Multimodal Sensor Data Fusion for Activity Recognition with Deep Neural Network Ensembles

Francesco Pegoraro

Supervisors: Andrew D. Bagdanov, Enrico Vicario
Correlator: Chris Nugent

Corso di Laurea Magistrale Ingegneria Informatica
Università degli Studi di Firenze
Dipartimento di Ingegneria dell'Informazione

Introduction and objectives

Introduction

Aging Population

Growth of the dependency index of 21% (2016 $\rightarrow$ 2070).

- Workforce reduction
- Decrease of mobility / increase in dependence
- Chronic and mental illness risk
- Increased Health Care costs

Introduction

Ambient Assisted Living

Aims to prolongate the time people can live in a decent way in their own home using innovative technologies.

- Increased autonomy and self-confidence
- Monitoring the elderly or ill persons
- Enhance the security and to save resources
- Delay institutionalization

Human Activity Recognition

Aims to recognize the actions and goals of one or more agents from a series of observations on the agents' actions and the environmental conditions by collecting sensors' data.

Introduction

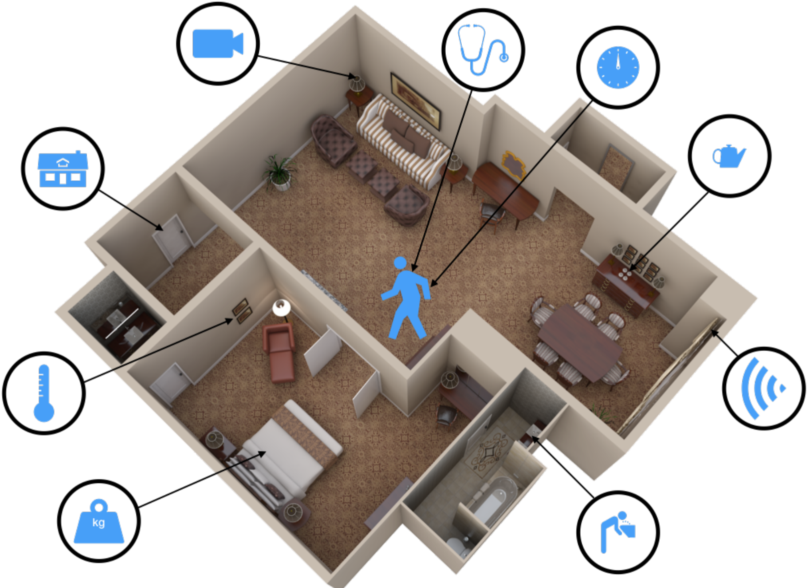
Objective

Framework for **Activity Learning** in sensor-rich environment.

- Collect data from environment and wearable sensors
- Data Processing
- Model design architecture
- Training e validation on UCAmI Cup Dataset

