Automatic detection of autonomic arousals in sleep

Michael D. Kirkegaard & Nicklas A. Hansen

Project aim

- Develop understanding of relevant clinical and technical topics
- Conduct literature study on current state-of-the-art of autonomic arousals
- Design, implement and evaluate feature extraction for a chosen modality
- Design, implement a multi-modal, data-driven autonomic arousal detection algorithm
- Extensive evaluation of the autonomic arousal detection algorithm
- Develop a prototype visual tool to aid diagnosis of sleep related disorders.

Agenda

1. Choice of methods
2. Choice of data
3. Algorithm description
4. File management
5. Results
6. User interface
7. Conclusion

Choice of methods: classifier

Chosen classifier: **RNN using GRU cells**

Feed Forward Neural Network
- Performs well on EEG + ECG
- Long training duration
- Not built for time series
- Requires advanced features

Curious Extreme Learning Machine
- Performs well on EEG
- Auto-optimising parameters
- Short training duration
- Not built for time series
- Requires advance features

RNN: Long Short-Term Memory
- Performs well on time series
- Performs well in related ECG study
- Simple feature extraction
- Slow training duration
- Unproven for arousals

RNN: Gated Recurrent Unit
- Same performance as LSTM
- Simple feature extraction
- Intermediate training duration
- Unproven for arousals

Choice of methods: modality

Chosen modality: **PPG**

- **PAT device shows multi-modal success**
- **Multimodal approach**

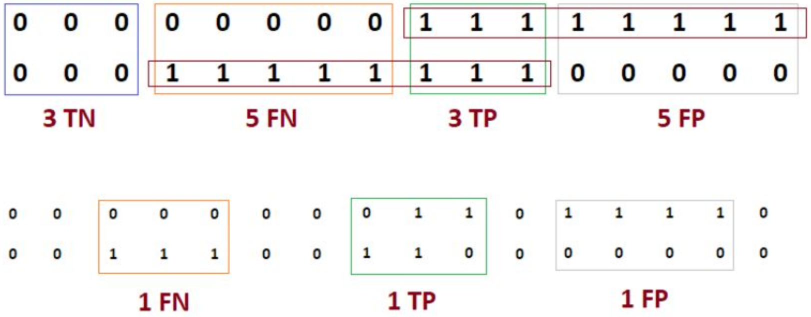
Choice of methods

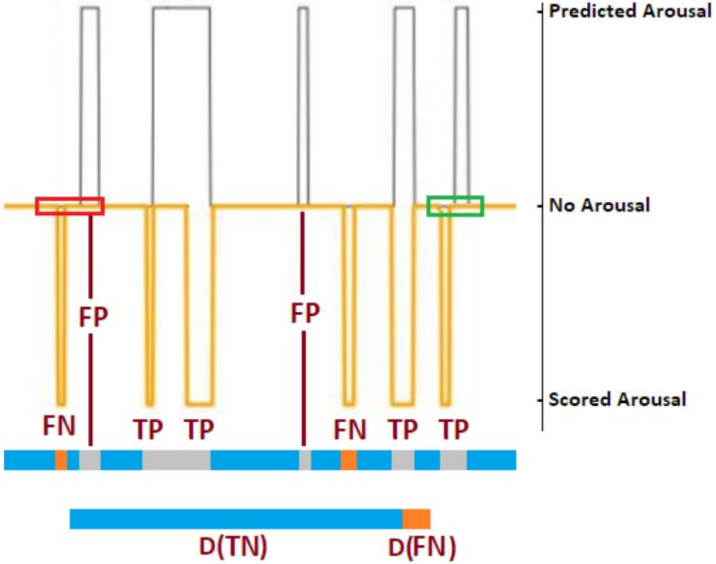
Metrics: **Accuracy, Specificity, <u>Sensitivity</u>, <u>Precision</u>, <u>F-Score</u>, MCC**

Method	**Abbreviation**	**Formula**
Accuracy	Acc	$\frac{TP+TN}{TP+FP+TN+FN}$
Specificity	Sp	$\frac{TN}{TN+FP}$
Sensitivity	Se	$\frac{TP}{TP+FN}$
Precision	P+	$\frac{TP}{TP+FP}$
F_1-Score	F1	$\frac{2 \cdot TP}{2 \cdot TP+FP+FN}$
MCC	–	$\frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$

Choice of methods

“True Positives”, “False Positive”, “True Negative” and “False Negative” in an arousal setting?





$$\mathrm{TN} = \mathrm{FN}\frac{D(\mathrm{TN})}{D(\mathrm{FN})} \quad , \quad \mathrm{TN} \in \mathbb{N} \; , \; \mathrm{FN} > 0$$

Choice of data: data-set

All files from MESA:

- 2054 PSG recordings + Sleep stage / arousal annotations (of 2237 subjects)
- Ethnicity: 830 Caucasian, 265 Asian, 616 African, 526 Hispanic
- Gender: 1039 Male, 1198 Female
- Age: 54-95, average of ~70

Choice of data: data-set

All files from MESA:

- 2054 PSG recordings + Sleep stage / arousal annotations (of 2237 subjects)
- Ethnicity: 830 Caucasian, 265 Asian, 616 African, 526 Hispanic
- Gender: 1039 Male, 1198 Female
- Age: 54-95, average of ~70

60 randomly selected files from SHHS 2:

- 2651 PSG recordings + Sleep stage / arousal annotations (of 4080 subjects)
- Ethnicity: 3587 Caucasian, 256 African, 237 undefined
- Gender: 1861 Male, 2219 female
- Age: 44-90, average ~68

Choice of data: feature selection

PPG:

