Sleep Study
– Detecting sleep in insomniacs

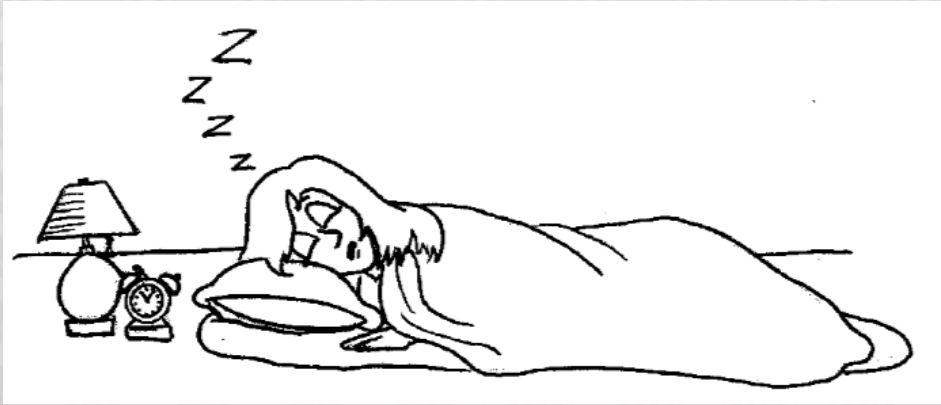
Presenter:
Qichan Feng

Agenda

- Project background
- Data preparation
- ML models and evaluation
- Conclusion and future work

Project background

- Are you sleeping well?



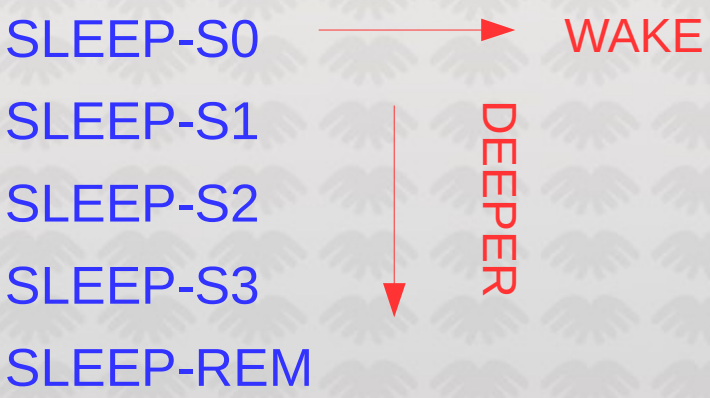
- Chronic insomnia can affect mood, concentration, memory and work performance
- Sleep staging is fundamental to insomnia treatment

Woolcock Institute

- Specialising in sleep research
- Sleep clinic
- Data collected for ~70 insomnia patients

Gold standard for sleep staging

Sleep staging through PSG recordings

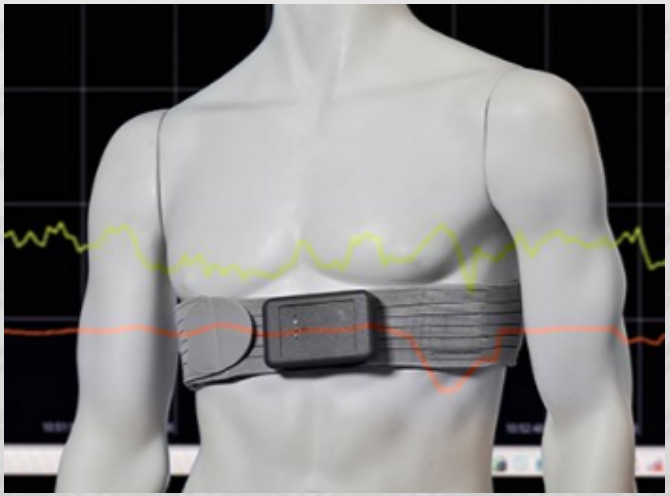


Disadvantages:

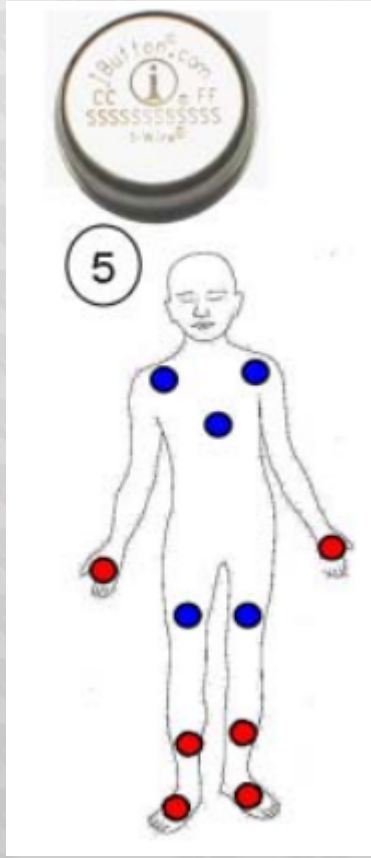
- Not portable
- Supervised
- Interruptive

