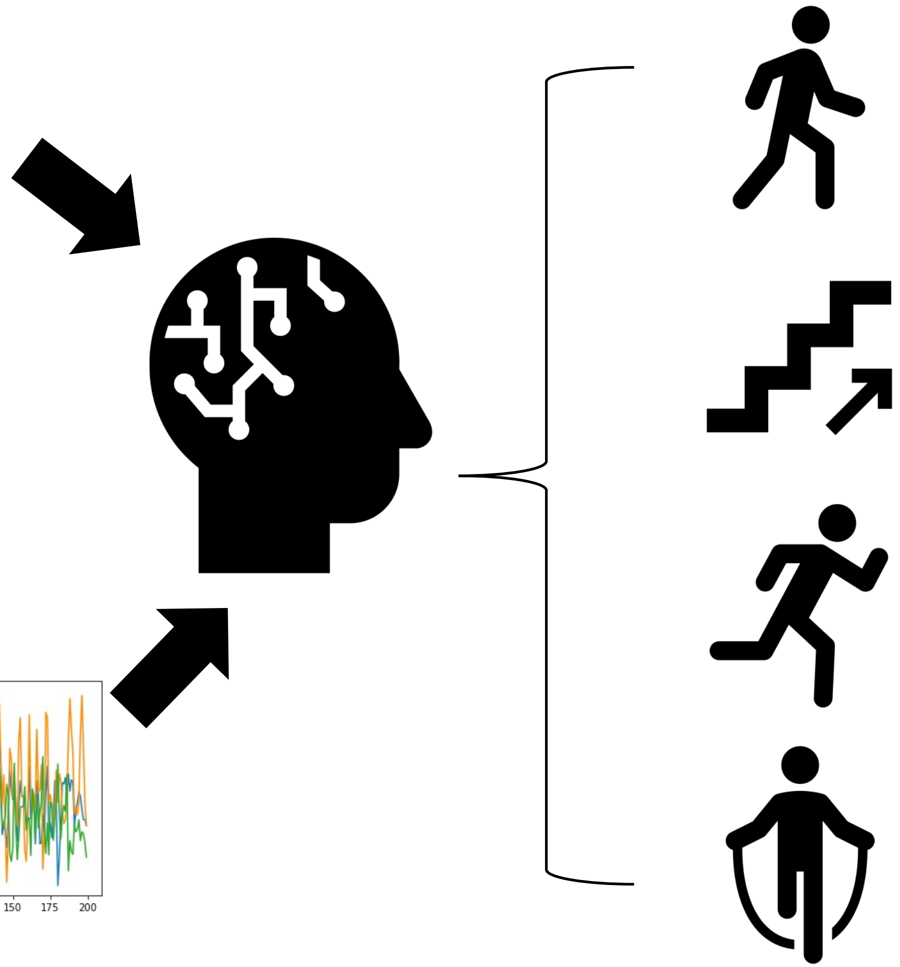
Pros

- Better mobility with conitnuous mointoring
- Less outer environmental intervention

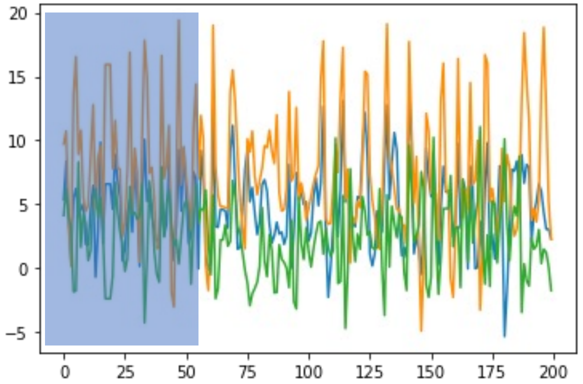
Cons

- Requires user to wear sensors
- Not much study performed

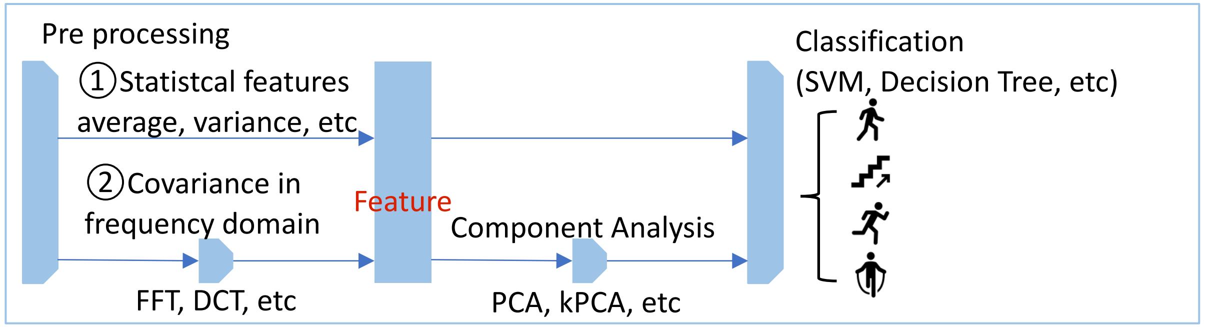


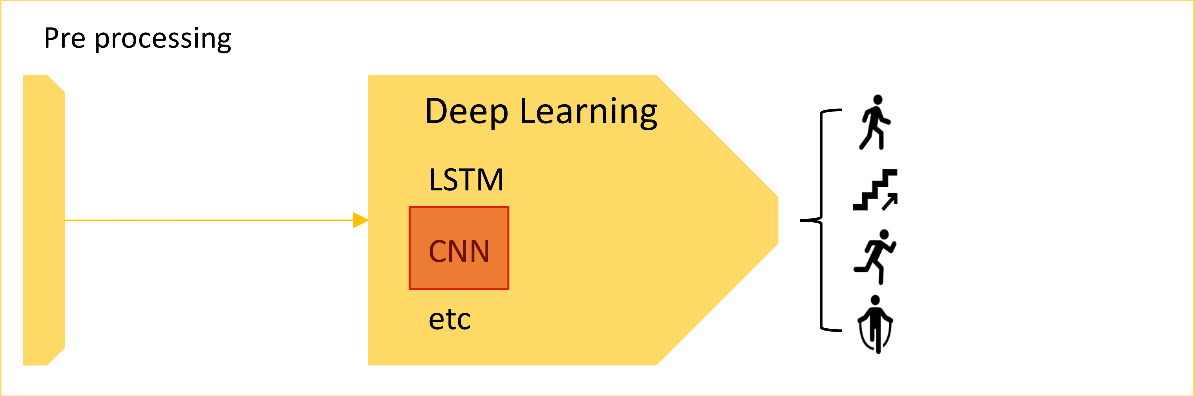


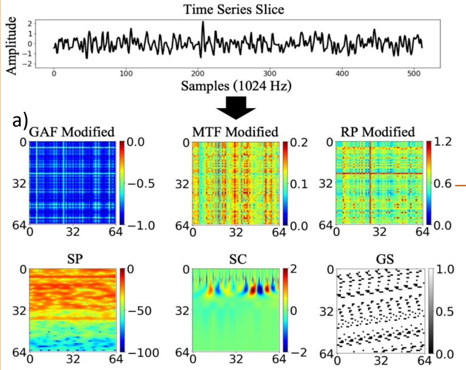
Approaches for Sensor Based HAR

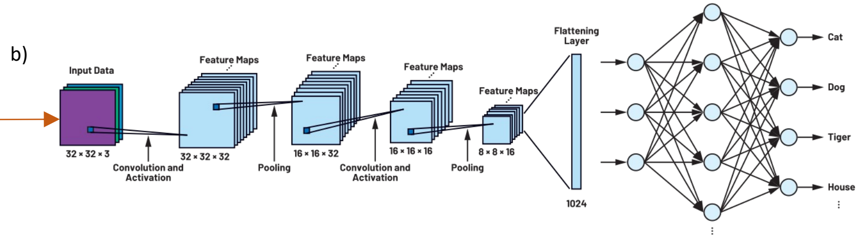


Window
Segment

① 

② 





CNN, ResNet, ImageNet, etc

This Study

- Does Image Encoding offer any improvements?
- Which image encoding offers the best performance?
- How to deal with multivariate time series data

a) Garcia, G. R., Michau, G., Ducoffe, M., Gupta, J. S., & Fink, O. (2022). Temporal signals to images: Monitoring the condition of industrial assets with deep learning image processing algorithms. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *236*(4), 617-627.

b) https://www.analog.com/jp/analog-dialogue/articles/training-convolutional-neural-networks-what-is-machine-learning-part-2.html

Objective

- Comparison of classificaiton performance of each image encoding method against conventional raw plot
- Comparison of classification performance of various approaches tackling multivariate time series data

Table of Contents

METHODOLOGY

- Dataset
- Image Encoding
- Model Architecture
- Evaluation Metrics
- Leave One Subject Out (LOSO)

Dataset - WISDM

- Number of subjects: 29
- Number of activities: 6

  (Walking, Jogging, Upstairs, Downstairs, Sitting, Standing)
- Collected data in a controlled laboratory environment
- Subjects performed instructed activity continually with smartphone in pockets

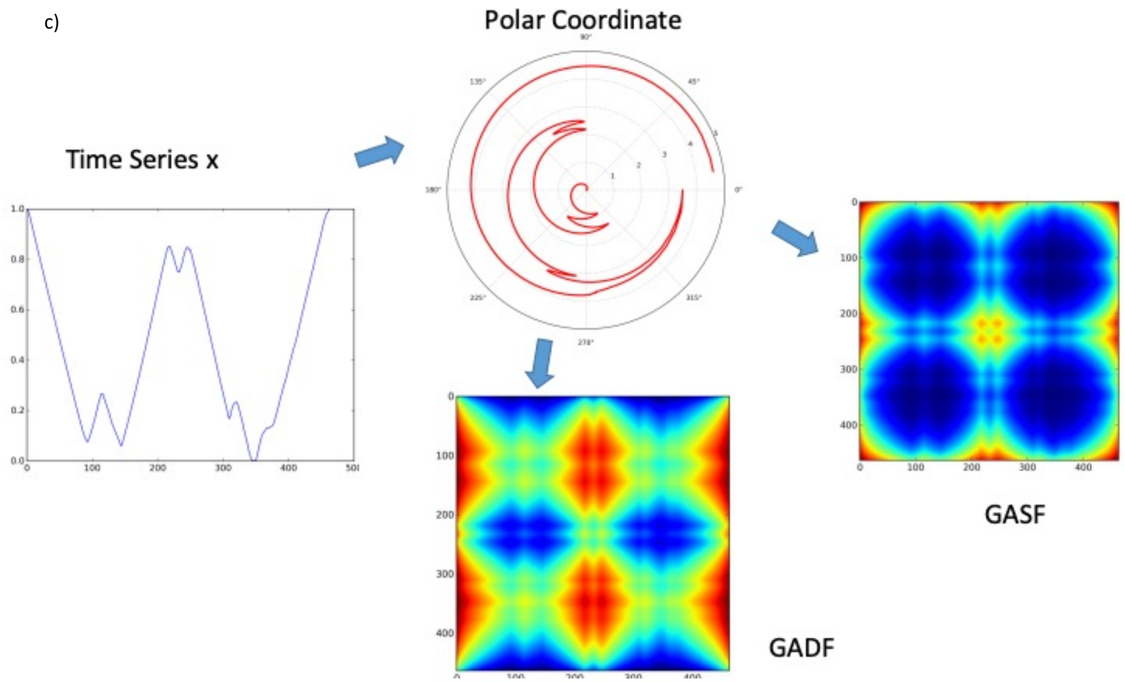
Dataset – Wheelchair Acceleration Data

- Number of subjects: 3 electric wheelchair + 3 manual wheelchair
- Number of labels: 5

  (Slope, Stop, Curb, Smooth, TI->Braille Blocks)
- Collected data recorded through sensor attached on wheelchair
- Subjects took total of 3 round trips (~1.4km/trip) around Yotsuya station

Image Encoding – Gramian Angular Fields (GAF)

c)