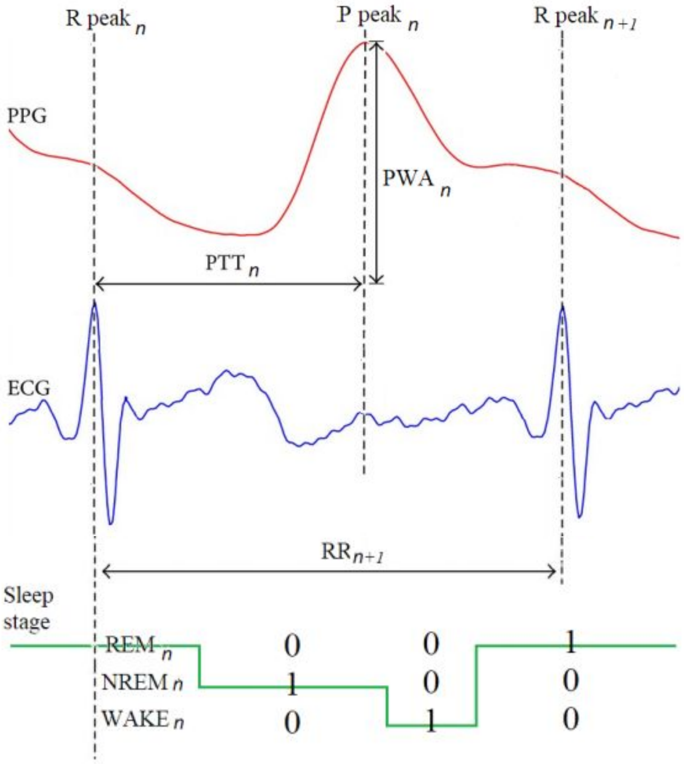
- Pulse Transit Time (PTT)
- Pulse Wave Amplitude (PWA)

ECG:

- RR-interval (RR)

Sleep state annotation:

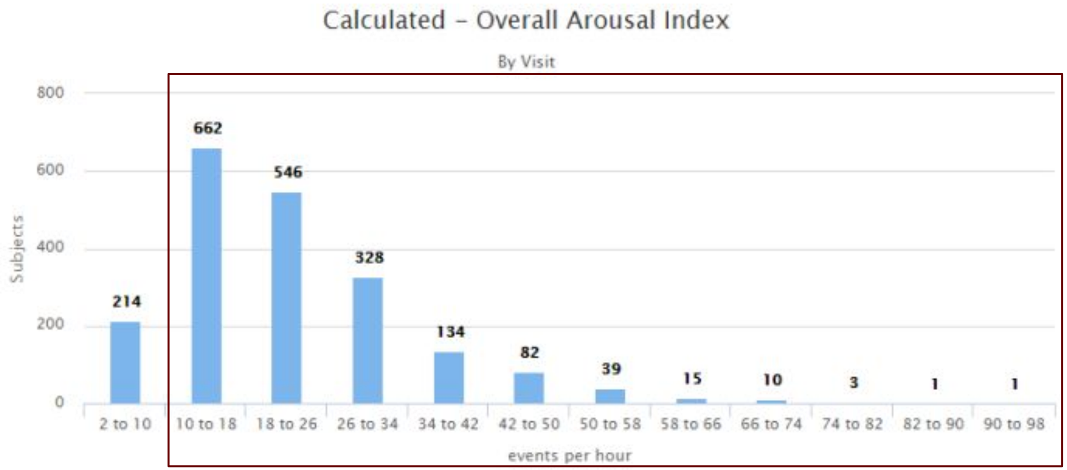
- Rem
- Non-Rem
- Wake



Choice of data: training data

Data trimming

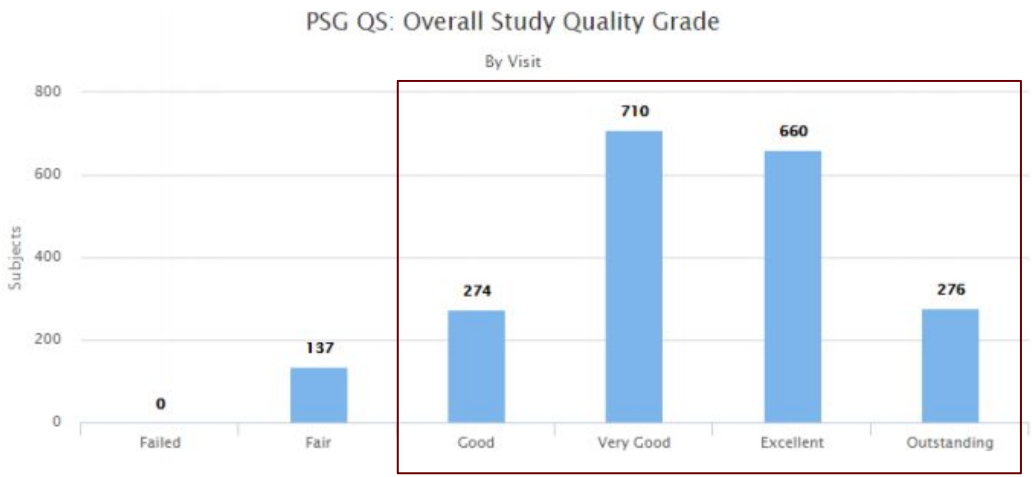
- MESA Variable Criteria
  - Arousal index



Choice of data: training data

Data trimming

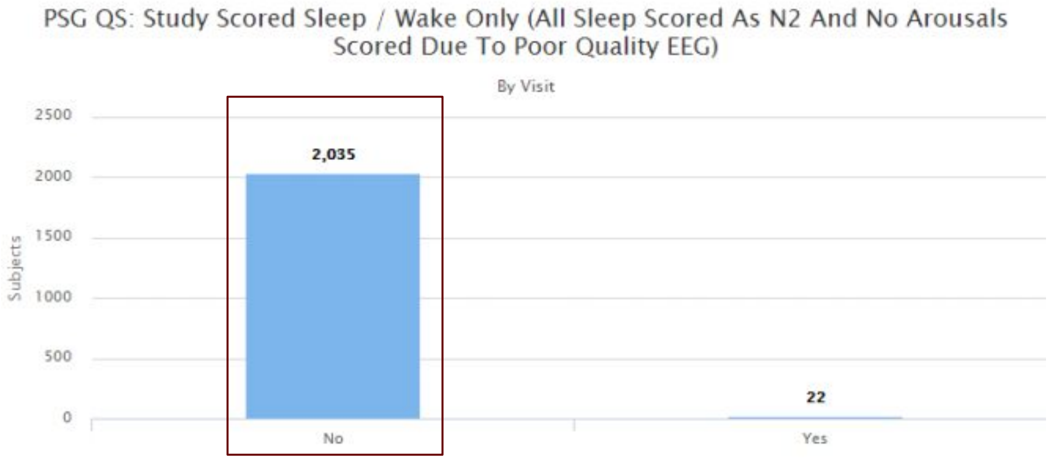
- MESA Variable Criteria
  - Arousal index
  - Overall study quality