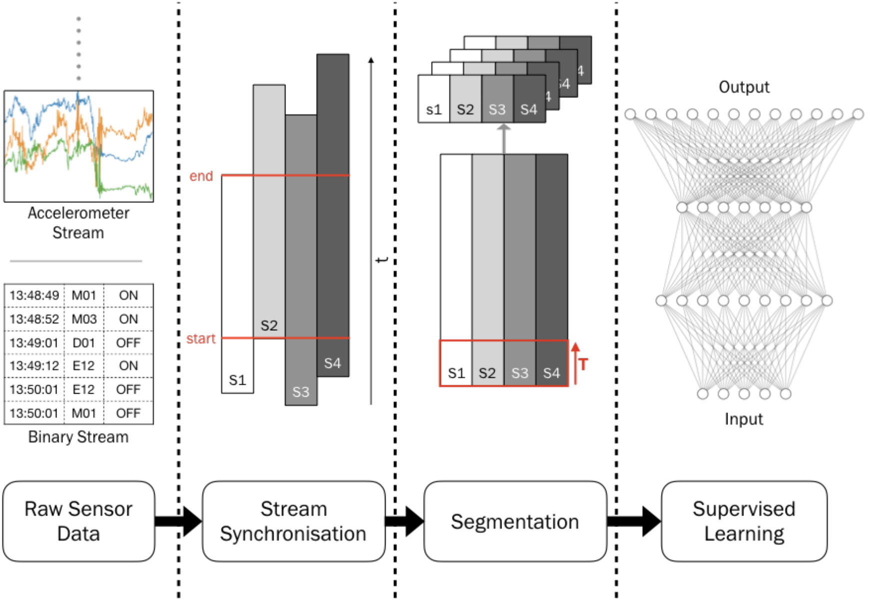
Introduction

Sensor-rich Environment

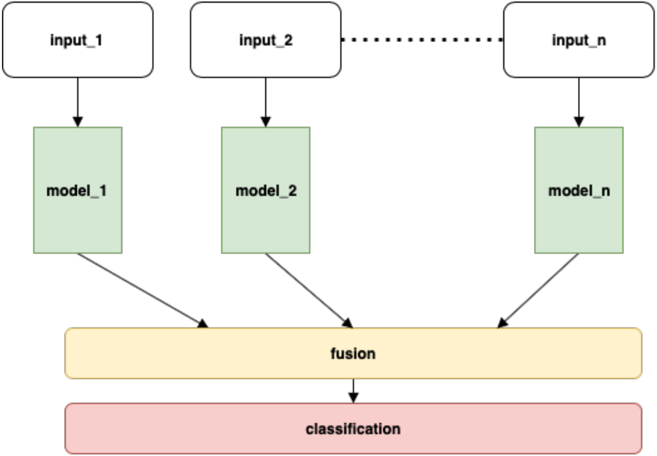


Model framework

Model Framework

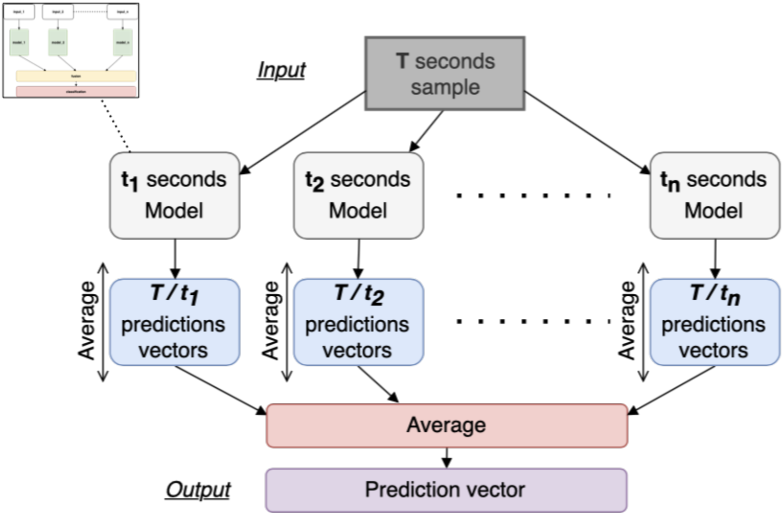


Supervised Learning - Multi Input



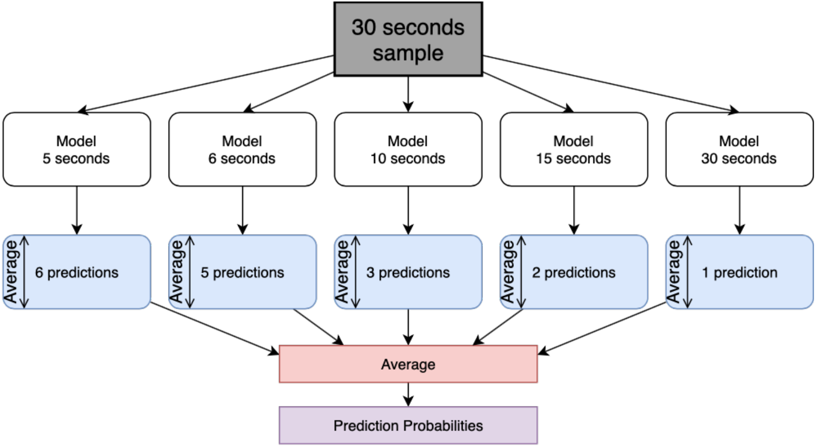
- Models for *TimeSeries*:
  - CNN (1D, 2D)
  - LSTM