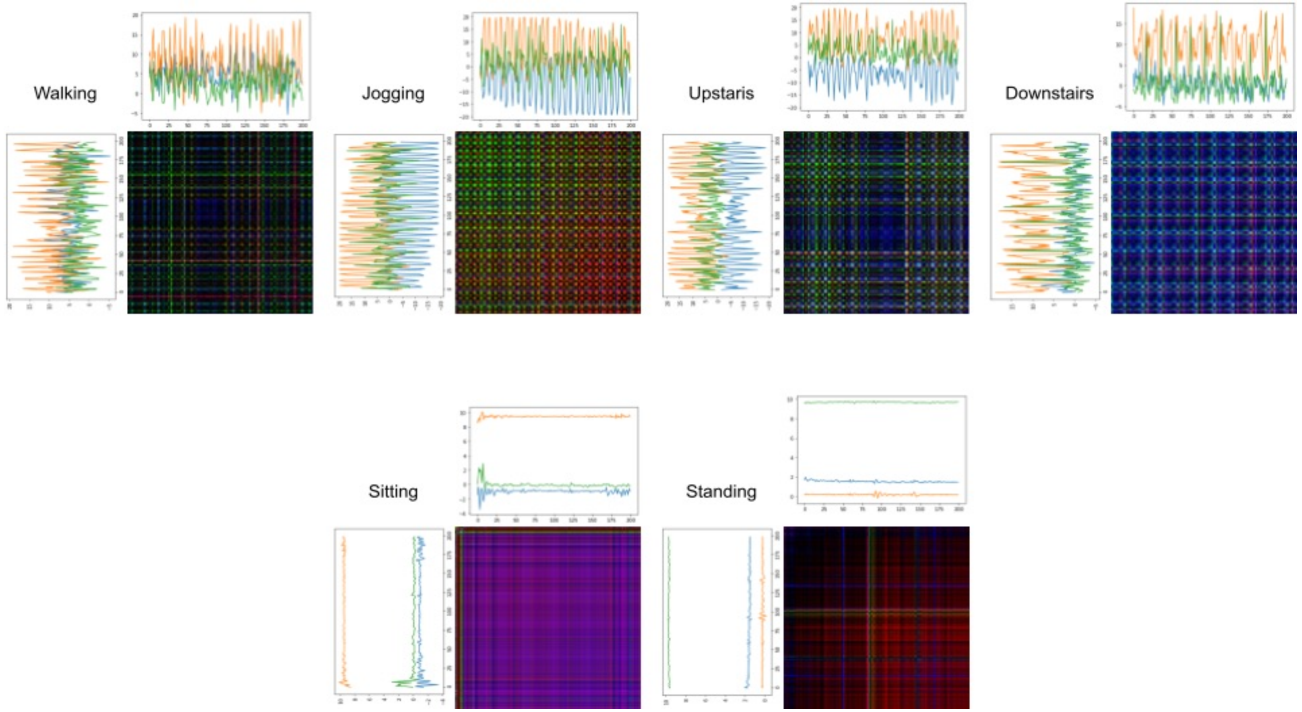
c) $$\begin{cases}\phi = \arccos(\tilde{x}_i), -1 \le \tilde{x}_i \le 1, \tilde{x}_i \in \tilde{X} \\ r = \frac{t_i}{N}, t_i \in \mathbb{N}\end{cases}$$

$\Longrightarrow$A time series X is encoded as the angular cosine and the time stamp as the radius

c) $$\begin{aligned} GASF &= [\cos(\phi_i + \phi_j)] \\ &= \tilde{X}' \cdot \tilde{X} - \sqrt{I - \tilde{X}^2}' \cdot \sqrt{I - \tilde{X}^2} \\[1em] GADF &= [\sin(\phi_i - \phi_j)] \\ &= \sqrt{I - \tilde{X}^2}' \cdot \tilde{X} - \tilde{X}' \cdot \sqrt{I - \tilde{X}^2} \end{aligned}$$

$\Longrightarrow$ i, j each are positions in the original time series array.

WISDM in GASF



- A matrix that represents the angular relationships between the different components of a time series.

c) Wang, Z., & Oates, T. (2015). Imaging Time-Series to Improve Classification and Imputation. *International Joint Conference on Artificial Intelligence*.

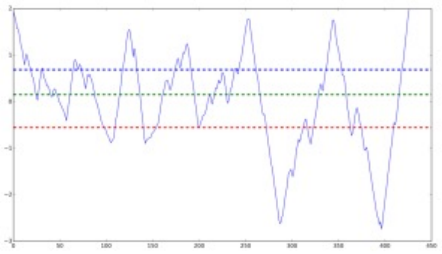
Image Encoding – Markov Transition Fields (MTF)

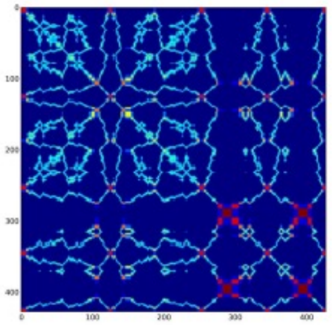
c)

$$\begin{array}{c|cccc} & A & B & C & D \\ \hline A & 0.944 & 0.056 & 0 & 0 \\ B & 0.056 & 0.860 & 0.084 & 0 \\ C & 0 & 0.085 & 0.840 & 0.075 \\ D & 0 & 0 & 0.075 & 0.925 \end{array}$$

Markov Transition Matrix

Time Series x





MTF

➢A matrix that represents the transition probabilities between different states of a time series.

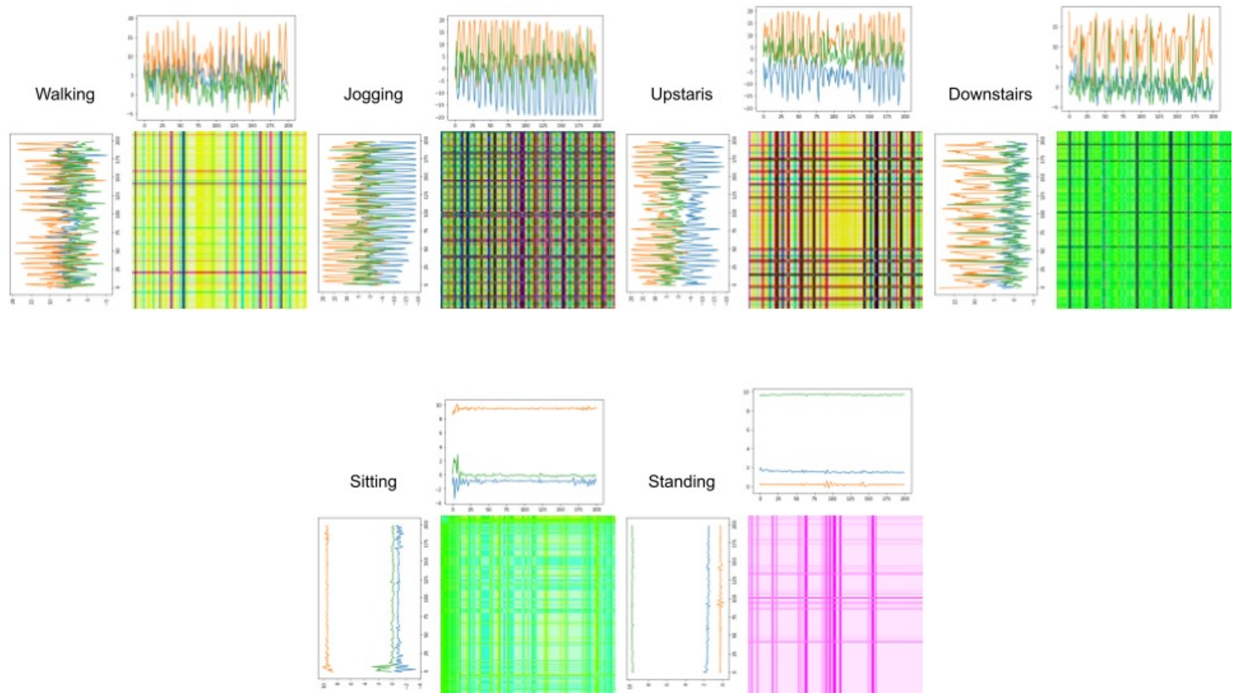
Q × Q weighted adjacency matrix is constructed by counting transitions among Q pre- assigned quantile bins in the manner of a first-order Markov chain along the time axis.

c)

$$M = \begin{bmatrix} w_{ij|x_1\in q_i, x_1\in q_j} & \cdots & w_{ij|x_1\in q_i, x_n\in q_j} \\ w_{ij|x_2\in q_i, x_1\in q_j} & \cdots & w_{ij|x_2\in q_i, x_n\in q_j} \\ \vdots & \ddots & \vdots \\ w_{ij|x_n\in q_i, x_1\in q_j} & \cdots & w_{ij|x_n\in q_i, x_n\in q_j} \end{bmatrix}$$

$\Longrightarrow q_i \longrightarrow q_j$ transitional probability consists the MTF matrix M.

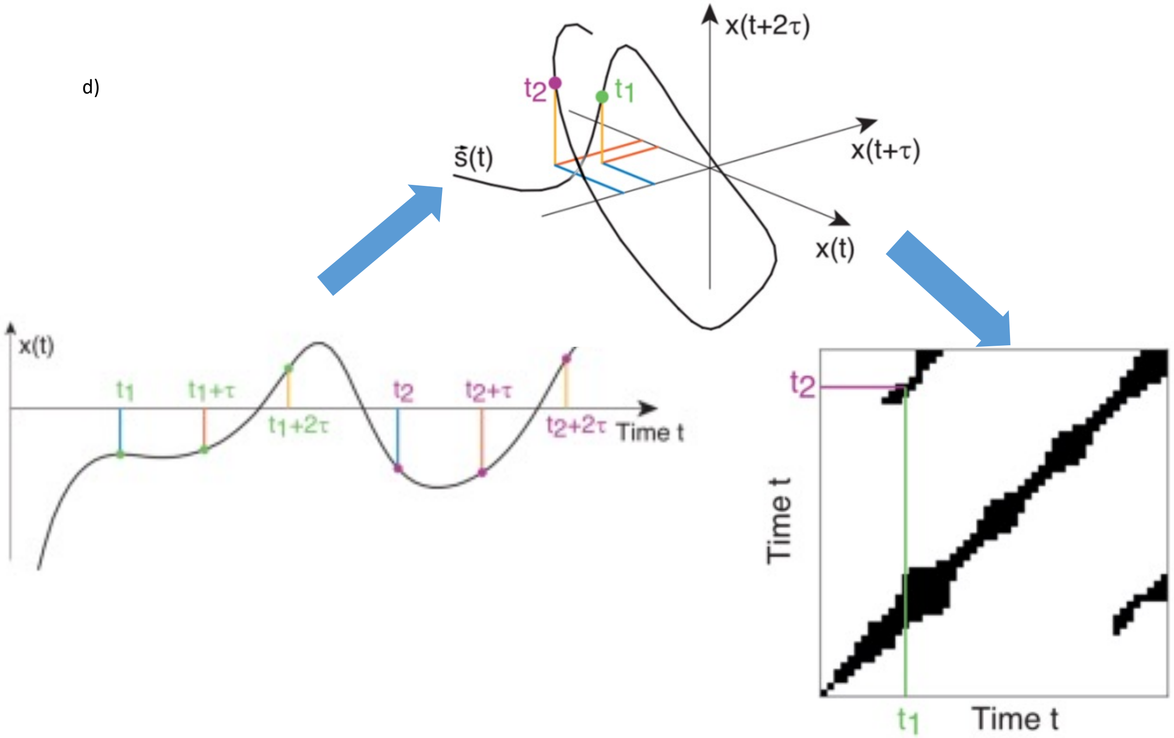
WISDM in MTF



c) Wang, Z., & Oates, T. (2015). Imaging Time-Series to Improve Classification and Imputation. *International Joint Conference on Artificial Intelligence*.