Choice of data: training data

Data trimming

- MESA Variable Criteria
  - Arousal index
  - Overall study quality
  - Lack of scored arousals and sleep stages



Choice of data: training data

Data trimming

- MESA Variable Criteria
  - Arousal index
  - Overall study quality
  - Lack of scored arousals and sleep stages

- Error Criteria
  - 5% error in feature series
  - 10% error in total file

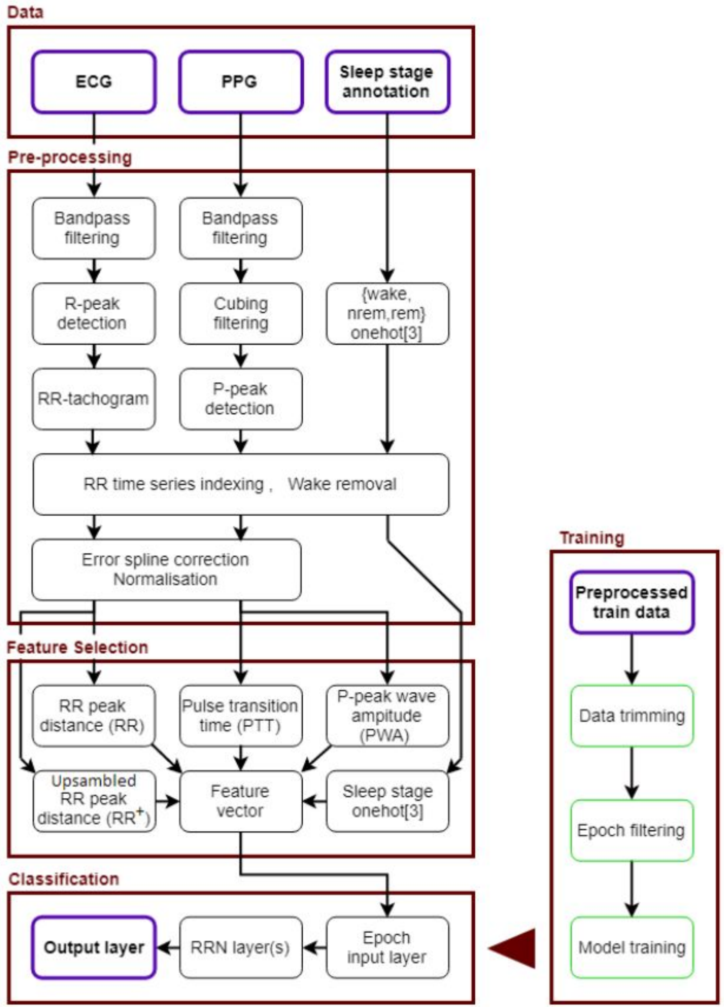
Criteria	**Amount**
MESA: ai_all5	232
MESA: overall5	137
MESA: slewake5	22
no arousals	24
E_{RR}	6
E_{RR+}	6
E_{PTT}	595
E_{PWA}	510
E^{+}	628

Group	**Amount**
All files	2054
Discarded files	871
Remaining files	1183

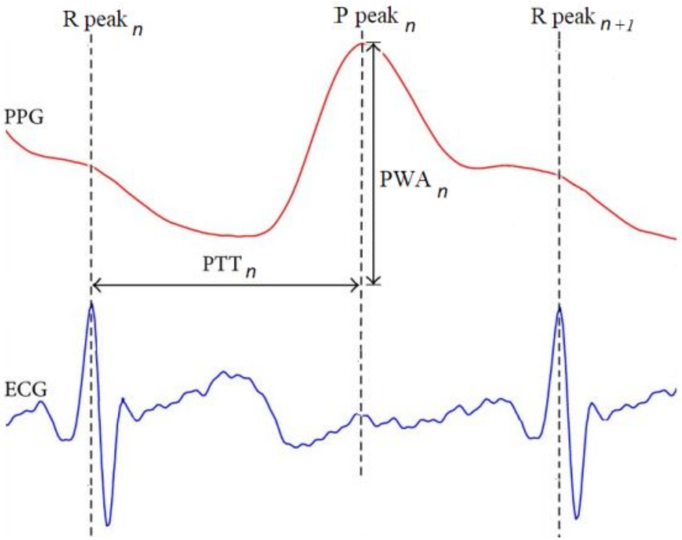
Group	**Amount**
All hours	13334.174
Discarded hours	5539.010
Remaining hours	7795.164

Algorithm description

- Preprocessing
  - PPG
  - ECG
- Cleaning
  - Error correction
  - Wake removal
  - Normalisation
- Training
  - Epochs
  - Validation

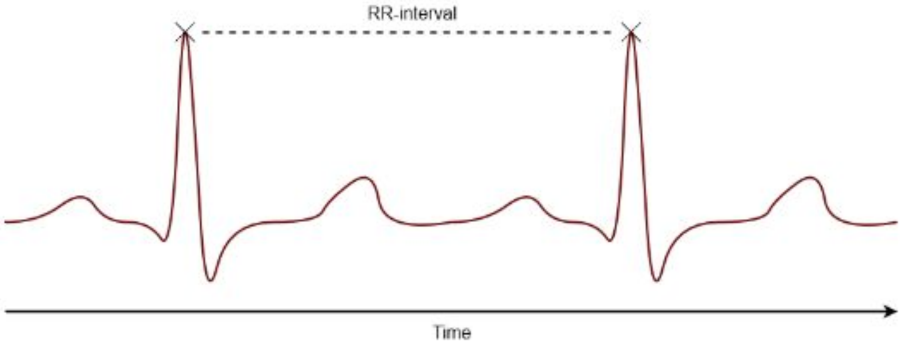


Algorithm description: PPG



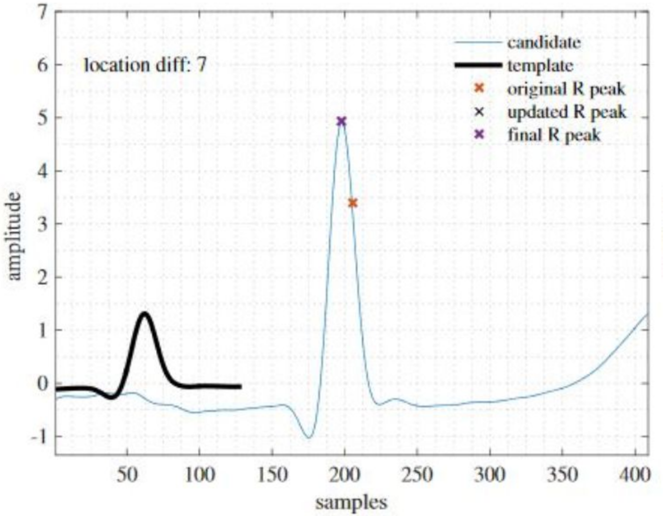
PTT and PWA are extracted from the signal using developed P-peak detection algorithm