Temporal Ensemble - General Case



Temporal Ensemble - Specific Case

- T = 30s
- t = [5, 6, 10, 15, 30]



Experimental Results

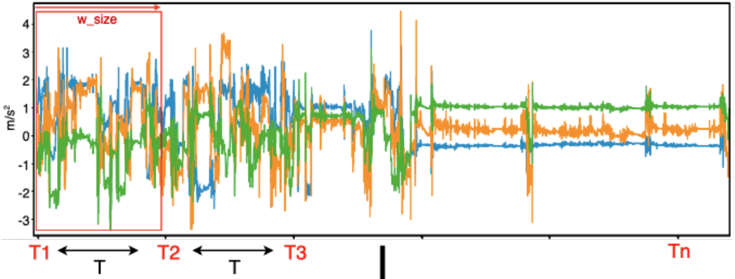
Dataset

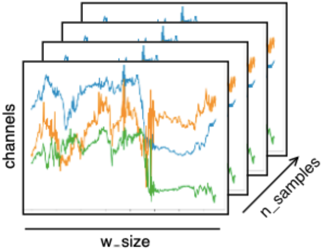
UCAmI Cup

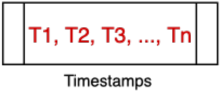
- Single inhabitant in Smart home
- 4 sources: accelerometer, binary sensors, pressure floor, proximity beacons
- 24 classes of activities

1. Medicine
2. Prep. breakfast
3. Prep. lunch
4. Prep. dinner
5. Breakfast
6. Lunch
7. Dinner
8. Snack
9. Watch TV
10. Enter
11. Play game
12. Relax sofa
13. Leave
14. Visit
15. Trash out
16. Wash hand
17. Brush teeth
18. Toilet
19. Wash dish
20. Laundry
21. Work
22. Dressing
23. Go to bed
24. Wake up

Accelerometer - Segmentation

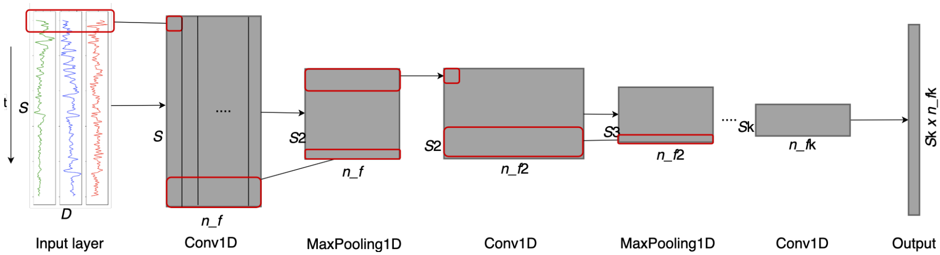






Accelerometer - Architecture

- X,y,z channels are processed by convolutional and pooling layers

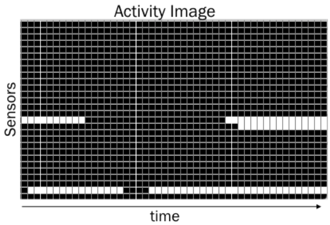


Binary Sensors - Segmentation

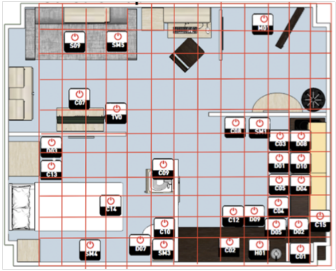
Raw Binary Stream

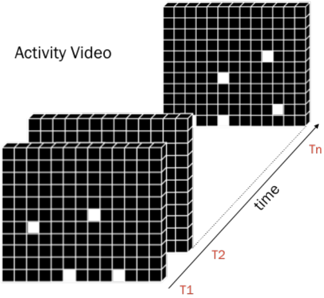
Timestamp	ID	Value
13:48:52	M3	ON
13:49:01	D1	OFF
13:49:12	E1	ON
13:50:01	E1	OFF
13:50:01	M1	OFF
...	...	...