Alternative measurements

Heart rate



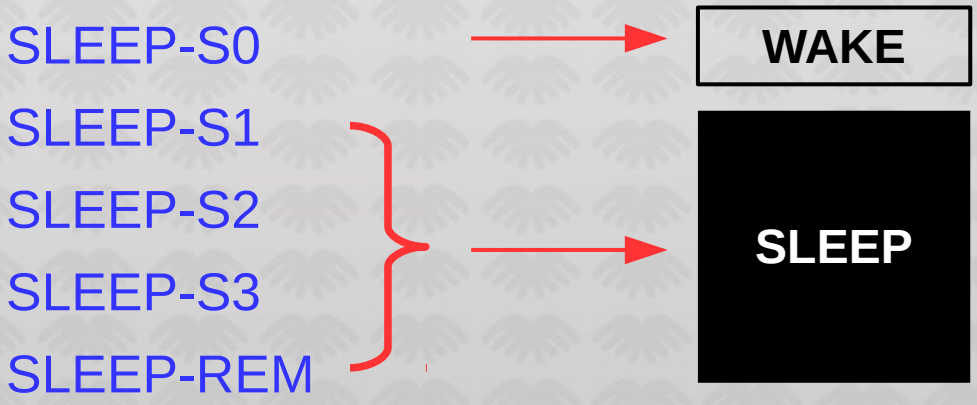
Skin temperature



Actigraphy

Goal of the project

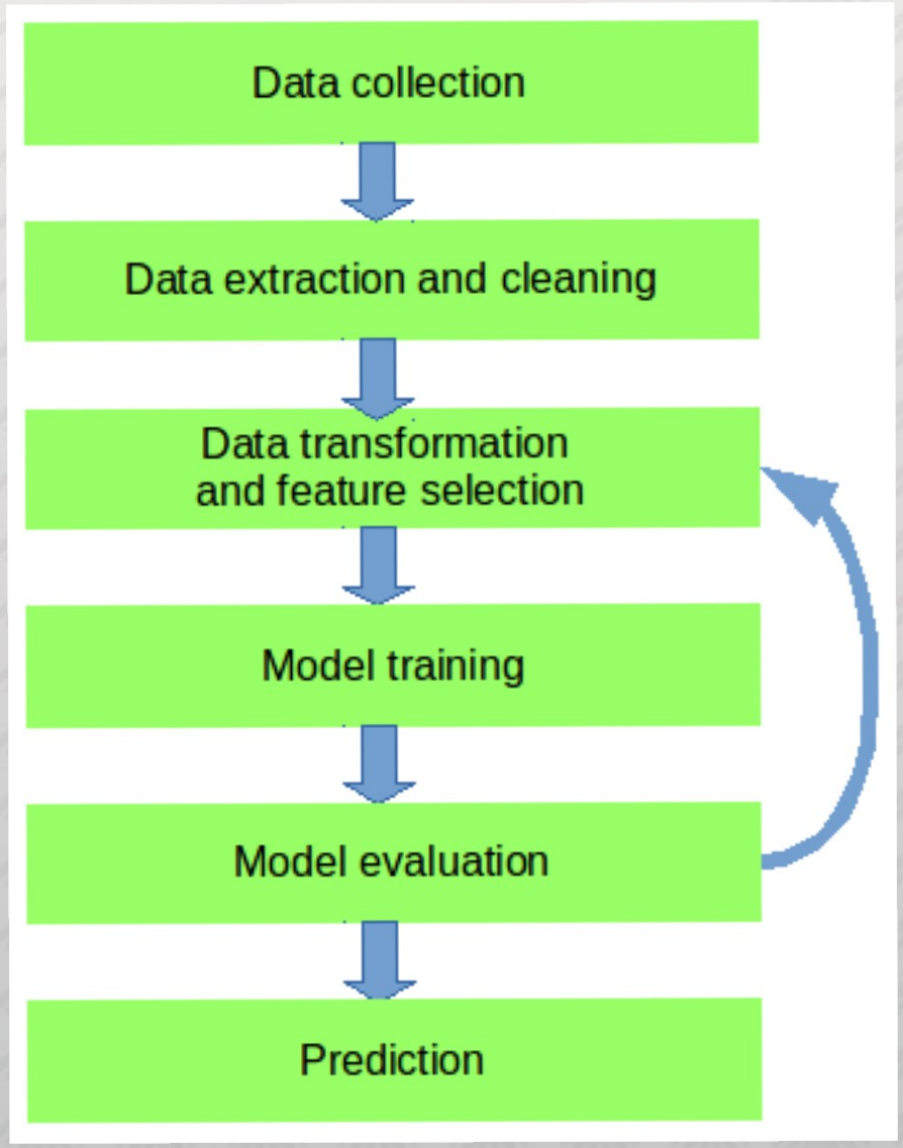
Apply machine learning methodologies to classify between wake and sleep with alternative measurements



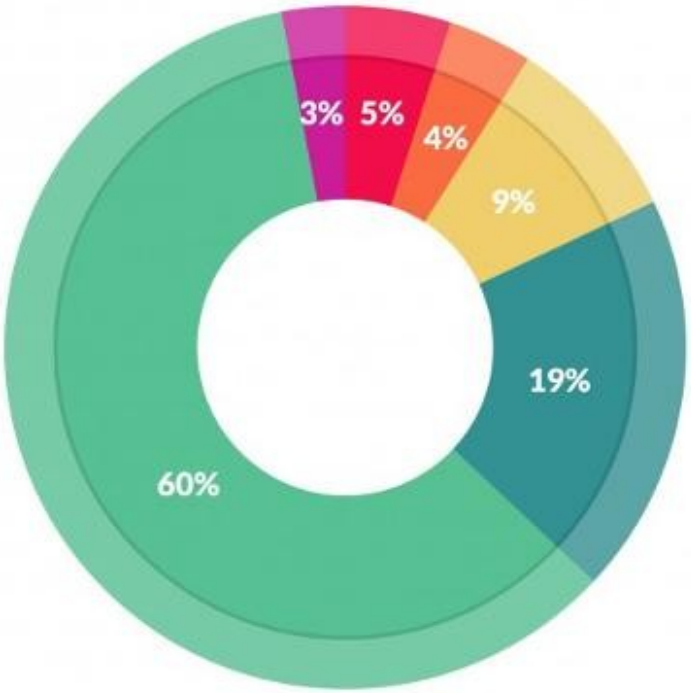
Python data frameworks

- Pandas
- Scikit-learn
- XGBoost
- Theano / Keras
- Matplotlib
- Jupyter notebook

Classification pipeline



Data preparation



What data scientists spend the most time doing

- *Building training sets: 3%*
- *Cleaning and organizing data: 60%*
- *Collecting data sets; 19%*
- *Mining data for patterns: 9%*
- *Refining algorithms: 4%*
- *Other: 5%*