Image Encoding – Recurrence Plot (RP)

d)



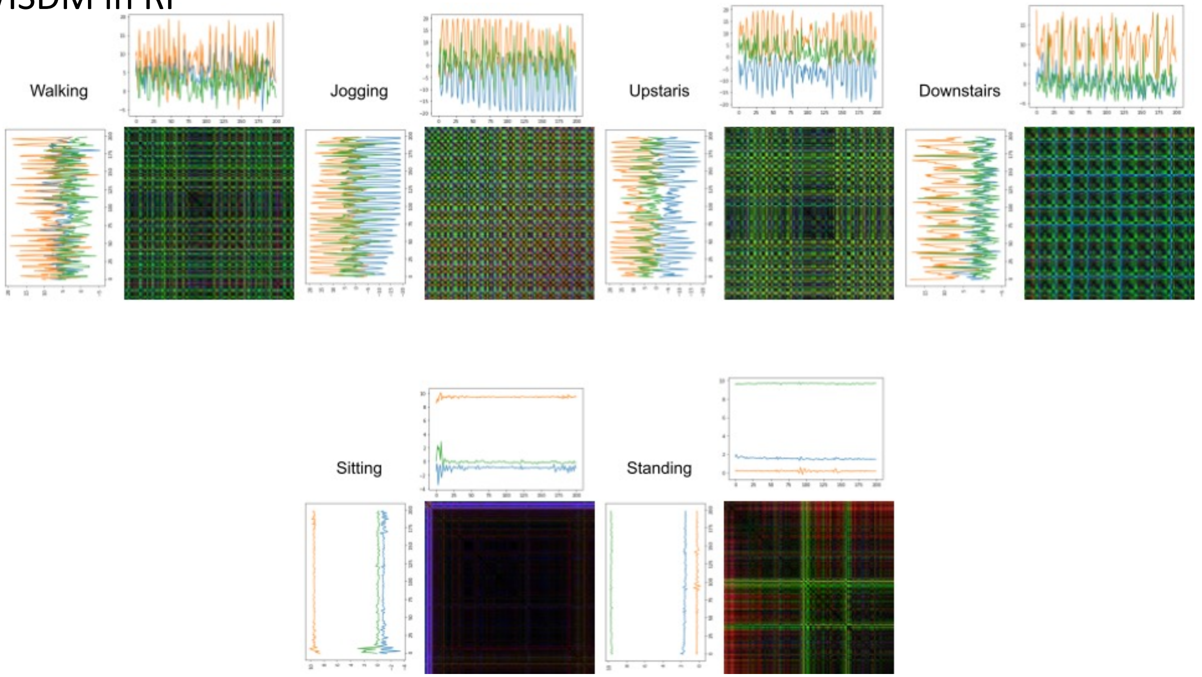
$x(t)$ is sepereated into several bins by time-delay embedding. $x(t), x(t+\tau), and\ x(t+2\tau)$ with a temporal seperation of τ are represented within the phase space.

e) $$R_{ij} = \begin{cases} 1 & \text{if} \quad \|\mathbf{X}_i - \mathbf{X}_j\| \le \varepsilon \\ 0 & \text{if} \quad \|\mathbf{X}_i - \mathbf{X}_j\| > \varepsilon. \end{cases}$$

$\Longrightarrow$ ||·|| refers to Euclidean distance and ε a small radius within which two points will be considered equal.

Such visualization of reccurent system is referred as Reccurence Plot.

WISDM in RP

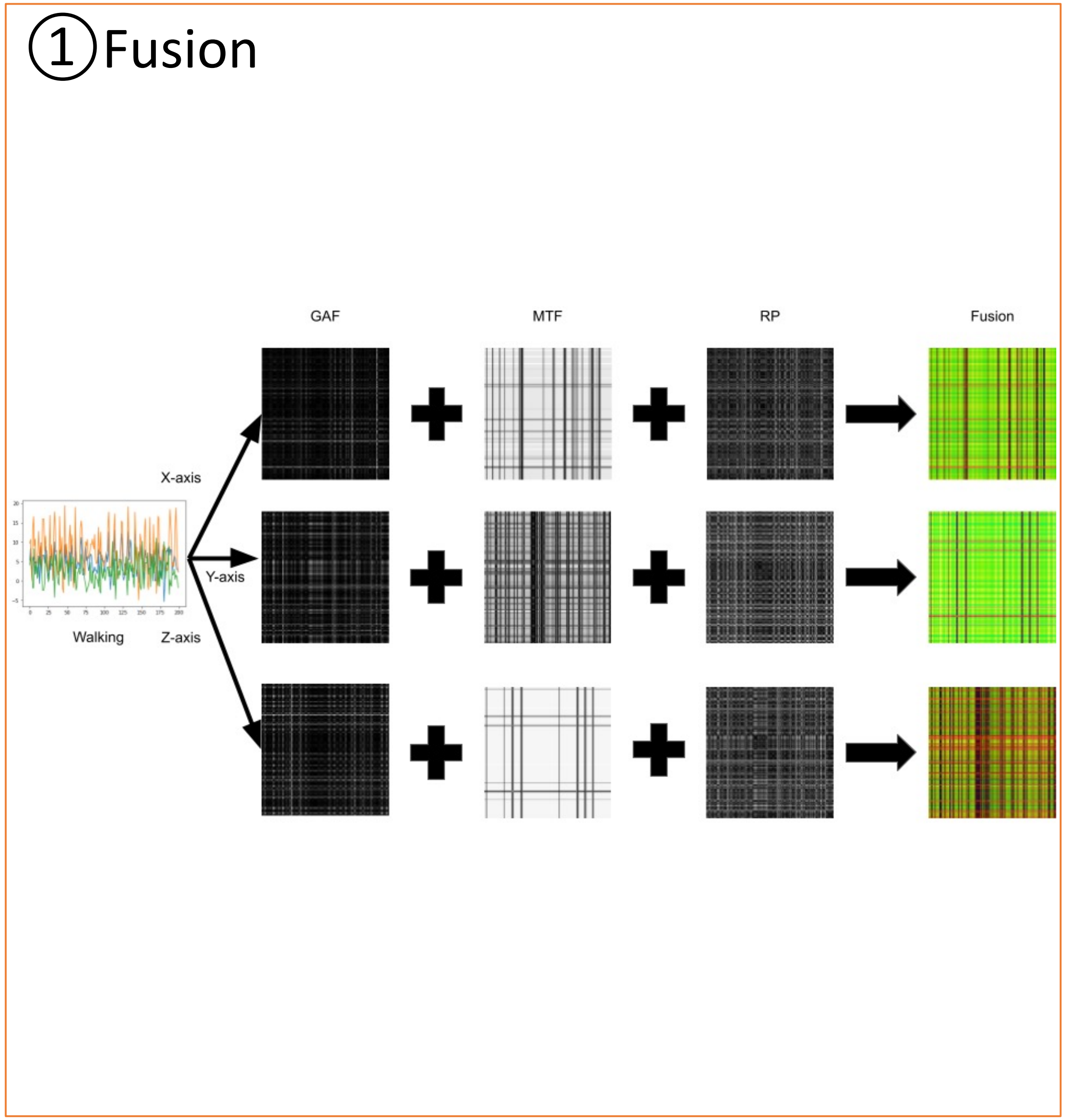


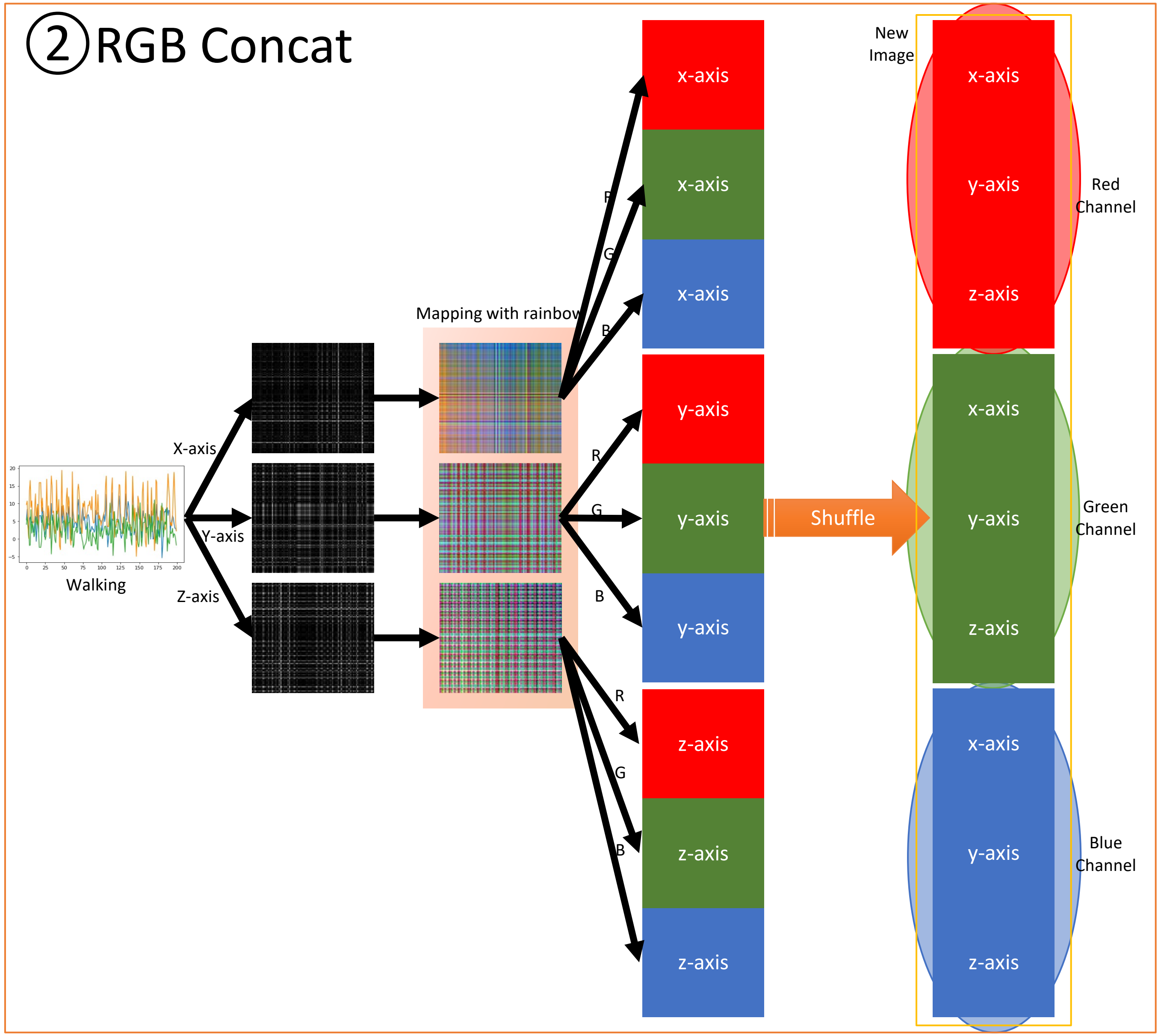
- A plot showing the reccurence of points in a time series data.

d) Trauth, M. H., Asrat, A., Duesing, W., Foerster, V., Kraemer, K. H., Marwan, N., ... & Schaebitz, F. (2019). Classifying past climate change in the Chew Bahir basin, southern Ethiopia, using recurrence quantification analysis. *Climate Dynamics*, *53*, 2557-2572.

e) Wallot, S., & Mønster, D. (2023). Multidimensional Joint Recurrence Quantification Analysis: detecting coupling between time series of different dimensionalities. *arXiv preprint arXiv:2303.16907*.