encoding

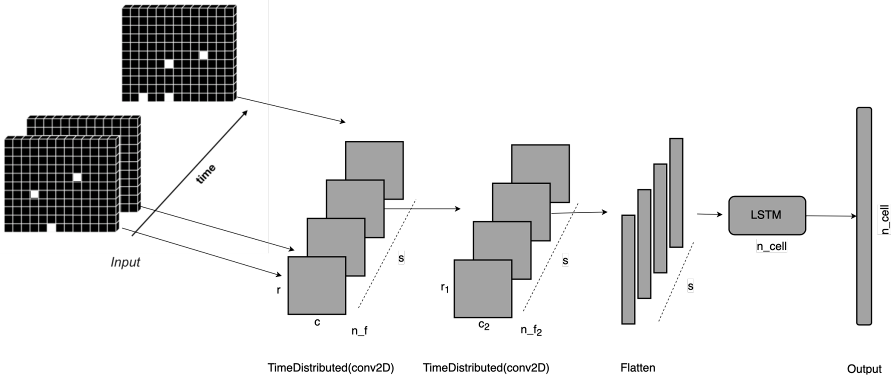


location





Binary Sensors - Architecture



- *Time Distributed*: This wrapper applies a layer to every temporal slice of an input.

Floor Sensors - Segmentation & Architecture

Pressure Stream

Timestamp	ID	Capacity
13:48:52	11	1.23
13:48:53	12	0.51
13:48:54	11	1.00
13:50:01	13	0.11
13:50:02	10	0.09
...	...	...

windowing

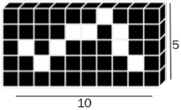
Timestamp	ID	Capacity
13:48:52	11	1.23
13:48:53	12	0.51
13:48:54	11	1.00
13:50:01	13	0.11
13:50:02	10	0.09
...	...	...

T1

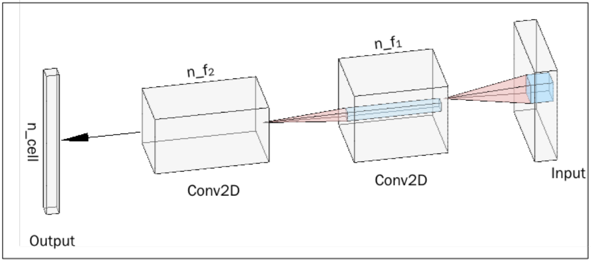
T2

Tn

location



Architecture



- *No Pooling*: CNN location variant

Proximity Sensor- Segmentation & Architecture

Proximity Stream

Timestamp	Obj	RSSI
13:48:52	Tap	-84
13:48:53	Bed	-70
13:48:54	Tap	-80
13:50:01	Pot	-60
13:50:02	Pot	-65
...	...	...

windowing

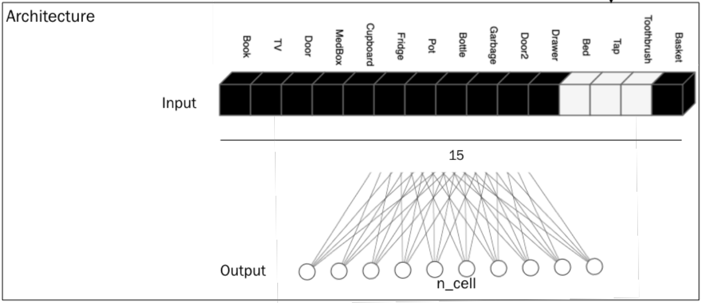
Timestamp	Obj	RSSI
13:48:52	Tap	-84
13:48:53	Bed	-70
13:48:54	Tap	-80
13:50:01	Pot	-60
13:50:02	Pot	-65
...	...	...