Algorithm description: ECG



RR-interval is extracted using provided state-of-the-art R-peak detection algorithm

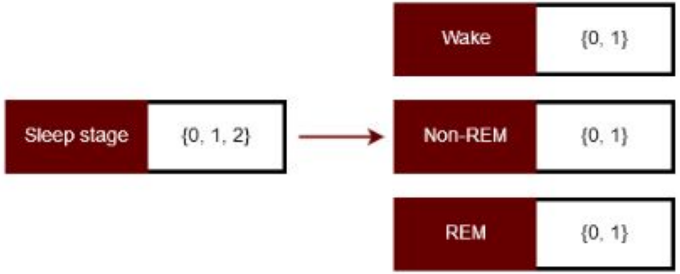
Algorithm description: ECG



RR-interval is adjusted by upsampling and template matching

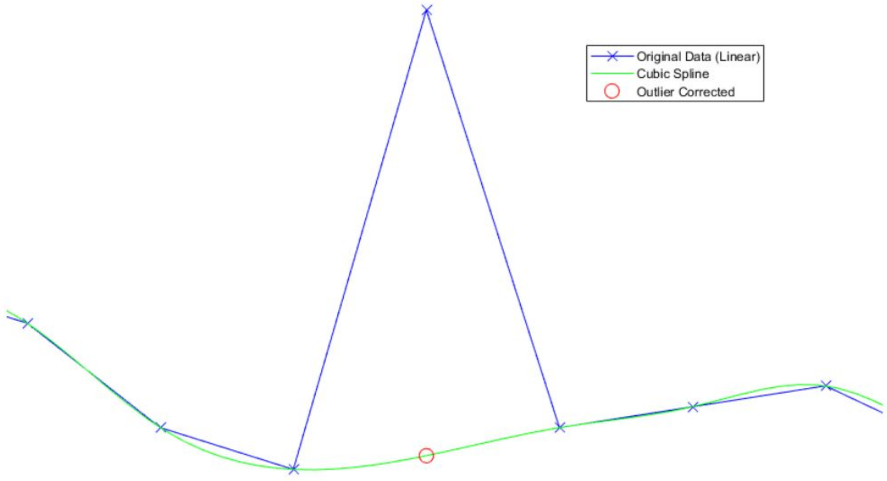
Source: M. Olsen et al. (2018), Sleep, zsy006

Algorithm description: sleep annotations



Sleep annotations are converted to three distinct features by one-hot transform

Algorithm description: error correction

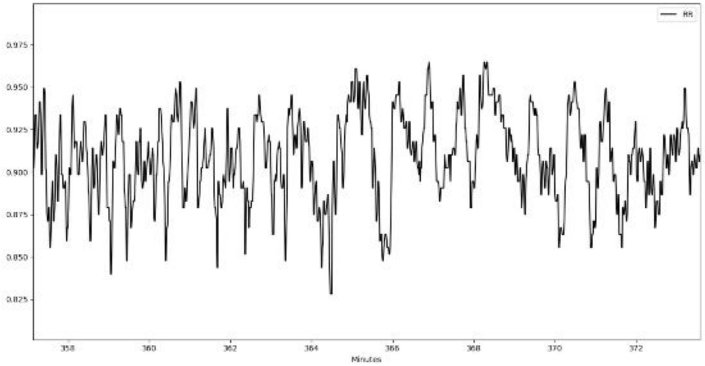


Outliers are found and corrected using a binary signal mask and cubic spline interpolation

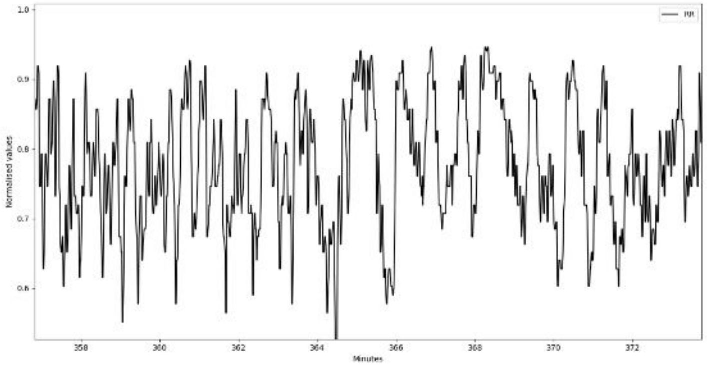
Algorithm description: wake state removal

Short durations of wakefulness are conserved, while segments of >60 s are removed except for the first and last 30 s

Algorithm description: quantile normalisation



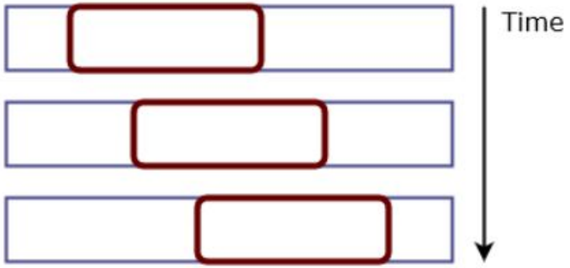
Before: value range dependent on patient

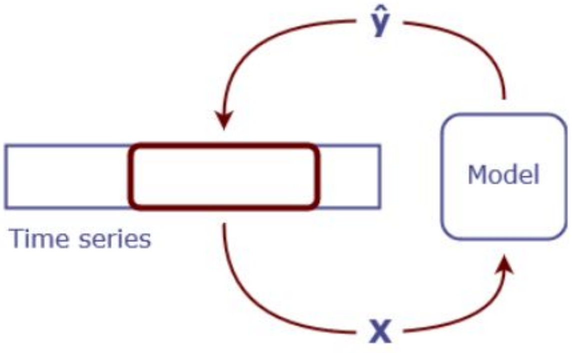


After: properties are kept, but value range is [0,1]