http://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says

Data collection and cleaning

The real-world data are **MESSY!**

- Various data sources

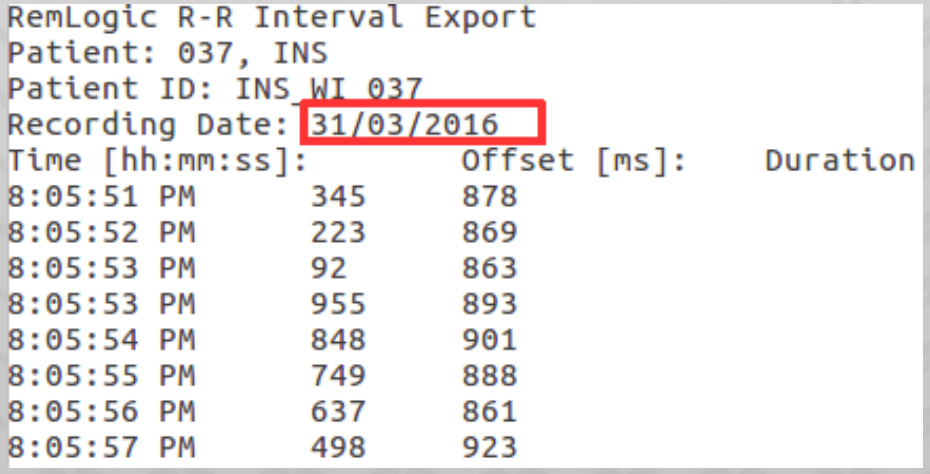
- Different file formats

- Data with odd-formats

2/09/2015 19:59	Log Values		L
2/09/2015 19:59	Reading	Values	R
2/09/2015 19:59			
2/09/2015 19:59	2/09/2015 19:59	34.402	
2/09/2015 20:00	2/09/2015 19:59	34.339	
2/09/2015 20:00	2/09/2015 19:59	34.339	
2/09/2015 20:00	2/09/2015 19:59	34.339	
2/09/2015 20:00	2/09/2015 20:00	34.402	
2/09/2015 20:01	2/09/2015 20:00	34.464	
2/09/2015 20:01	2/09/2015 20:00	34.526	
2/09/2015 20:01	2/09/2015 20:00	34.526	

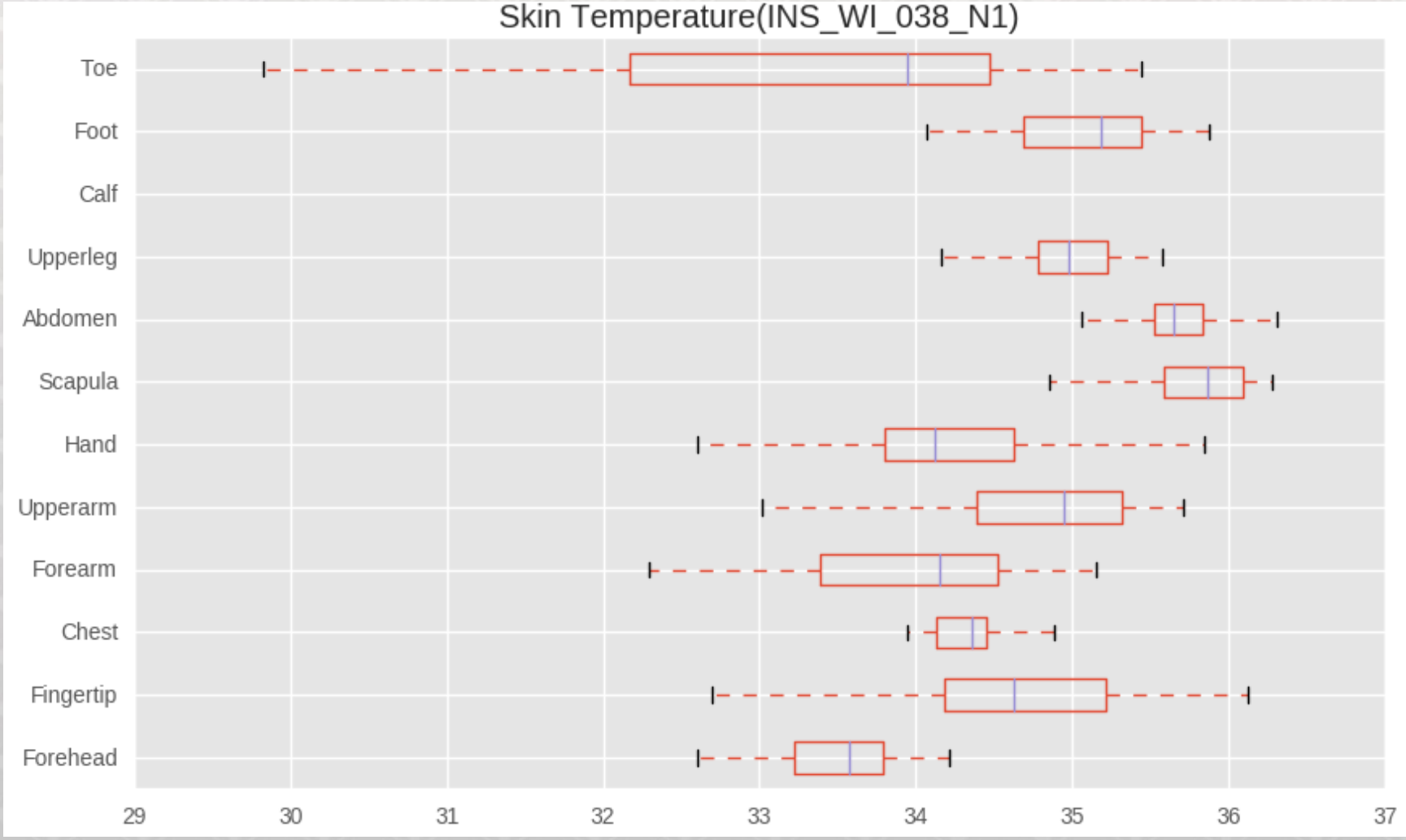
Challenges of time-series data

- Uneven lengths
  - Sleep staging / Actigraphy: 30s
  - Skin temparature: 15s
  - Heart rate: beat-by-beat interval (ms)
- Irregular sampling times
- Interruption of data
- Next day flipping over