Image Encoding – Others





Model Architectures - Kadota

<u>Input</u>

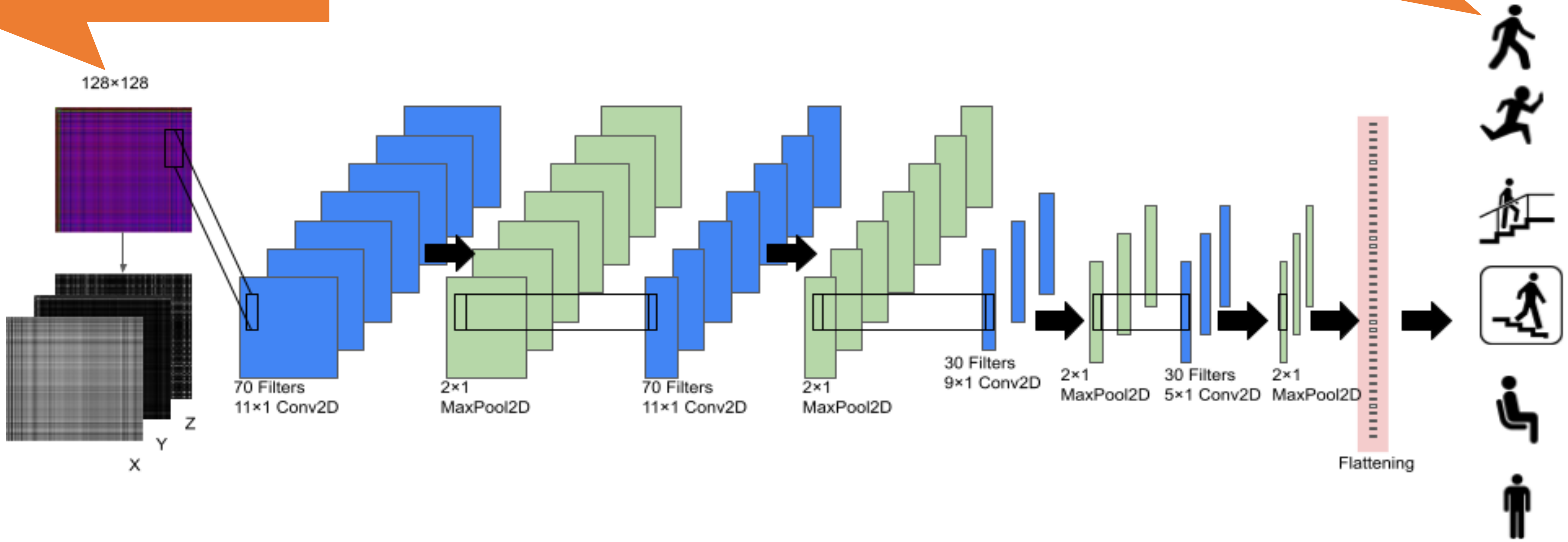
Width × Height × Channel

(128 × 128 × 3)

<u>Output</u>

WISDM: 6 predictions

Wheelchair: 5 predictions



Model Architectures - Multihead

Input1

Width × Height × Channel

(128 × 128 × 1)

Input2

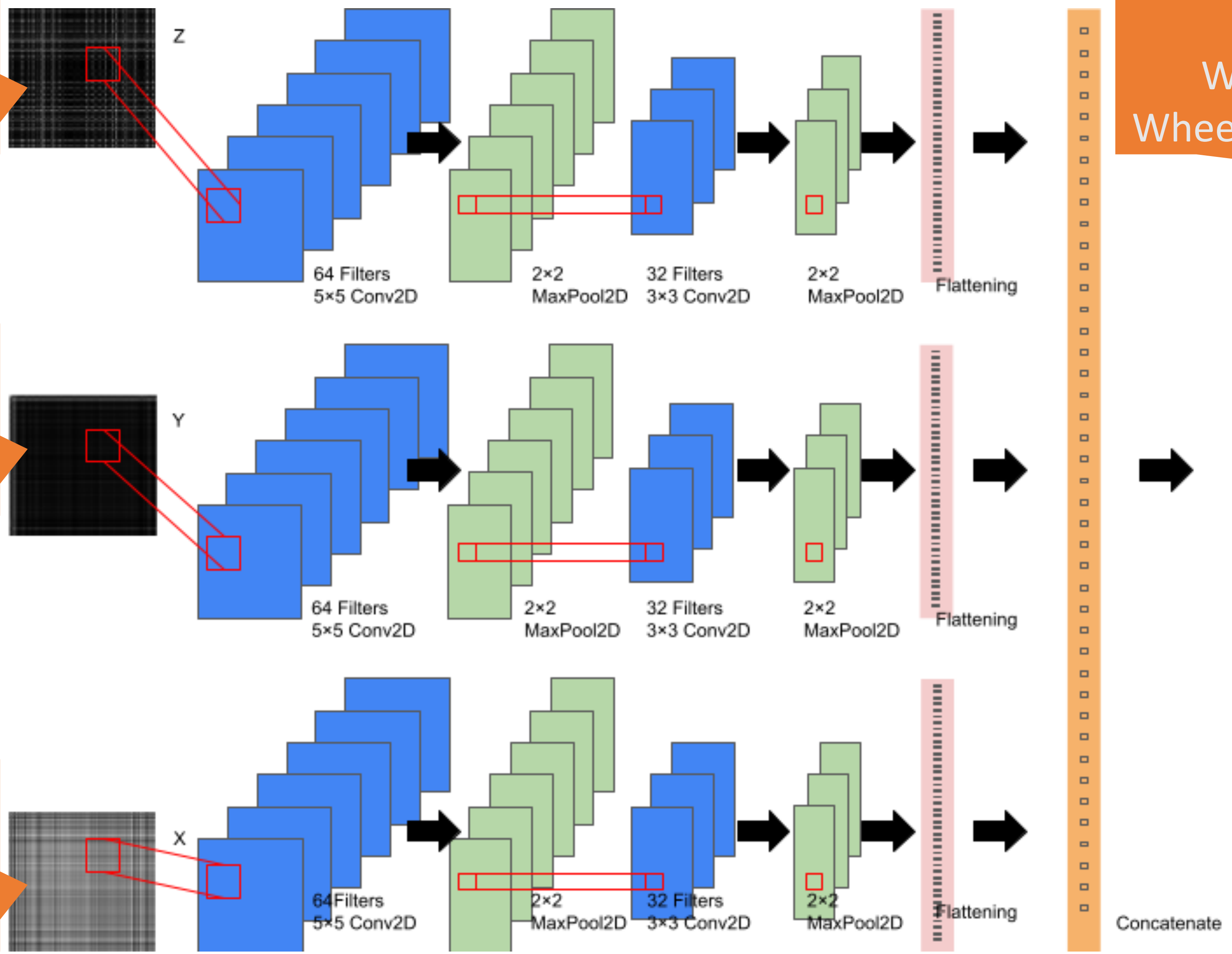
Width × Height × Channel

(128 × 128 × 1)

Input3

Width × Height × Channel

(128 × 128 × 1)



Output

WISDM: 6 predictions

Wheelchair: 5 predictions

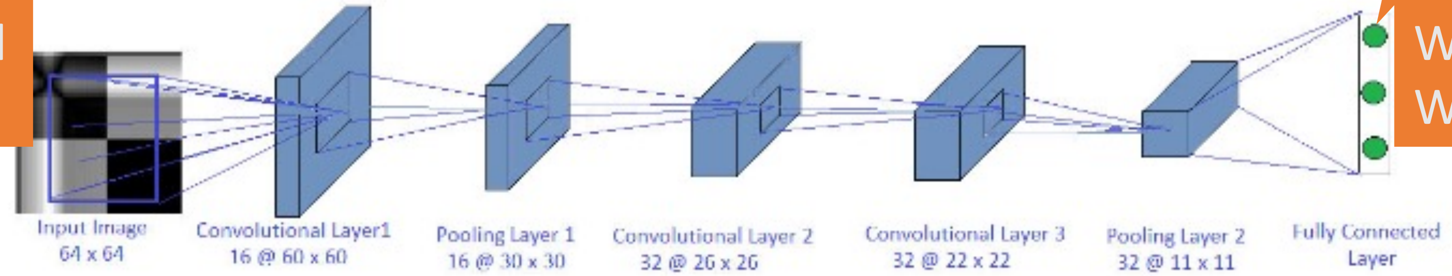
Model Architectures – Others

Three Inputs (x, y, z)

Width × Height × Channel
(128 × 128 × 1)

①Fusion

f)

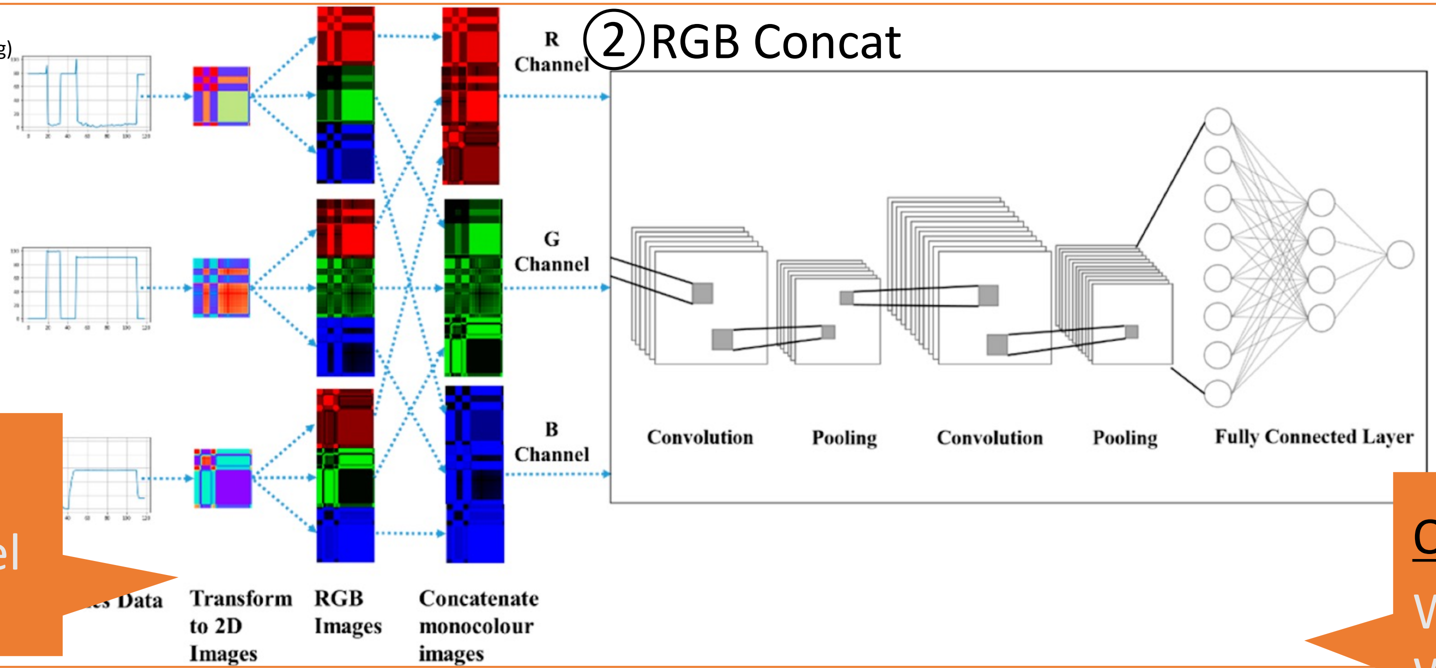


Output

WISDM: 6 predictions
Wheelchair: 5 predictions

②RGB Concat

g)



Input

Width × Height × Channel
(128 × 128 × 3)

Output

WISDM: 6 predictions
Wheelchair: 5 predictions

f) Ahmad, Z., Tabassum, A., Guan, L., & Khan, N. M. (2021). ECG heartbeat classification using multimodal fusion. *IEEE Access*, *9*, 100615-100626.
g) Yang, C. L., Chen, Z. X., & Yang, C. Y. (2019). Sensor classification using convolutional neural network by encoding multivariate time series as two-dimensional colored images. *Sensors*, *20*(1), 168.