Algorithm description: epochs

Epoch: small time window created with ⅔ time overlap



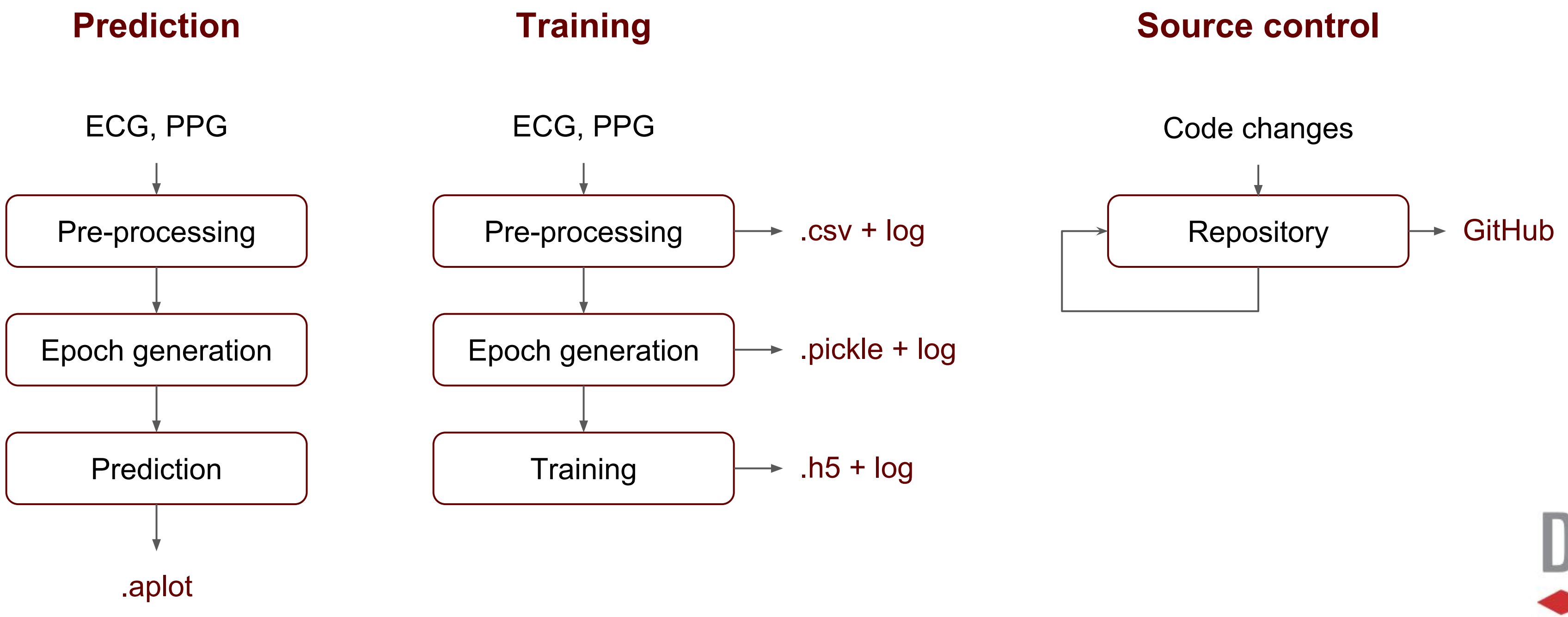


Algorithm description: training

Epoch filtering criteria

Continuous	Acceptable	Uncut	Arousal	Total	Remains
601 ± 289	620 ± 285	540 ± 268	474 ± 95	634 ± 290	349 ± 181

File management



File management

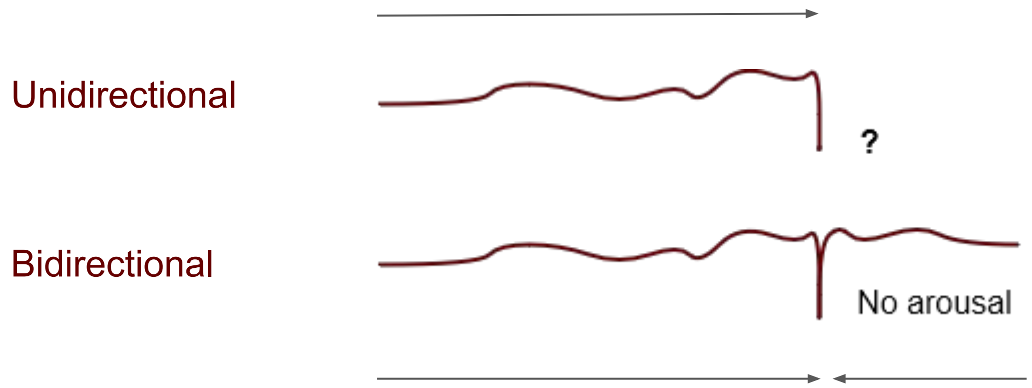
```
Preprocessed files:  2054
Reliable files:      1183
Removed by ai\_all5:  232
Removed by overall5: 137
Removed by slewake5: 22
Removed by sum(y)=0: 24
Removed by mask RR:  6
Removed by mask RR+: 6
Removed by mask PTT: 595
Removed by mask PWA: 510
Removed by mask all: 628
-----------------------------------
mesa-sleep-0010 -- T,F,T,T,T,T,F,F,F
mesa-sleep-0016 -- T,F,T,T,T,T,T,T,T
mesa-sleep-0021 -- T,T,T,T,T,T,T,T,F
mesa-sleep-0027 -- T,T,T,T,T,T,F,F,F
mesa-sleep-0028 -- T,T,T,T,T,T,F,F,F
mesa-sleep-0033 -- T,T,T,T,T,T,F,F,F
mesa-sleep-0038 -- F,T,T,T,T,T,T,T,T
mesa-sleep-0050 -- T,T,T,T,T,T,F,F,F
mesa-sleep-0074 -- T,T,T,T,T,T,T,T,F
mesa-sleep-0081 -- T,T,T,T,T,T,T,T,F
mesa-sleep-0085 -- T,T,T,T,T,T,F,T,F
mesa-sleep-0087 -- F,T,T,T,T,T,T,T,T
mesa-sleep-0107 -- T,T,T,T,T,T,F,T,F

[...]
```

```
Total files:                1009
Reliable files:             845
-----------------------------------
Files already completed:    907
Files remaining:            36
-----------------------------------
mesa-sleep-0001 already completed
mesa-sleep-0002 already completed
mesa-sleep-0006 already completed

[...]

mesa-sleep-4099 already completed
-----------------------------------
mesa-sleep-2987 pre-processed in 154s
mesa-sleep-3264 pre-processed in 131s
mesa-sleep-3267 pre-processed in 143s
mesa-sleep-3275 pre-processed in 124s
mesa-sleep-3280 pre-processed in 123s
mesa-sleep-3290 pre-processed in 123s
mesa-sleep-3293 pre-processed in 123s

[...]
```

```
mesa-sleep-0299 -- Se: 0.56, P+: 0.83
mesa-sleep-5563 -- Se: 0.68, P+: 0.82
mesa-sleep-1152 -- Se: 0.57, P+: 0.68
mesa-sleep-5810 -- Se: 0.22, P+: 0.84
mesa-sleep-6680 -- Se: 0.68, P+: 0.55
mesa-sleep-2172 -- Se: 0.46, P+: 0.73
mesa-sleep-6789 -- Se: 0.69, P+: 0.73
mesa-sleep-3850 -- Se: 0.68, P+: 0.60
mesa-sleep-2677 -- Se: 0.33, P+: 0.75
mesa-sleep-1541 -- Se: 0.73, P+: 0.84
[...]
```