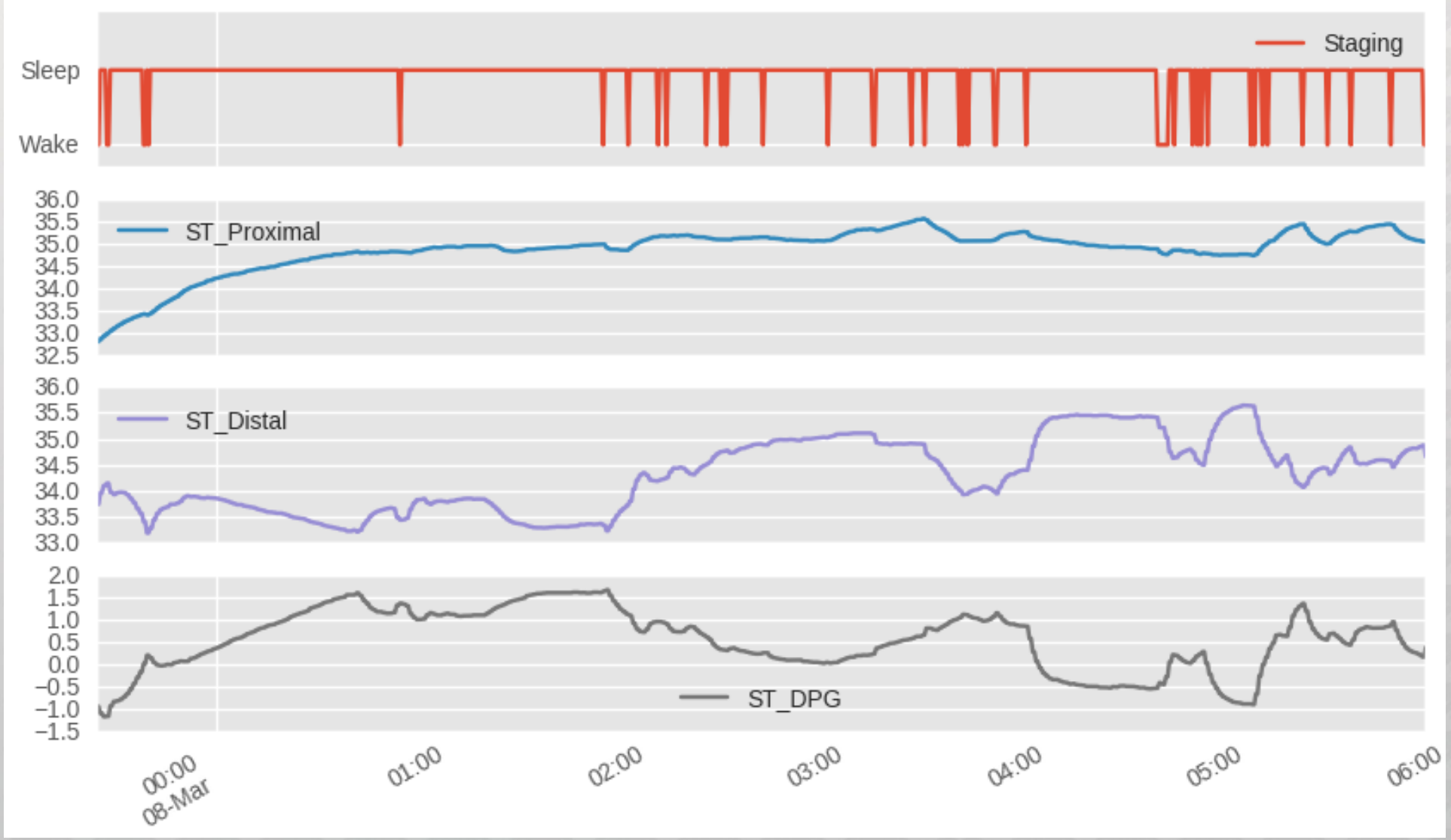
```
RemLogic R-R Interval Export
Patient: 037, INS
Patient ID: INS_WI_037
Recording Date: 31/03/2016
Time [hh:mm:ss]:        Offset [ms]:    Duration
8:05:51 PM      345     878
8:05:52 PM      223     869
8:05:53 PM      92      863
8:05:53 PM      955     893
8:05:54 PM      848     901
8:05:55 PM      749     888
8:05:56 PM      637     861
8:05:57 PM      498     923
```

Features – skin temparature



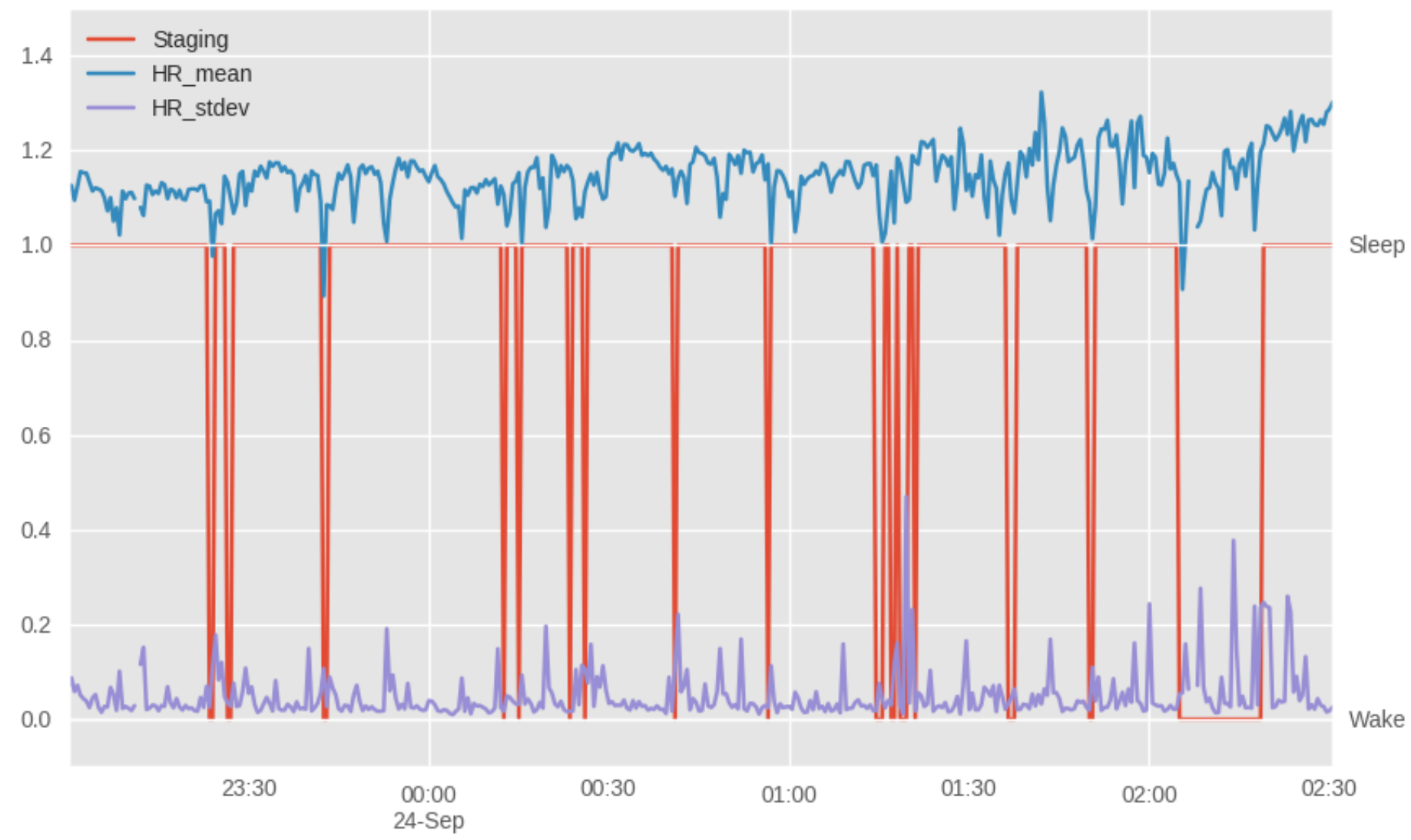
- **Proximal: abdomen, chest, upperarm, upperleg**
- **Distal: fingertip, toe, hand, foot**
- **DPG: gradient between the above**