Evaluation Metrics

- Recall: Ability of the model to correctly identify positive instances

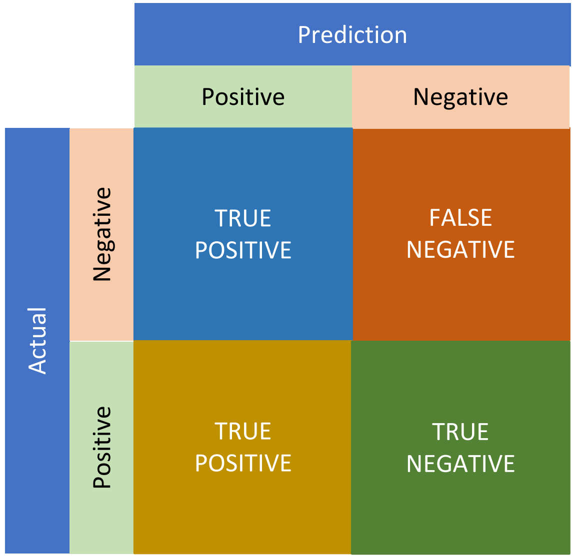
⟹ True Positives / (True Positives + False Negatives)

- Precision: Accuracy of the model's positive predictions

⟹ True Positives / (True Positives + False Positives)

- Accuracy: Overall correctness of the model's predictions

⟹(True Positives + True Negatives) / (True Positives + True Negatives + False Positives + False Negatives)



- F1 score: Balanced measure of precision and recall, particularly useful when dealing with imbalanced classes

Leave One Subject Out (LOSO)

➢ Splitting the test data on a subject basis to evaluate the accuracy of the model

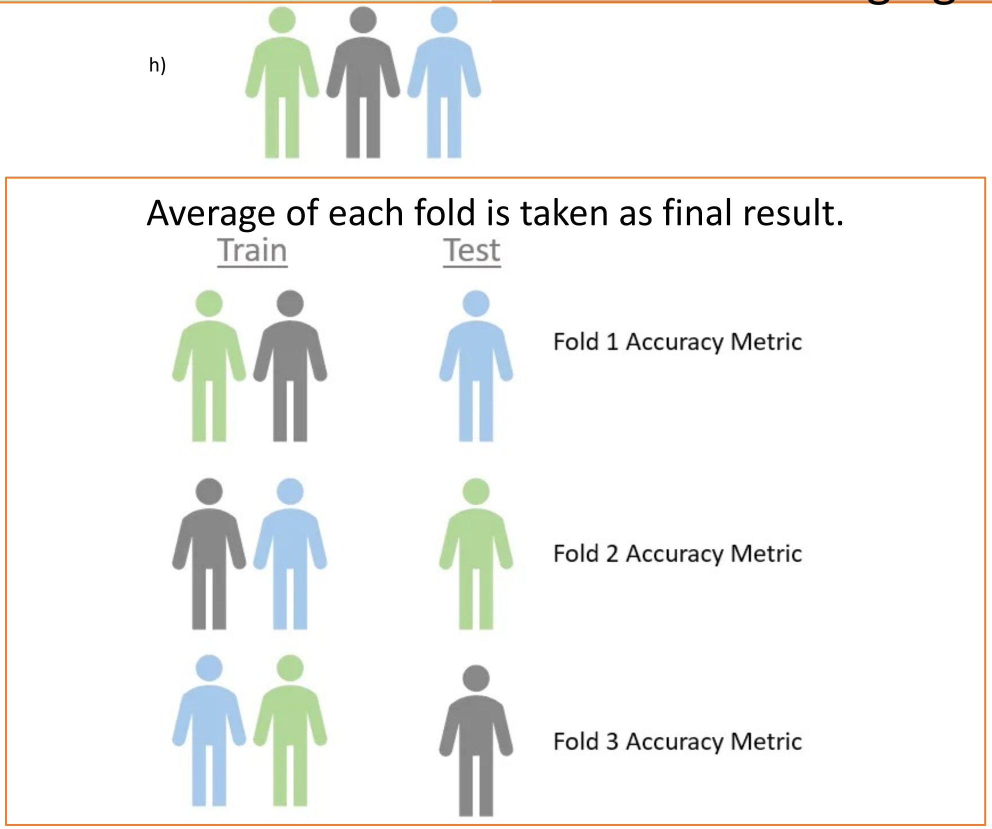
Pros

- Maximization of data utilization
- Reduction of bias

Cons

- High computational cost
- Possible biases due to data dependency
- Limited averaging

h)



h) Gholamiangonabadi, D., Kiselov, N., & Grolinger, K. (2020). Deep neural networks for human activity recognition with wearable sensors: Leave-one-subject-out cross-validation for model selection. *Ieee Access*, *8*, 133982-133994.

Table of Contents

EXPERIMENT

- Study 1
- Study 2
- Discussion

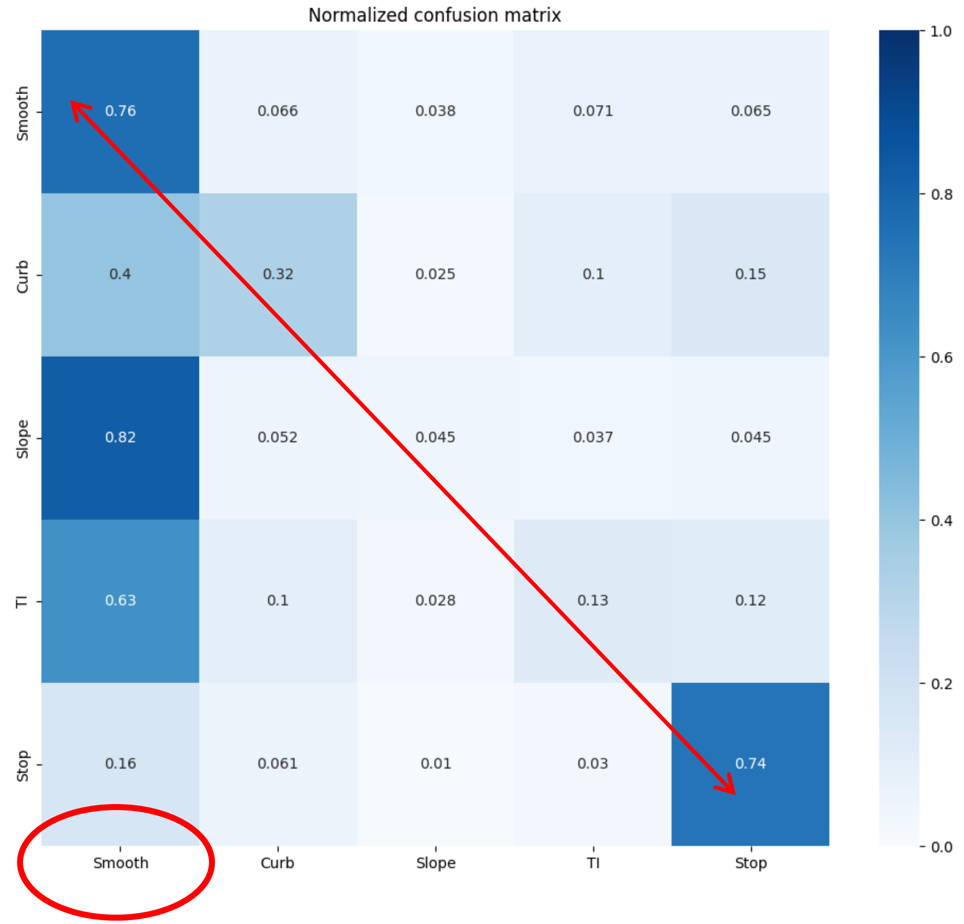
STUDY 1 – Image Encoding performance compared to raw plot

Data	Wheelchair		WISDM	
	Accuracy[%]	F-score[%]	Accuarcy[%]	F-score[%]
RAW	83.8	70.1	88.6	78.7
GADF	35.01	35.99	81.68	82.09
GASF	62.59	60.32	83.04	83.43
MTF	52.34	49.72	73.89	73.37
RP	30.14	28.76	72.41	70.67

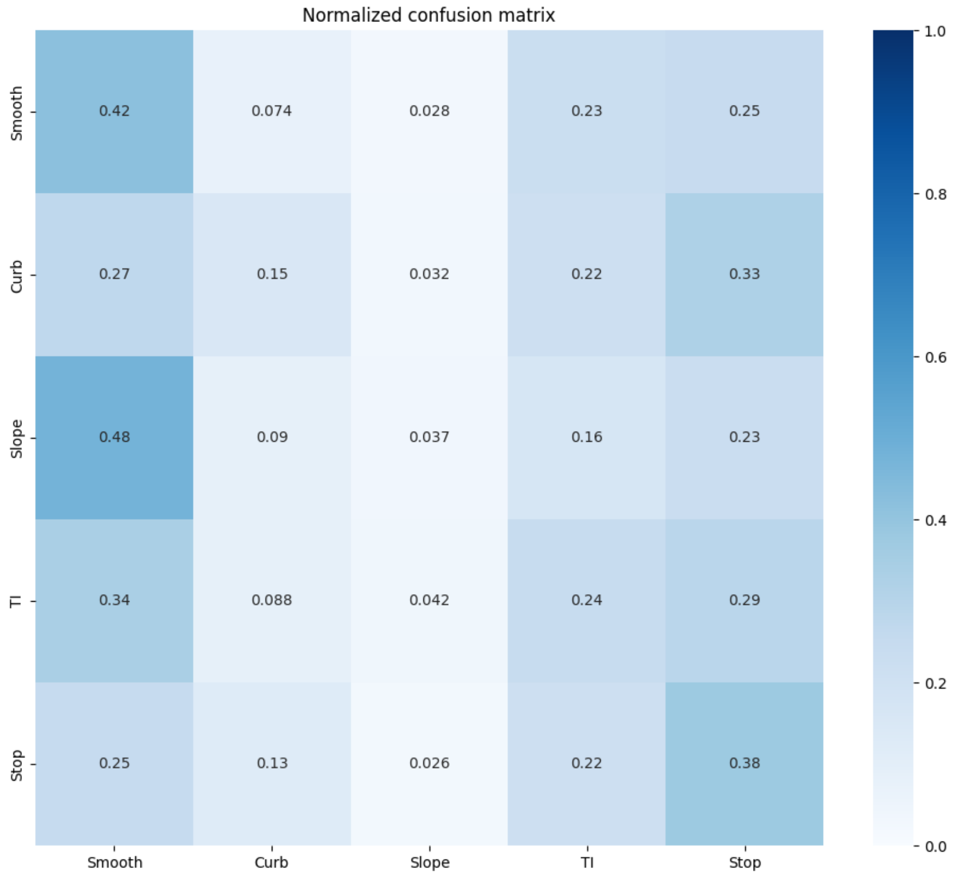
STUDY 1 – Image Encoding performance compared to raw plot