Hypotheses

A **bidirectional** RNN has higher precision than a unidirectional RNN

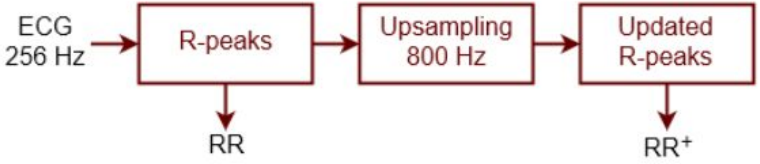


Hypotheses

Inclusion of **PPG** features increases performance

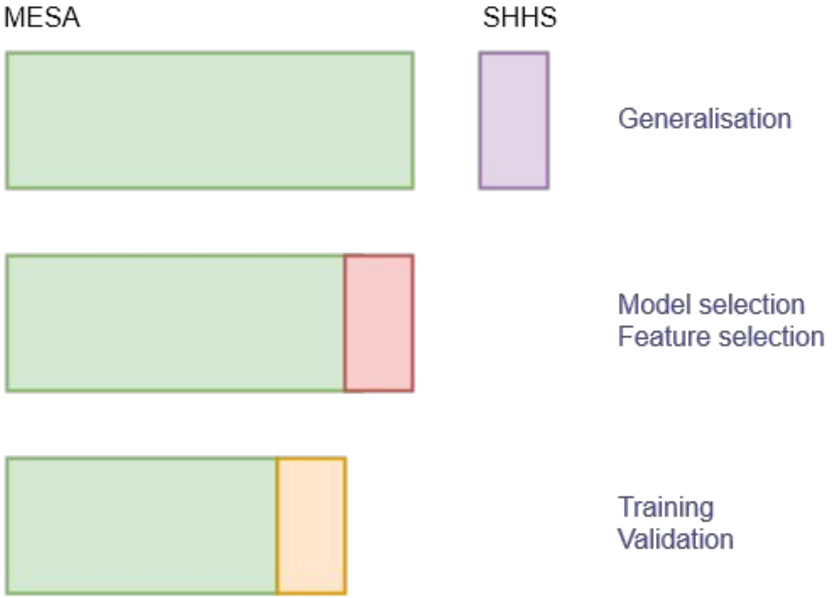
RR | Sleep annotations + PTT | PWA = Performance ⬆