Features - skin temperature



Features – Heart rate

- Aggregate into 30s epoch
- Calculate mean and stdev



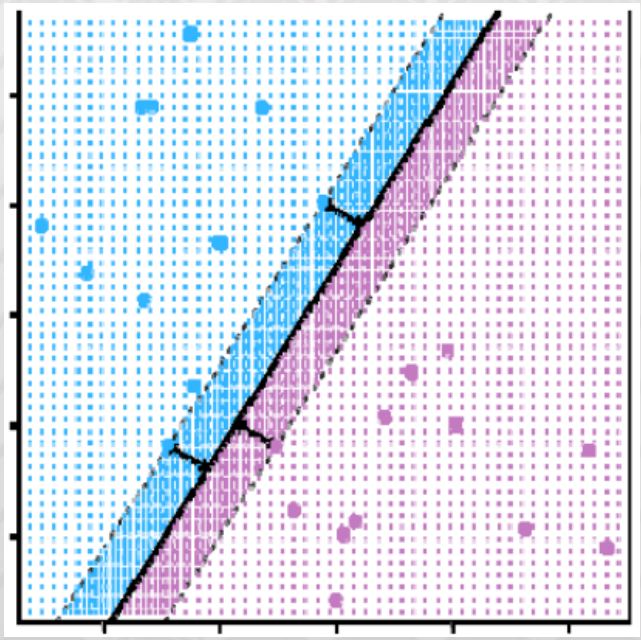
Cleaned datasets

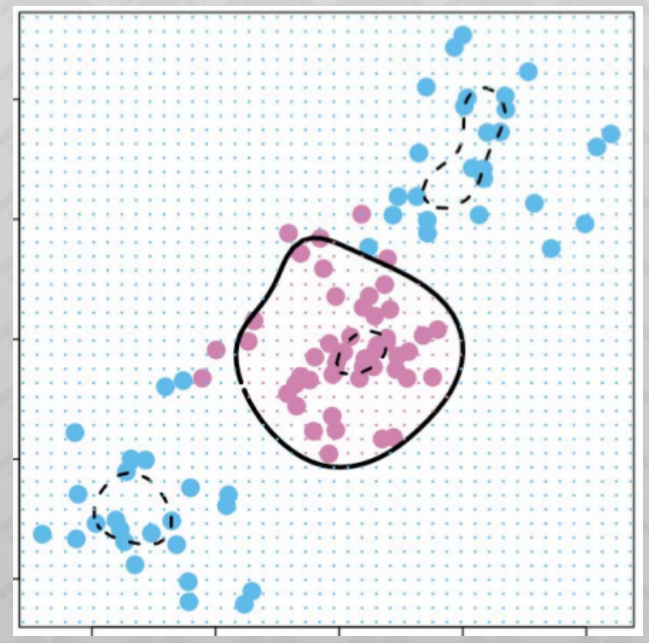
- 20 patients with Actigraphy, 19,369 samples
- 59 patients without Actigraphy, 56,174 samples

				mess format				
				device not attached				not to process
Patient_ID	Staging N1	HR N1	ACT N1	ST N1	Staging Samples	HR Samples	Diff	Notes
INS_WI_001	2015-09-02	2015-09-02	Not available	2015-09-02	889	801	88	
INS_WI_002	2015-09-16	2015-09-16	Not available	2015-09-16	1033	1017	16	
INS_WI_003	2015-09-23	2015-09-23	Not available	2015-09-23	987	982	5	
INS_WI_004	2015-09-09	2015-09-09	2015-09-09	2015-09-09	1287	1198	89	
INS_WI_005	2015-09-09	2015-09-09	Not available	2015-09-09				Toe too mess, not to process
INS_WI_006	2015-09-16	2015-09-16	Not available	2015-09-16	1026	1020	6	
INS_WI_007	2016-02-18	2016-02-18	2016-02-18	2016-02-18	844	843	1	
INS_WI_008	2015-09-23	2015-09-23	2015-09-23	2015-09-23	924	907	17	
INS_WI_009	2015-10-08	2015-10-08	Not available	2015-10-08	1074	1057	17	
INS_WI_011	2015-10-01	2015-10-01	2015-10-01	2015-10-01	729	452	277	
INS_WI_012	2015-10-29	2015-10-29	2015-10-29	2015-10-29	1024	985	39	Toe missing
INS_WI_013	2015-11-04	2015-11-04	2015-11-04	2015-11-04	919	910	9	
INS_WI_014	2015-10-29	2015-10-29	Not available	2015-10-29	1048	897	151	
INS_WI_015	2015-11-04	2015-11-04	Not available	2015-11-04	1002	877	125	
INS_WI_017	2015-12-16	2015-12-16	2015-12-16	2015-12-16	877	875	2	
INS_WI_018	2015-12-09	2015-12-09	Not available	2015-12-09	1064	895	169	
INS_WI_019	2015-12-02	2015-12-02	2015-12-02	2015-12-02	1129	609	520	
INS_WI_020	2015-12-09	2015-12-09	2015-12-09	2015-12-09	1049	846	203	
INS_WI_021	2016-01-20	2016-01-20	2016-01-20	2016-01-20	999	998	1	
INS_WI_022	2016-01-13	2016-01-13	2016-01-13	2016-01-13	780	685	95	
INS_WI_023	2016-04-13	2016-04-13	Not available	2016-04-13	1041	1031	10	
INS_WI_024	2016-04-06	2016-04-06	Not available	Not available				N1 is actually N2, not to proces
INS_WI_025	2016-02-03	2016-02-03	2016-02-03	2016-02-03	900	900	0	
INS_WI_026	2016-02-03	2016-02-03	Not available	2016-02-03	852	852	0	
INS_WI_028	2016-03-02	2016-03-02	Not available	2016-03-02	696	696	0	
INS_WI_030	2016-02-10	2016-02-10	2016-02-10	2016-02-10	1009	790	219	