Wheelchair

GASF



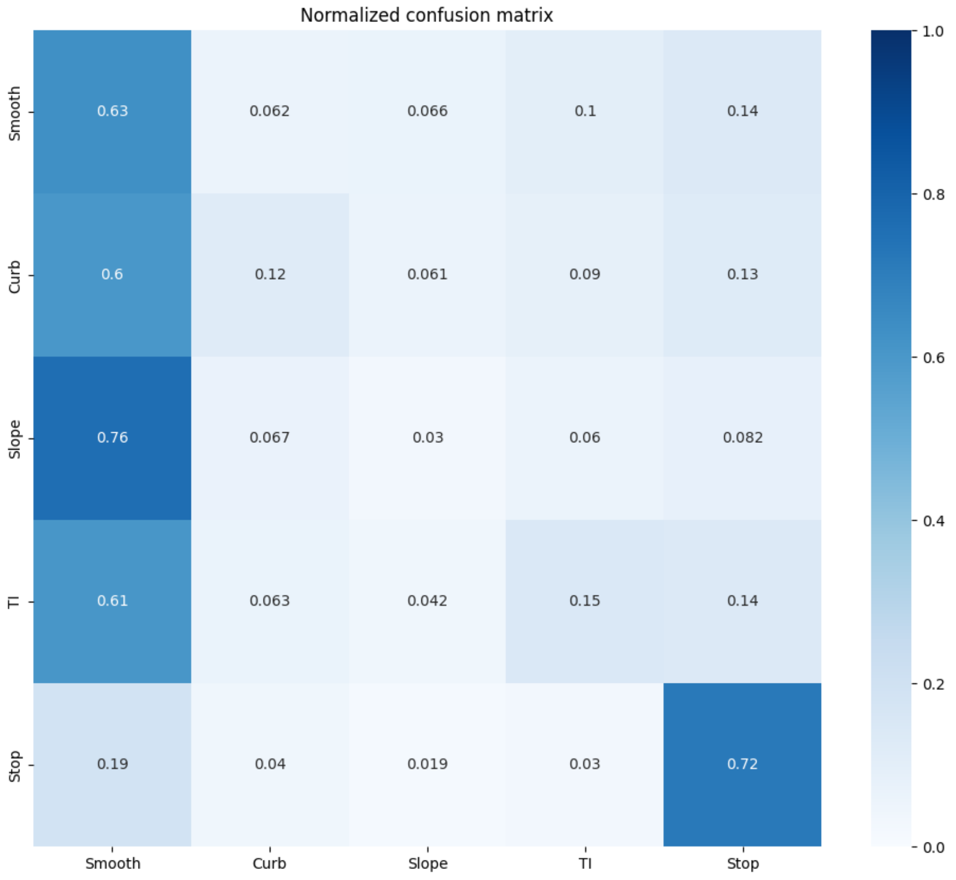
GADF



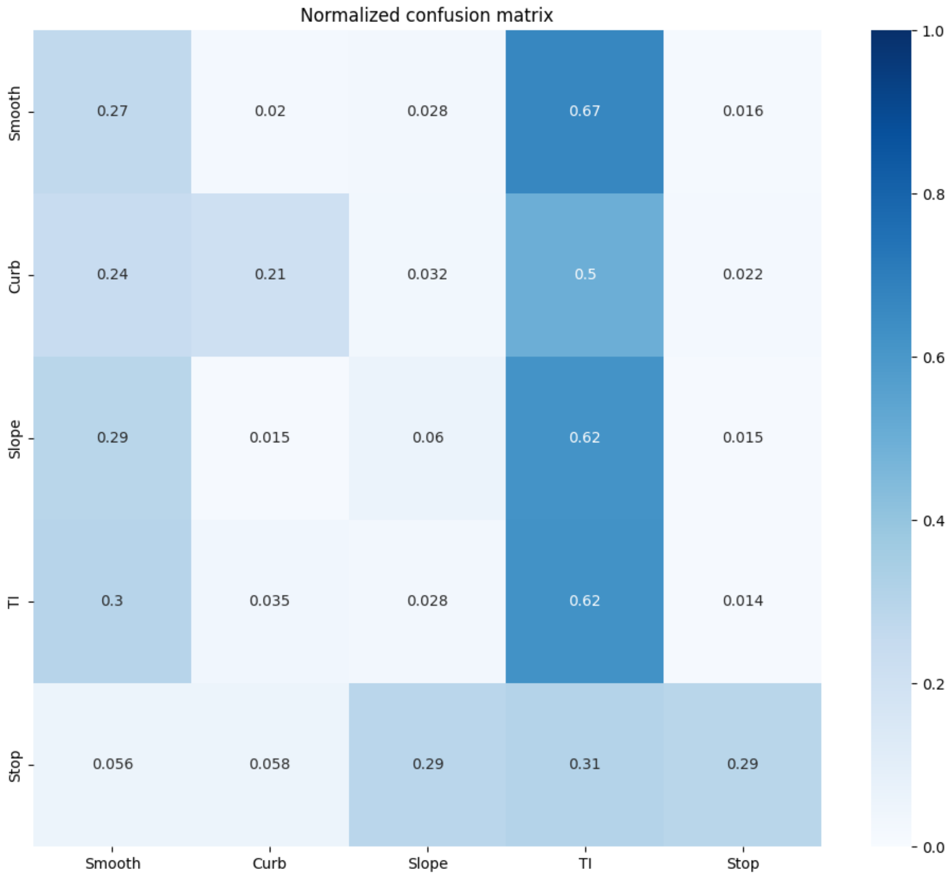
STUDY 1 – Image Encoding performance compared to raw plot

Wheelchair

MTF



RP

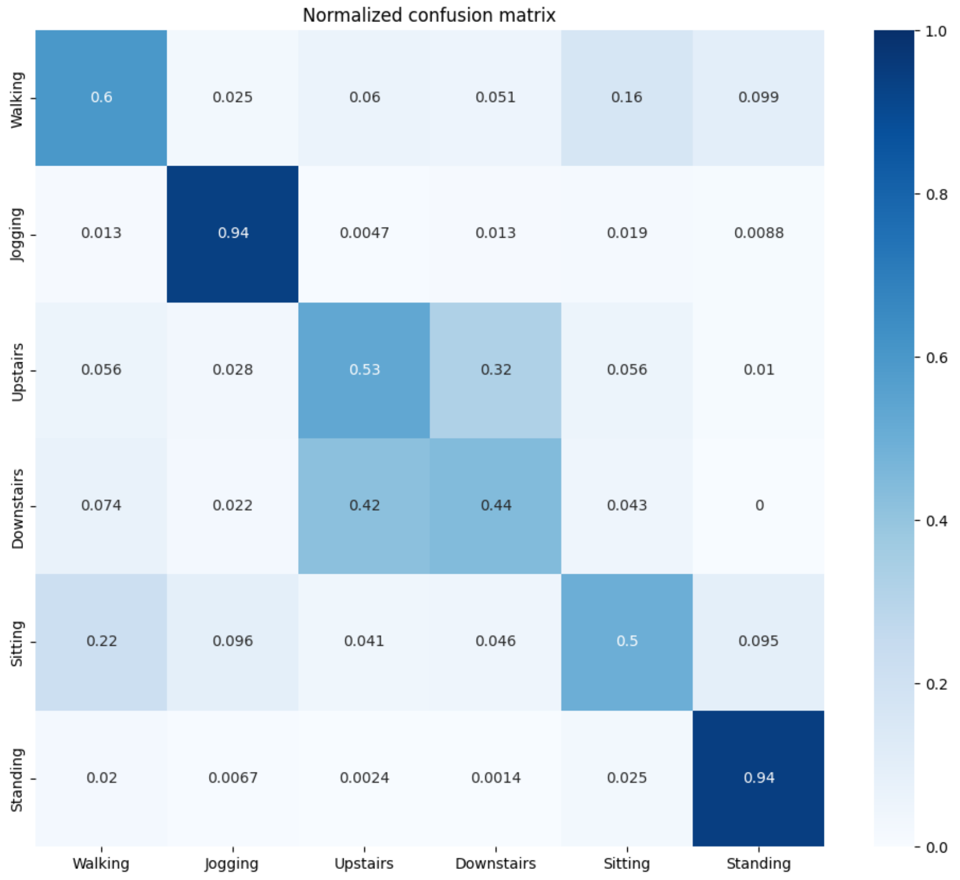


STUDY 1 – Image Encoding performance compared to raw plot

WISDM

GASF

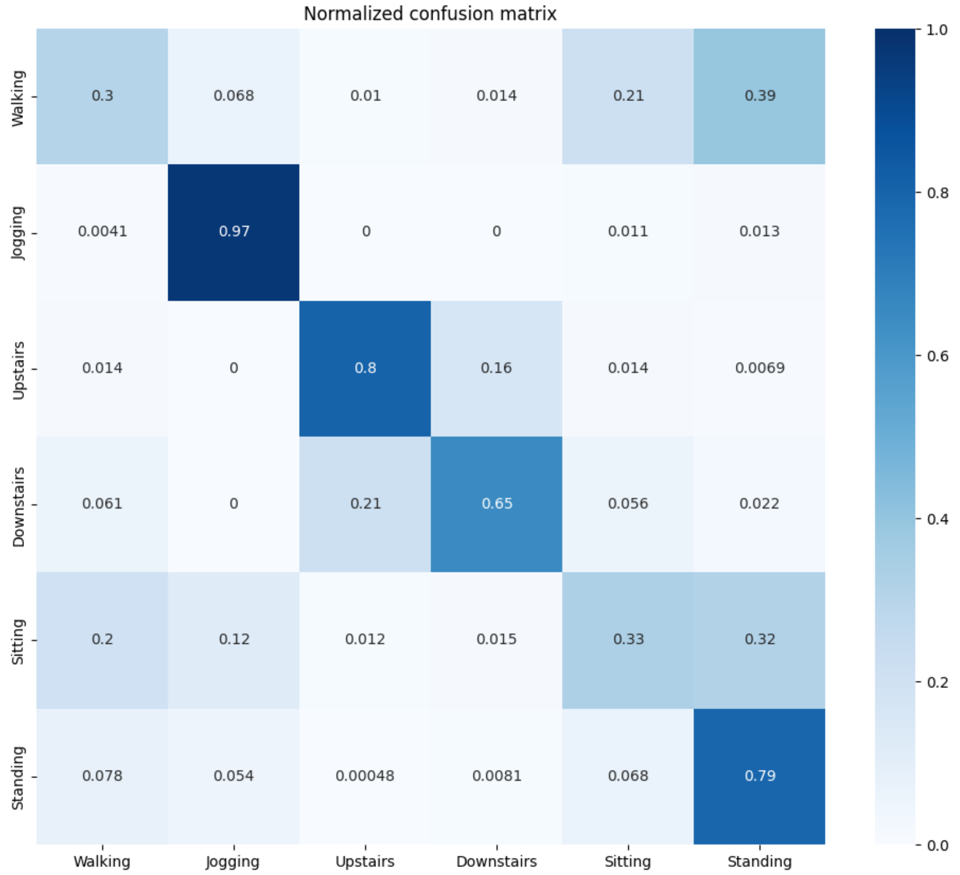
GADF



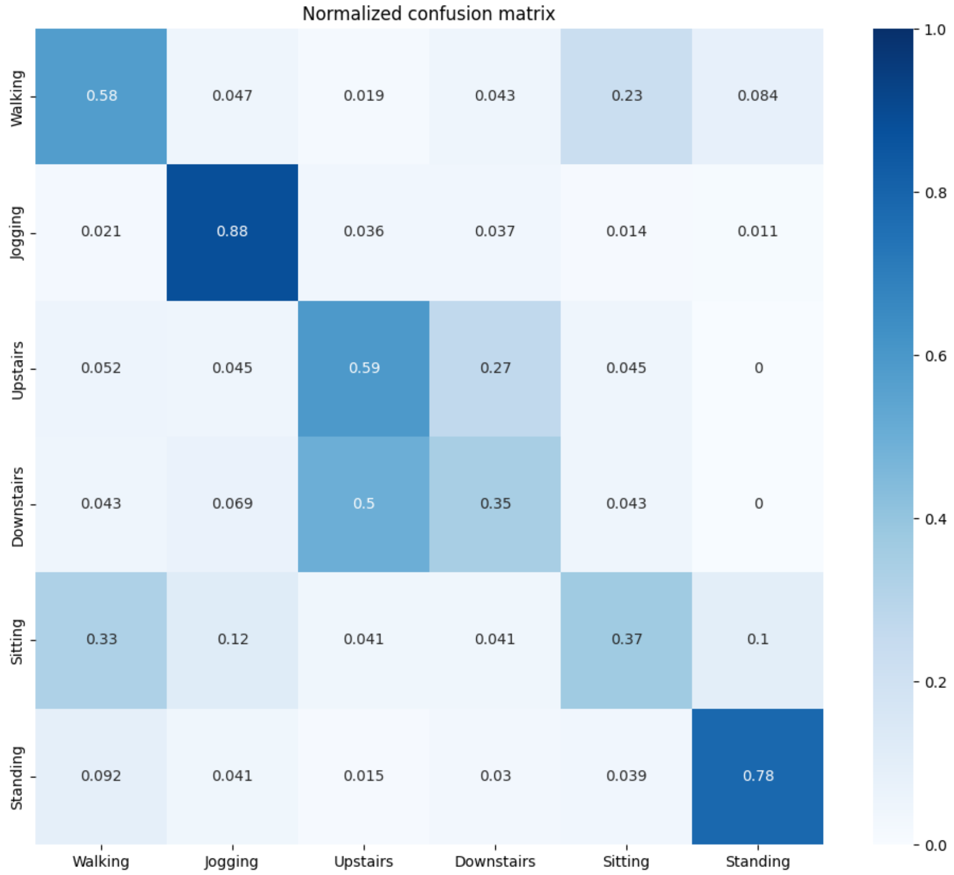
STUDY 1 – Image Encoding performance compared to raw plot