Hypotheses



Upsampling and **template matching** increases performance

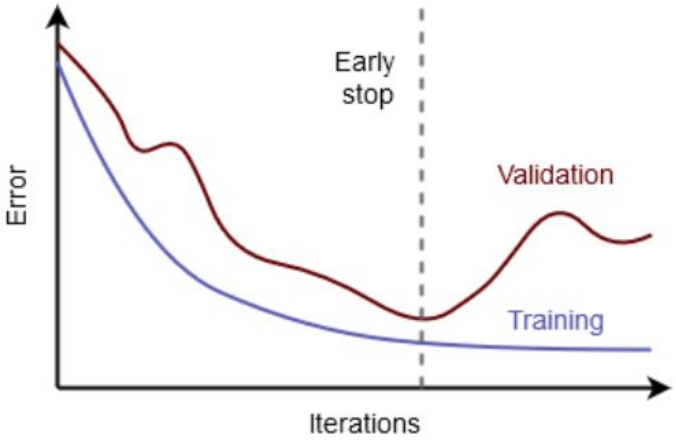
Results

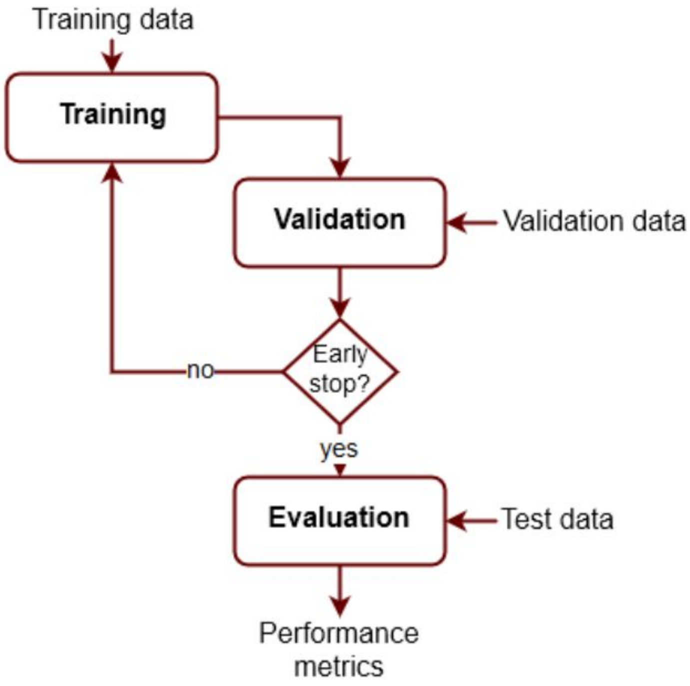


3-layer cross-validation is used to produce an accurate estimate of performance

Results: first two layers

Illustration of early stopping

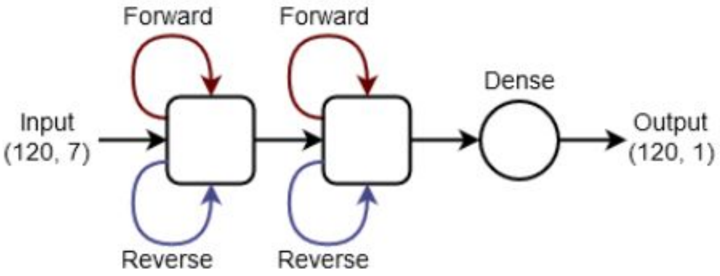




Results: model selection

Recurrent layers	Bi-directional	Sensitivity	Precision
1	No	56.02%	68.24%
2	No	59.30%	61.87%
1	Yes	58.63%	65.95%
2	**Yes**	**57.90%**	**68.08%**
3	Yes	53.74%	71.93%
4	Yes	54.77%	70.60%

Best model:
2-layered bidirectional RNN with GRU cells



Results: feature selection

Features	Sensitivity	Precision	F-score
RR	56.96%	66.59%	0.614
RR^+	**59.06%**	**67.92%**	**0.632**
PTT, PWA	29.93%	80.05%	0.436
RR, PTT, PWA	54.70%	71.21%	0.619

Inclusion of PTT and PWA increases precision, but decreases sensitivity slightly

Results: feature selection

Features	Sensitivity	Precision	F-score
RR	56.96%	66.59%	0.614
RR$^+$	**59.06%**	**67.92%**	**0.632**
PTT, PWA	29.93%	80.05%	0.436
RR, PTT, PWA	54.70%	71.21%	0.619

PTT and PWA alone achieves high precision, but only 30% sensitivity

Results: feature selection

Features	Sensitivity	Precision	F-score
RR	56.96%	66.59%	0.614
RR$^+$	**59.06%**	**67.92%**	**0.632**
PTT, PWA	29.93%	80.05%	0.436
RR, PTT, PWA	54.70%	71.21%	0.619

Upsampling & template matching increases overall performance

Results: generalisation

SHHS

Accuracy	Specificity	Sensitivity	Precision	F1-score	MCC
98.32 %	99.38 %	58.53 %	71.68 %	0.6444	0.6395

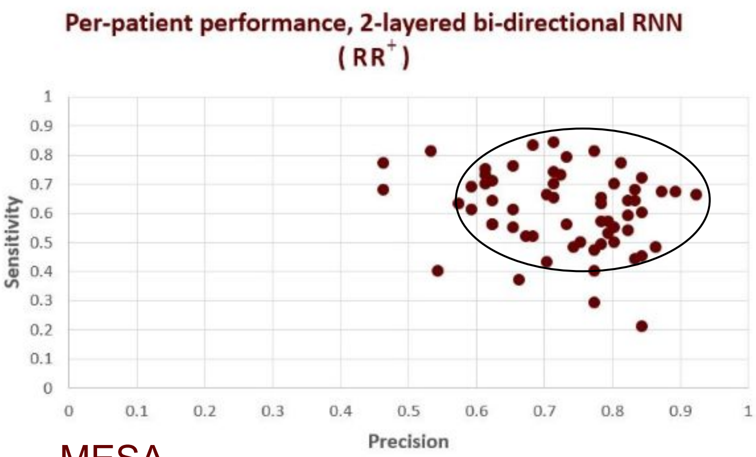
Model generalises well to other data-sets

Results: generalisation

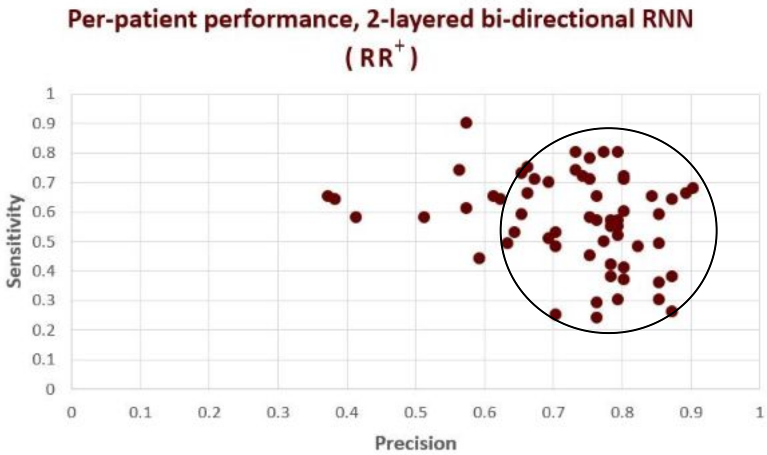
SHHS

Accuracy	Specificity	Sensitivity	Precision	F1-score	MCC
98.32 %	99.38 %	58.53 %	71.68 %	0.6444	0.6395