T1

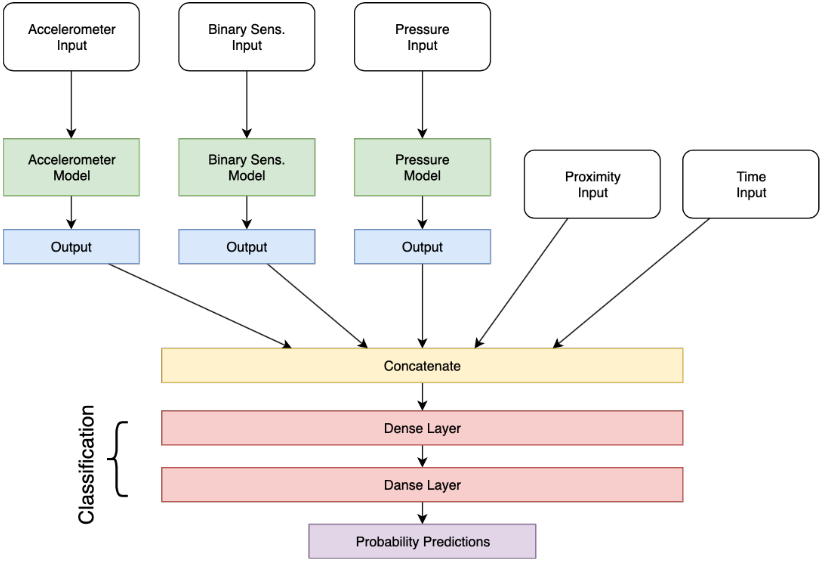
T2

Tn

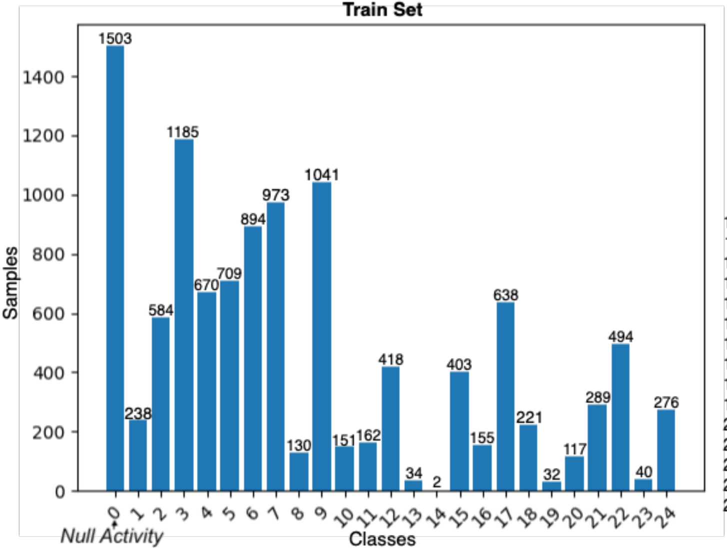
encoding



Model - Multi Input



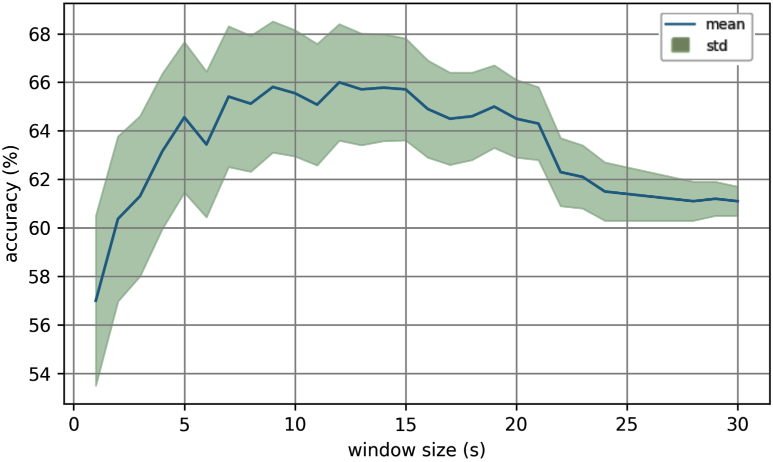
TrainSet - Histogram with T=5s



1. Take medication
2. Prepare breakfast
3. Prepare lunch
4. Prepare dinner
5. Breakfast
6. Lunch
7. Dinner
8. Eat a snack
9. Watch TV
10. Enter the SmartLab
11. Play a video game
12. Relax on the sofa
13. Leave the SmartLab
14. Visit in the SmartLab
15. Put waste in the bin
16. Wash hands
17. Brush teeth
18. Use the toilet
19. Wash dishes
20. Laundry
21. Work at the table
22. Dressing
23. Go to the bed
24. Wake up