WISDM

MTF



RP



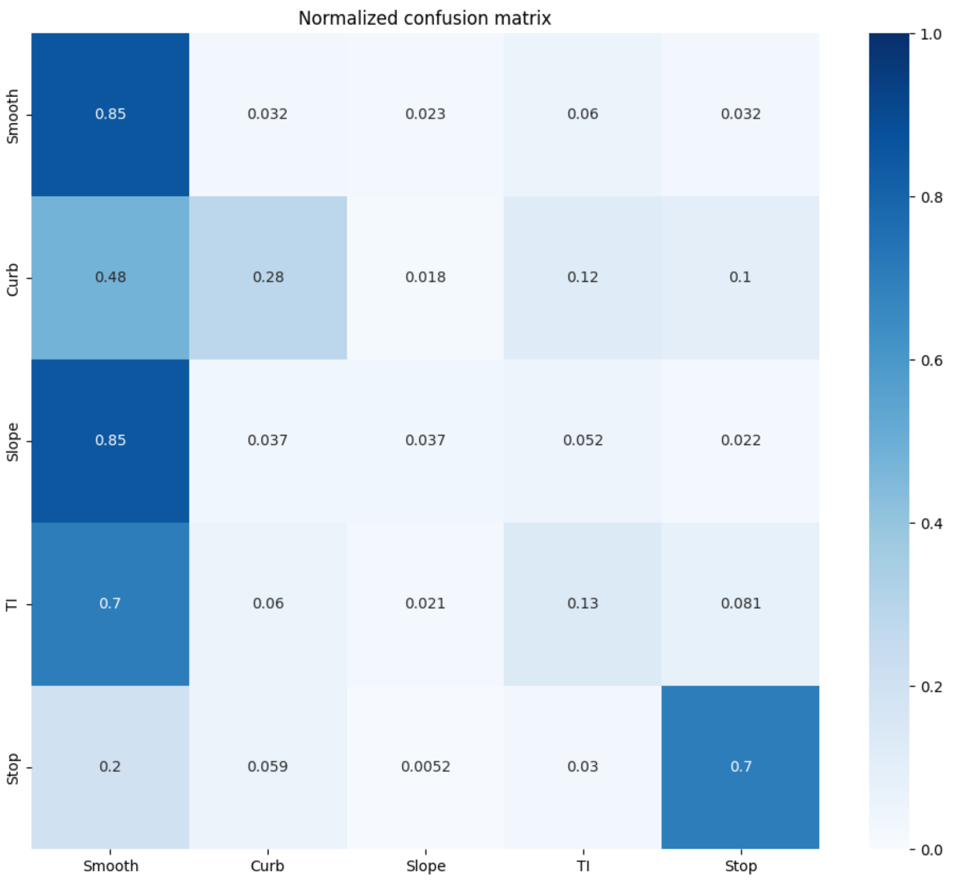
STUDY 2 – Best CNN architecture for multivariate time series

Approach	Encoding	Wheelchair		WISDM	
		Accuracy[%]	F-score[%]	Accuarcy[%]	F-score[%]
Multihead Model	GASF	52.31	50.64	78.87	78.68
Kadota	GASF	62.59	60.32	83.04	83.43
RGB Concat	GASF	59.95	59.72	73.97	72.94
Fusion	Fusion1 (GADF)	56.95	54.98	74.63	73.64

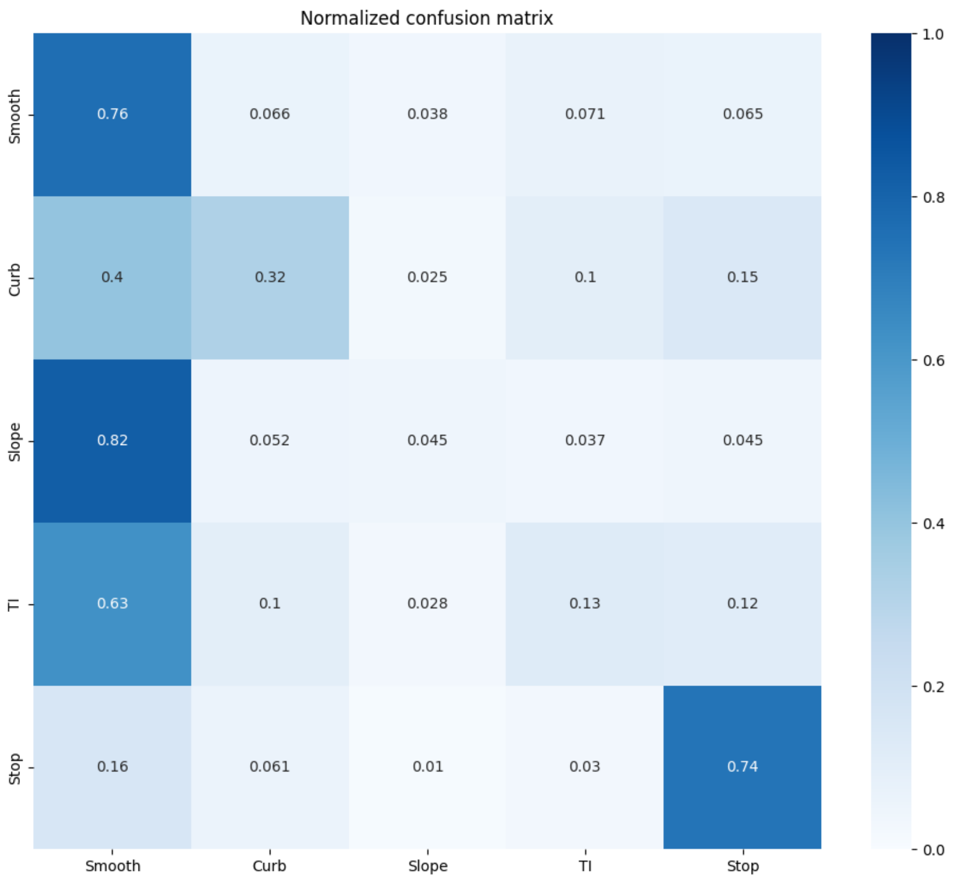
STUDY 2 – Best CNN architecture for multivariate time series

Wheelchair

Multihead



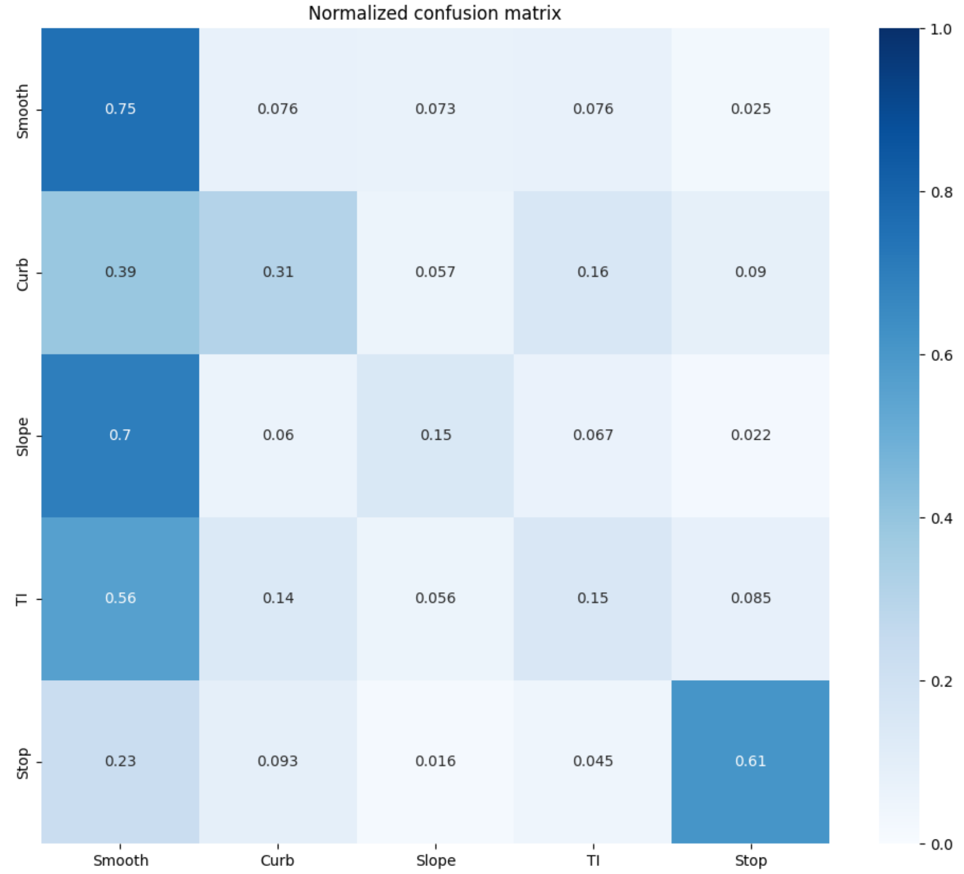
Singlehead



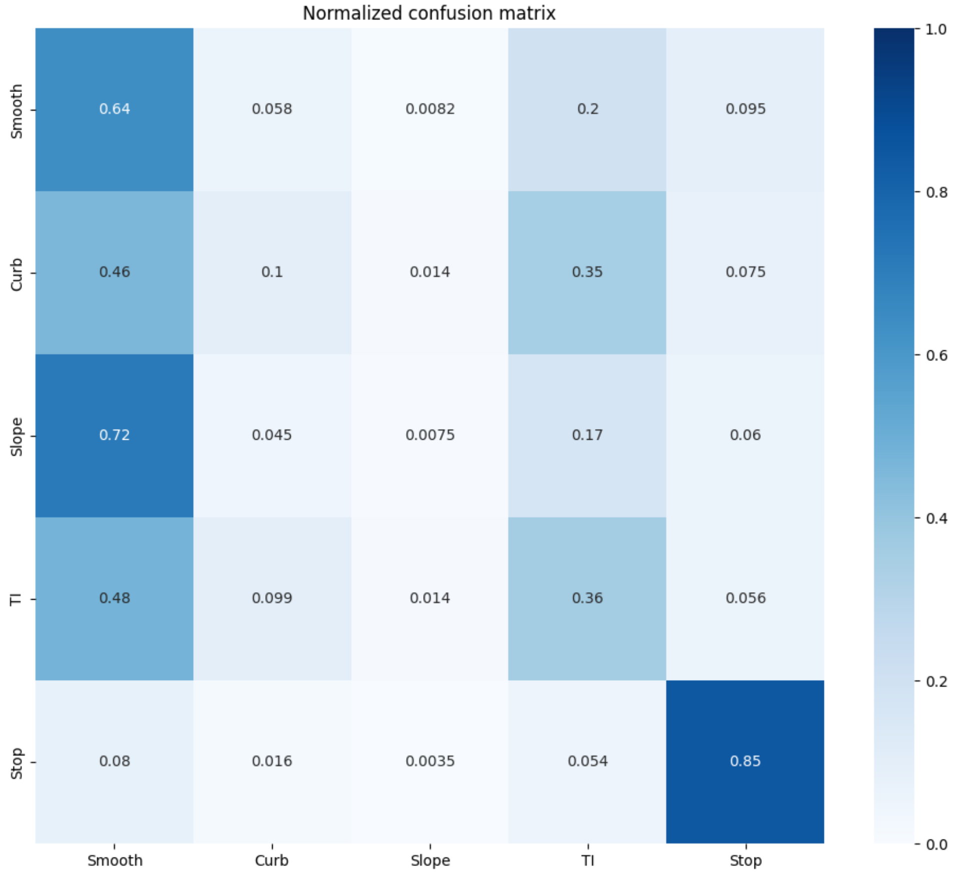
STUDY 2 – Best CNN architecture for multivariate time series