Support Vector Machine (SVM)

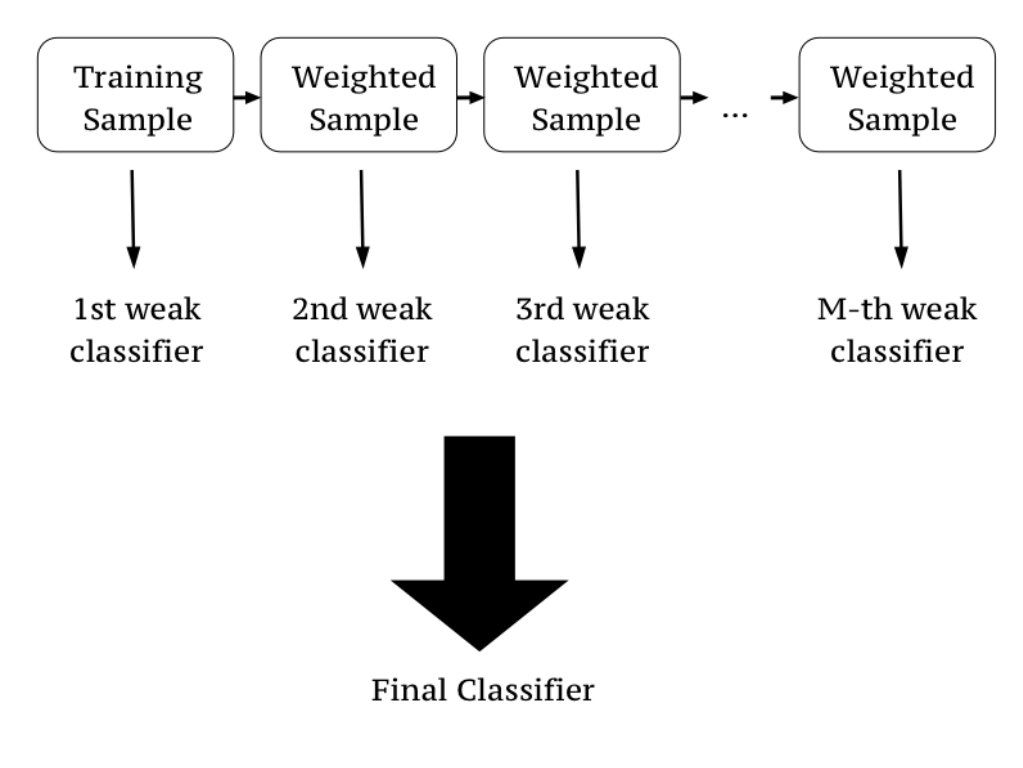
- Hyperplane with maximal margin
- Use kernel trick for non-linear hyperplane
- Forward / backward fill and standardise features





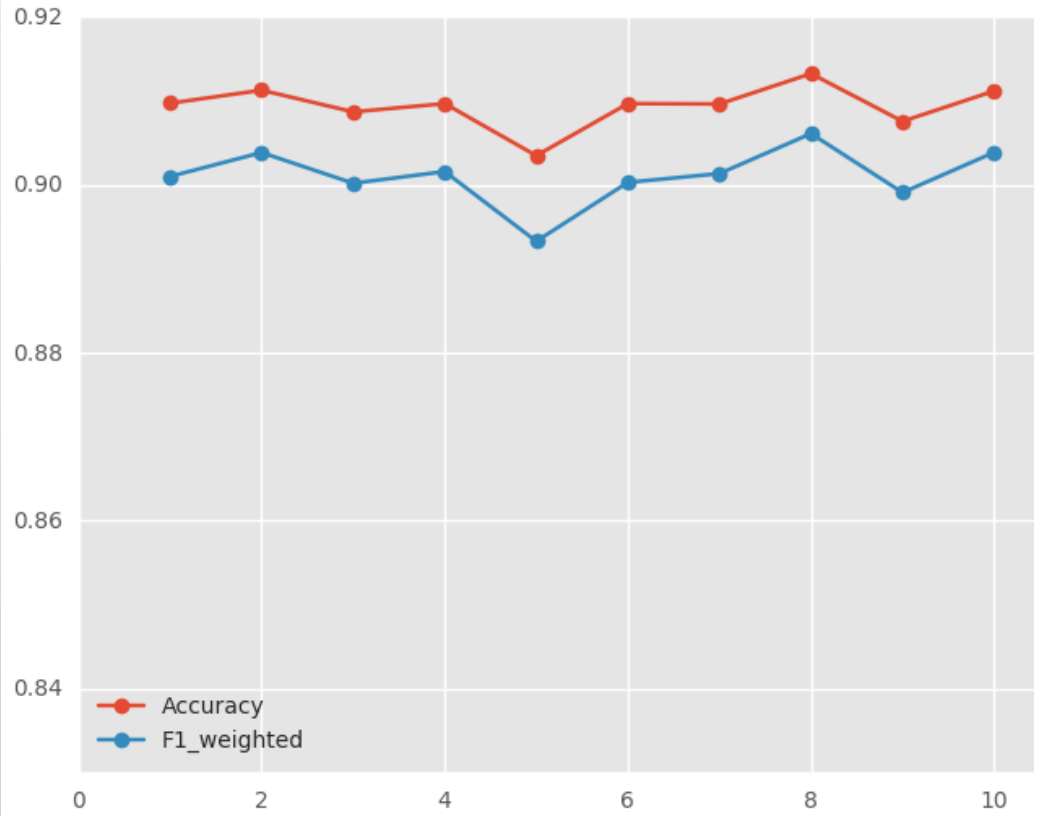
XGBoost: Gradient Booting Model

- Tree based ensemble method
- Speed and performance
- Handle missing data
- Tunable parameters