Results - Cross Validation

- 5-fold on TrainSet

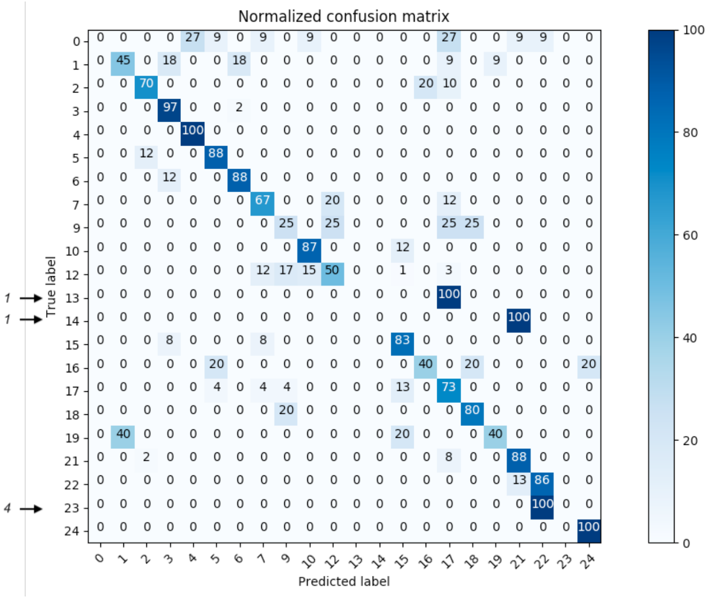


Results

$$F_1 = \sum_i 2 \times \omega_i \frac{precision_i \cdot recall_i}{precision_i + recall_i} \tag{1}$$

Method	**Accuracy**	**F-1 Score**
Single model *(T=5s)* w/Activity 0	55%	0.5212
Single model *(T=5s)*	63%	0.6543
Average Ensemble model w/Activity 0 *T=5,6,10,15,30*	64%	0.6237
Average Ensemble model *T=5,6,10,15,30*	**73.9%**	**0.7567**

Results


Normalized confusion matrix
True label
Predicted label
1
1
4

Challenge

- Other participants results

Method	Accuracy
Markov Model + NN	45%
Random Forest	47%
NeuralNetwork	60.10%
Naive Bayes Classifier	60.5%
Average Ensemble model ***T=5,6,10,15,30***	73.9%
Finite Automata	**89**%

Conclusions

Conclusions

In this work

- Generic Framework
- Scalable/Modular
- Does not require *feature engineering*

Future work

- Train at the same time the single models in the Temporal Ensemble
- Collect/Test balanced dataset for HAR with more data