Wheelchair

RGB Concat



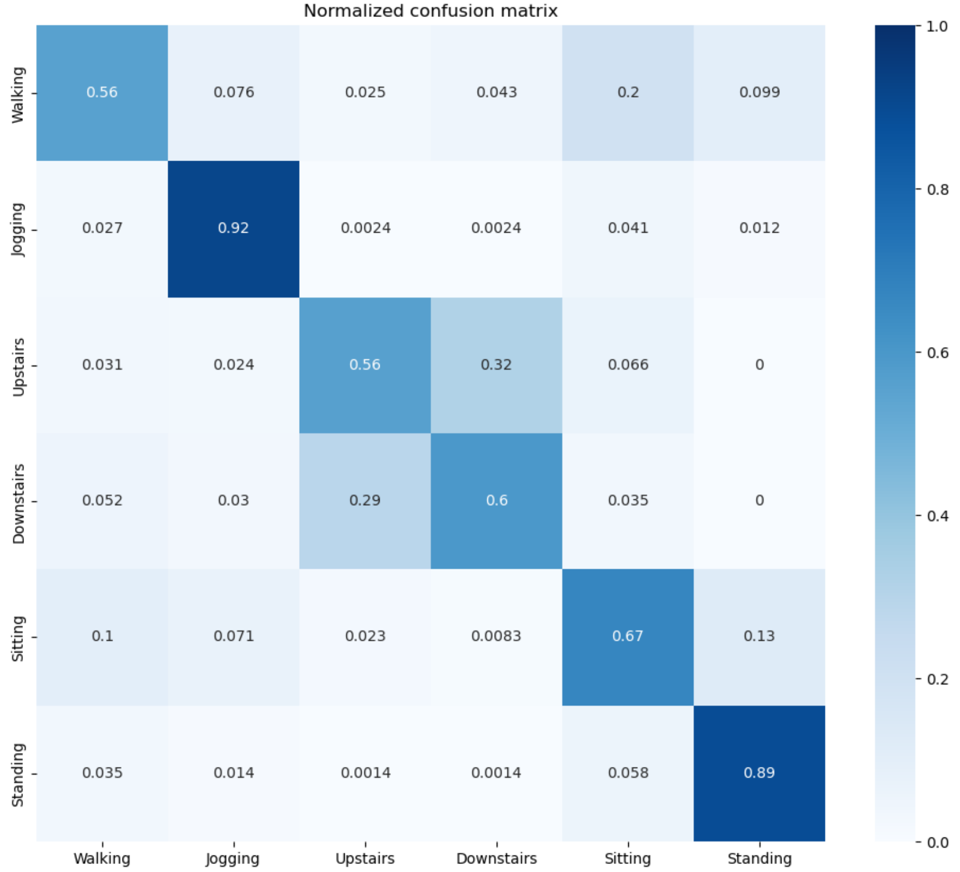
Fusion



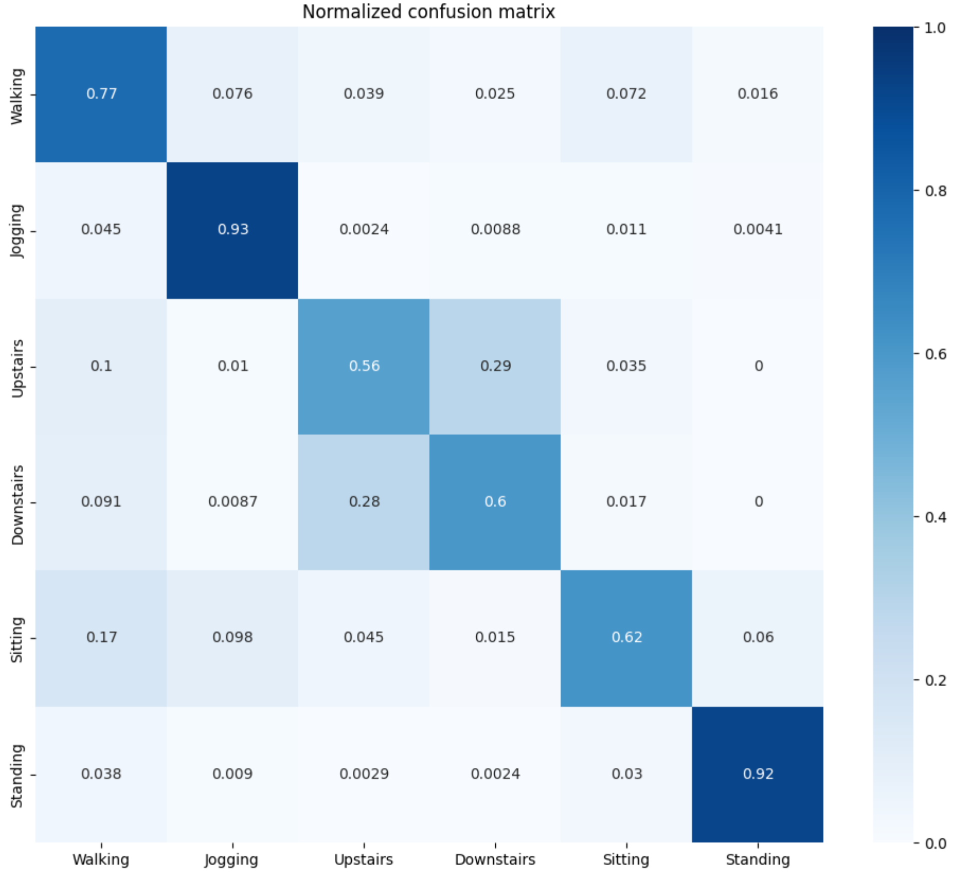
STUDY 2 – Best CNN architecture for multivariate time series

WISDM

Multihead



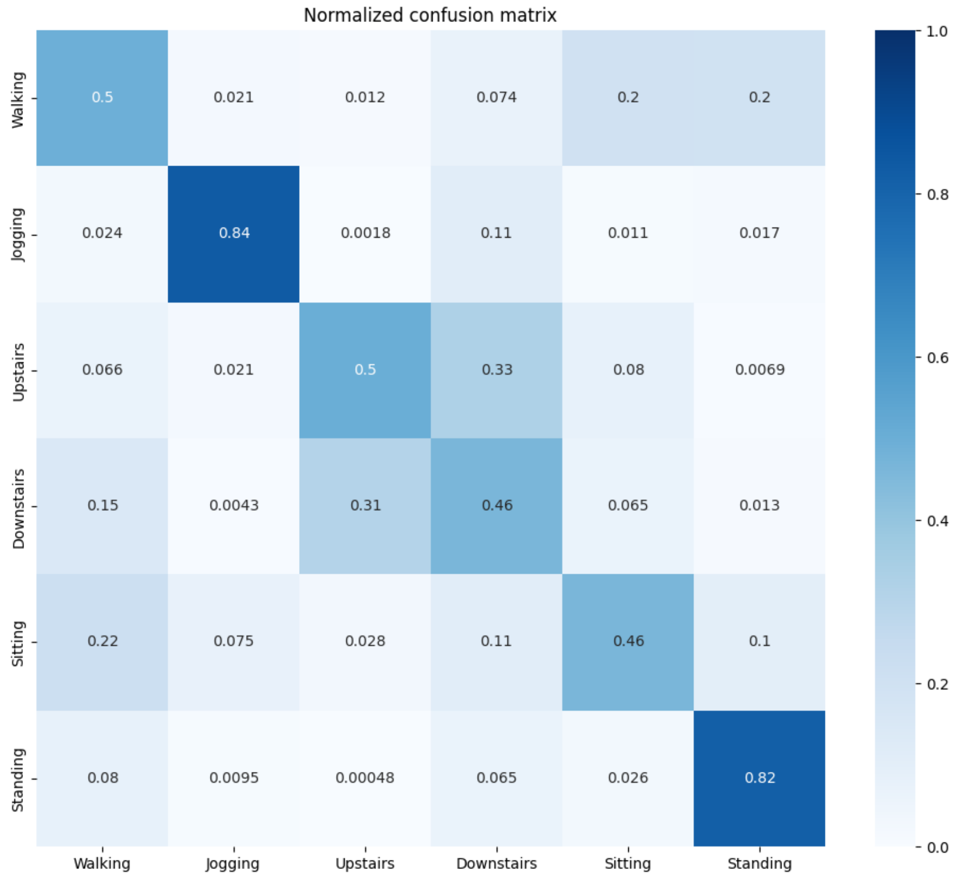
Singlehead



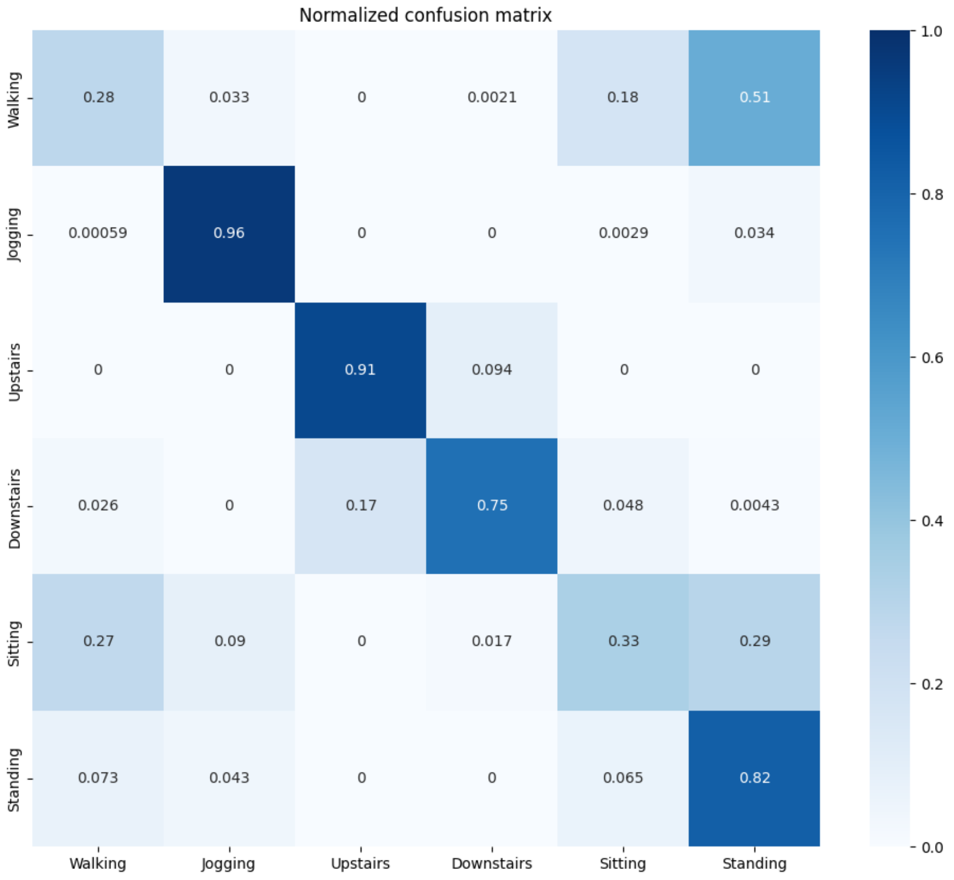
STUDY 2 – Best CNN architecture for multivariate time series

WISDM

RGB Concat



Fusion



Discussion

- WISDM is collected in controlled laboratory environment.
- Wheelchiar Acceelration is collected on the real road surface.

Smooth: 60%

Curbing Stone: 50%

Stop: 30%

⋮

Image: https://www.japan-rail-pass.com/plan-your-trip/travel-by-train/around-tokyo/yotsuya-station

Table of Contents

Conclusion

- GASF has shown the best performance compared to other image encoding methods combined with simple CNN multichannel method.

- Through various investigations of encoding multivariate time series data as images, GASF is re-confirmed to perform better than RP and MTF.

- ①Not enough amount of dataset
  ②Class imbalance
  Due to the above two issues, the noise has been amplified hindering better performance.

- Data augmentation technique could be employed in order to overcome the nature of datasets.

Thank you for listening!

Yairi Laboratory

A1777489 Dong Hyeon Park