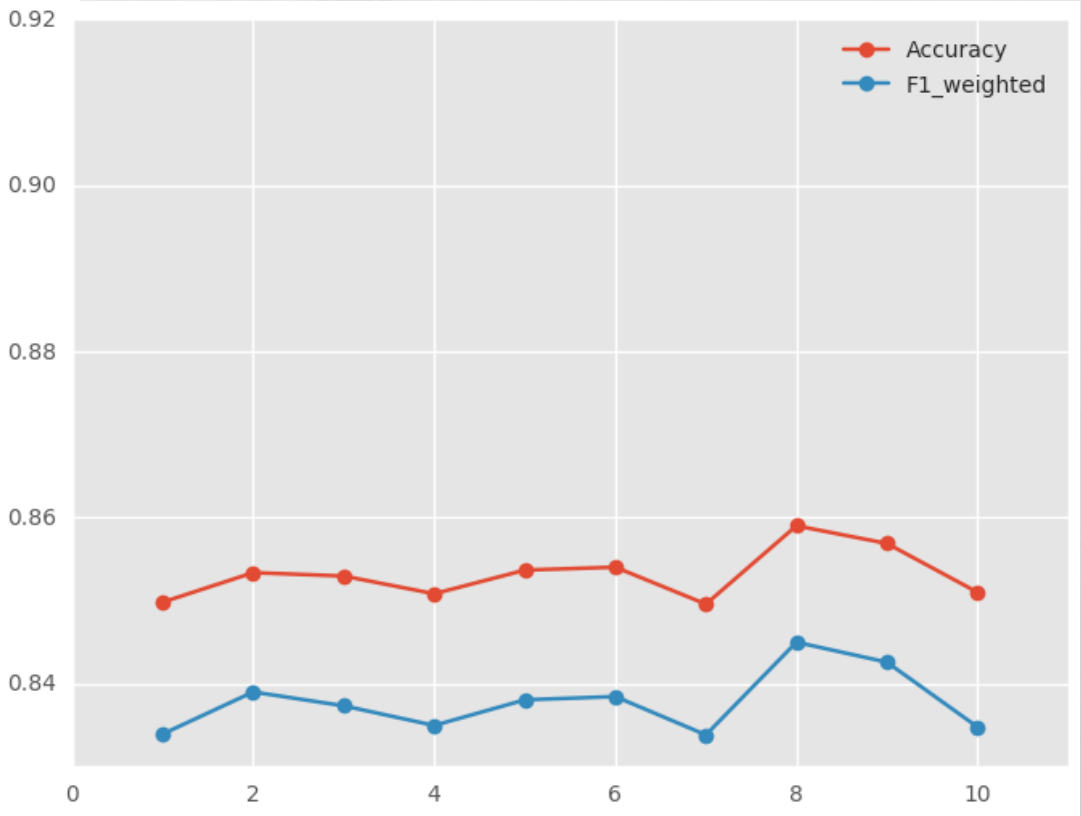
SVM – 10 fold CV

20 subjects



Avg Accuracy	0.9093
Avg F1 score	0.9010

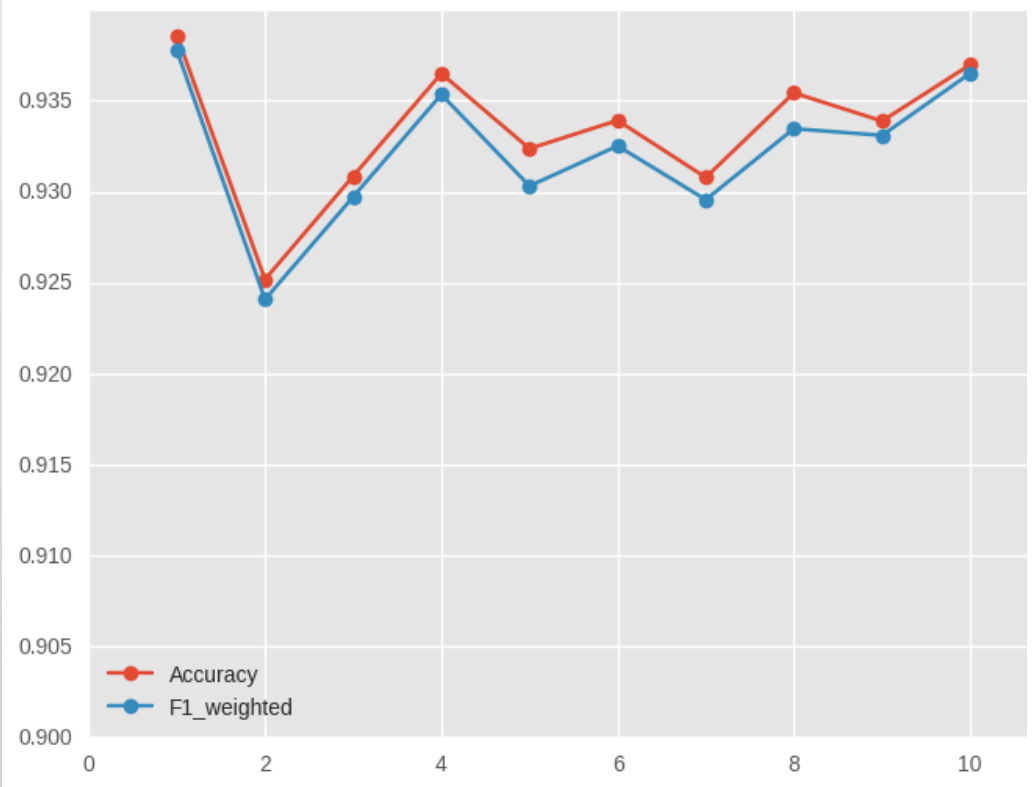
59 subjects



Avg Accuracy	0.8531
Avg F1 score	0.8378

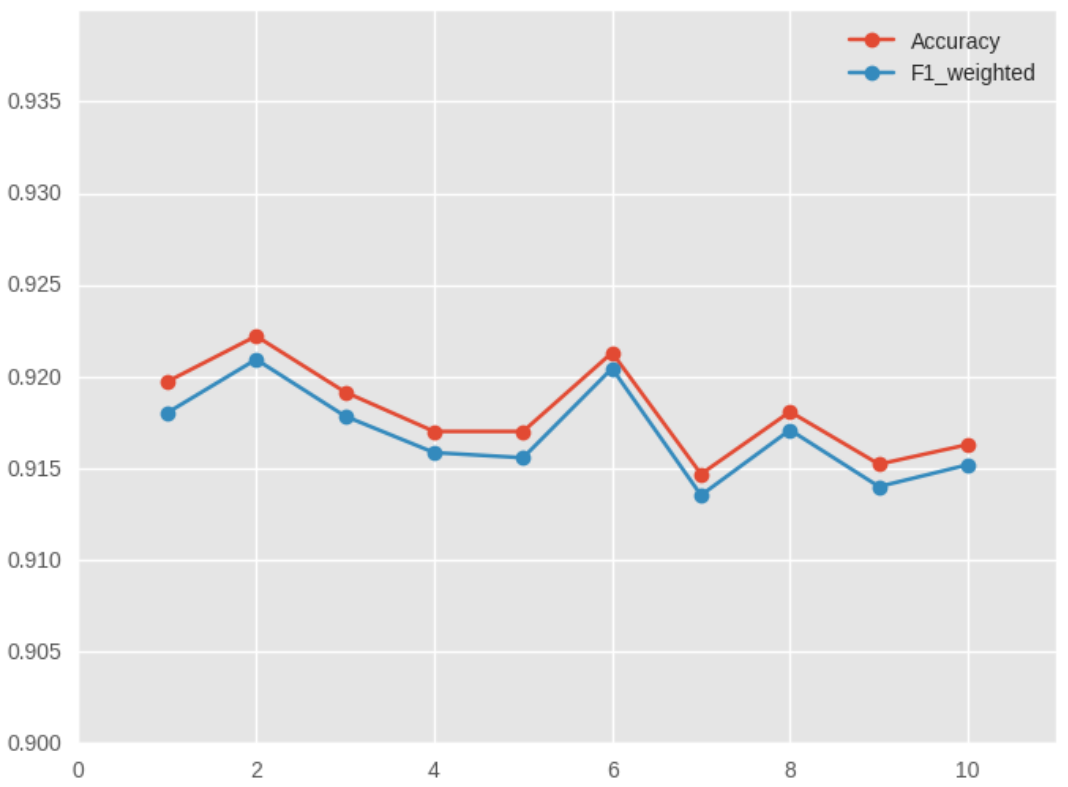
XGBoost – 10 fold CV

20 subjects



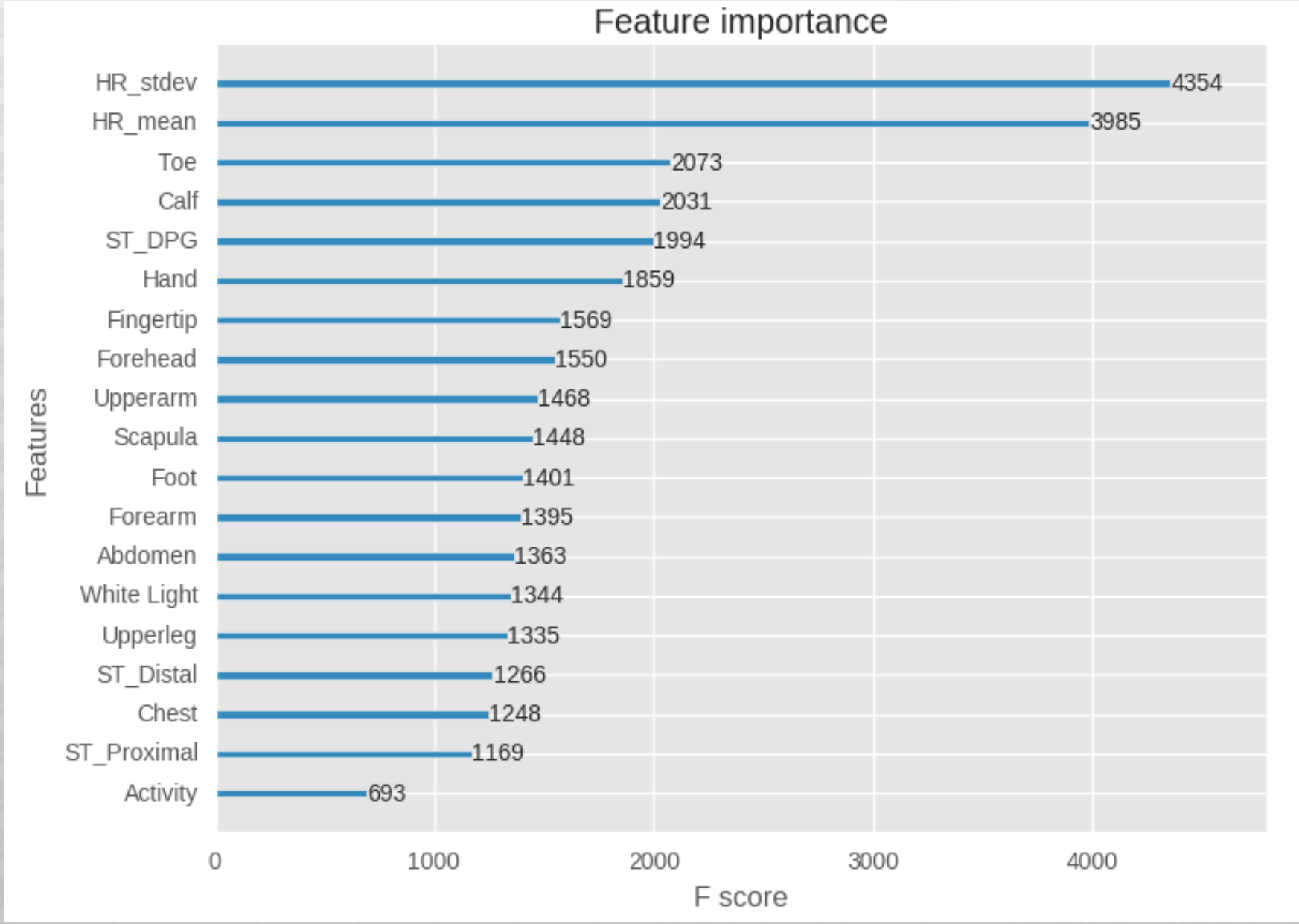
Avg Accuracy	0.9334
Avg F1 score	0.9322

59 subjects



Avg Accuracy	0.9181
Avg F1 score	0.9168

XGBoost - Feature importance



Conclusion and future work

- Decent results achieved
- More data for robust models
- Integrity of Heart rate data
- Futher study on LSTM / CNN
- Classification of 5 sleep stages

Thank you!