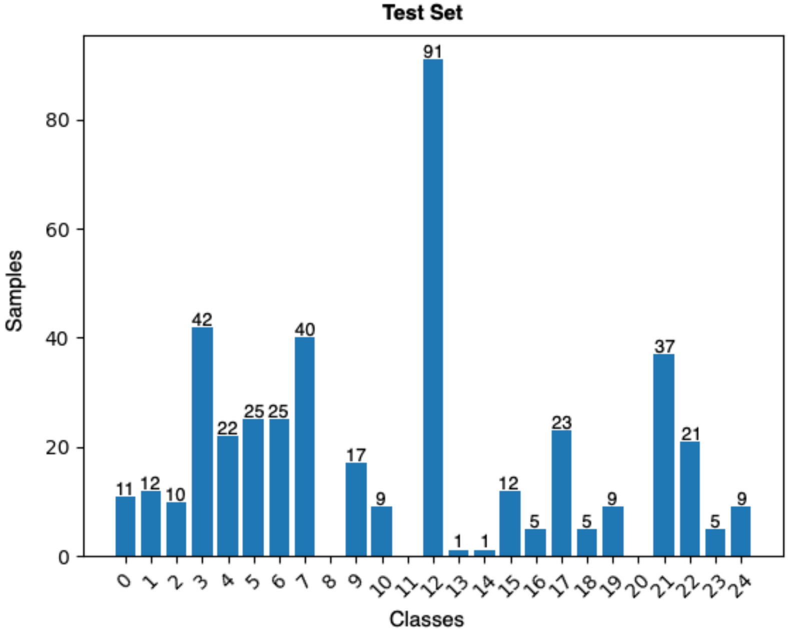
Multimodal Sensor Data Fusion for Activity Recognition with Deep Neural Network Ensembles

Francesco Pegoraro

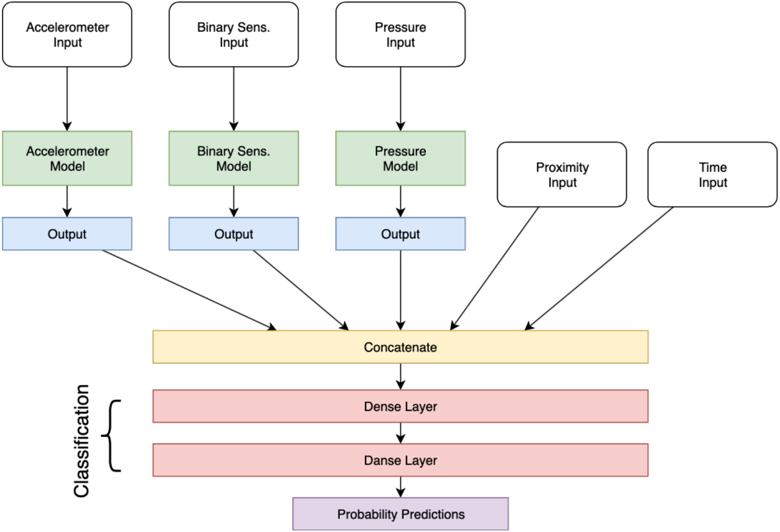
Supervisors: Andrew D. Bagdanov, Enrico Vicario
Correlator: Chris Nugent

Corso di Laurea Magistrale Ingegneria Informatica
Università degli Studi di Firenze
Dipartimento di Ingegneria dell'Informazione

TestSet - Histogram with T=30s

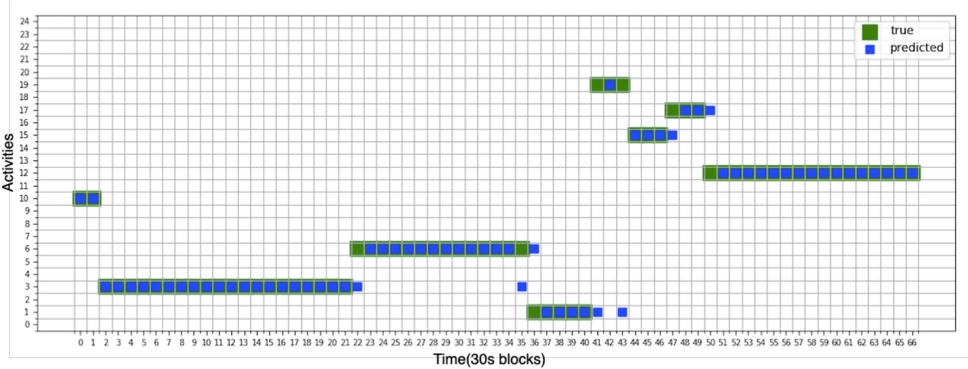


Complete Model



$$time\_input = \left[\sin\left(\frac{2\pi \times \text{hour}}{24}\right), \cos\left(\frac{2\pi \times \text{hour}}{24}\right)\right]$$

Sequenza predizioni



- Criticality at the extremes of the activities