Model generalises well to other data-sets

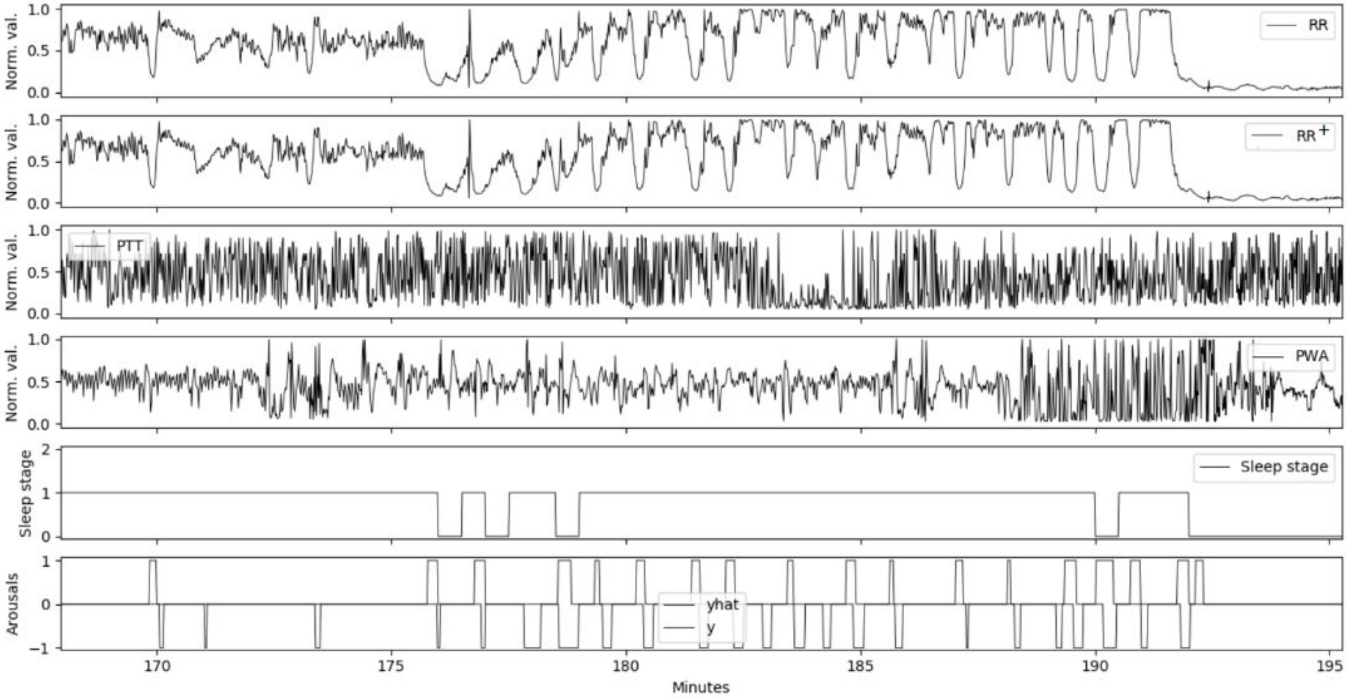


MESA

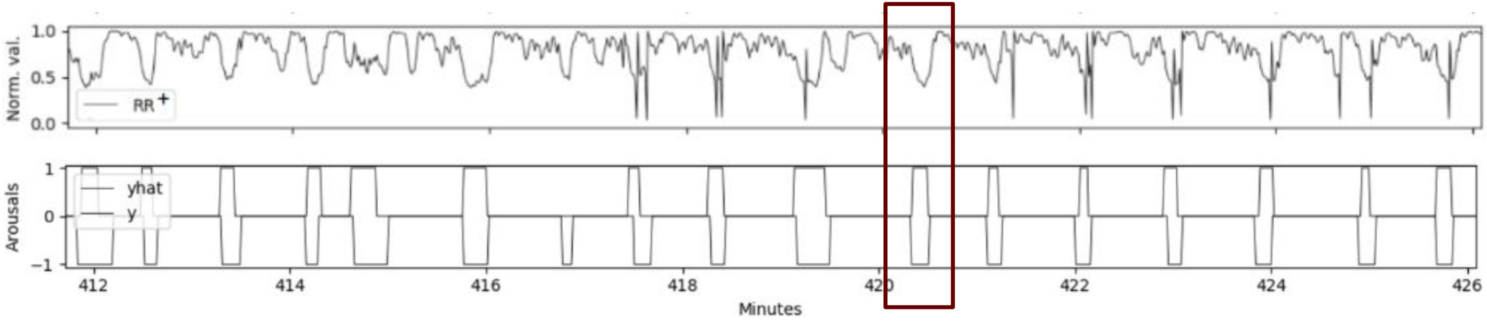


SHHS

Results: performance

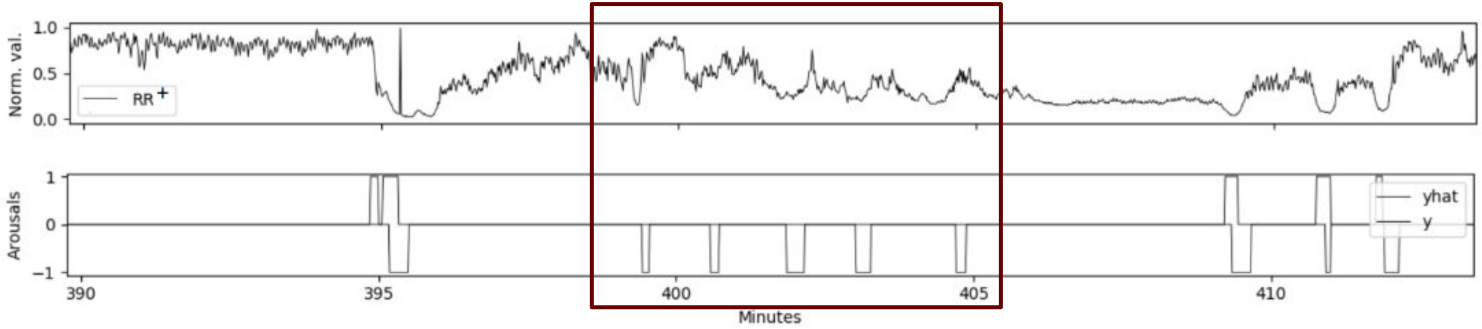


Results: performance



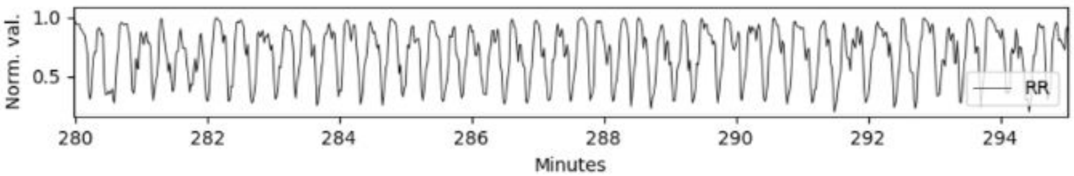
Model learns that **significant decrease in RR = arousal likely**

Results: performance

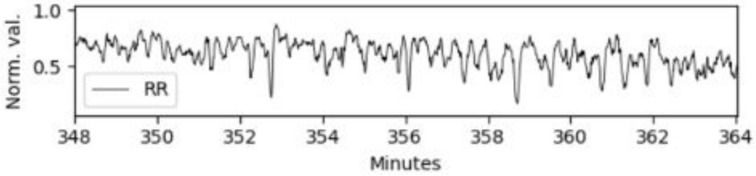


But does not always detect **irregular drops ➜ lower sensitivity**

Results: abnormal heart rate



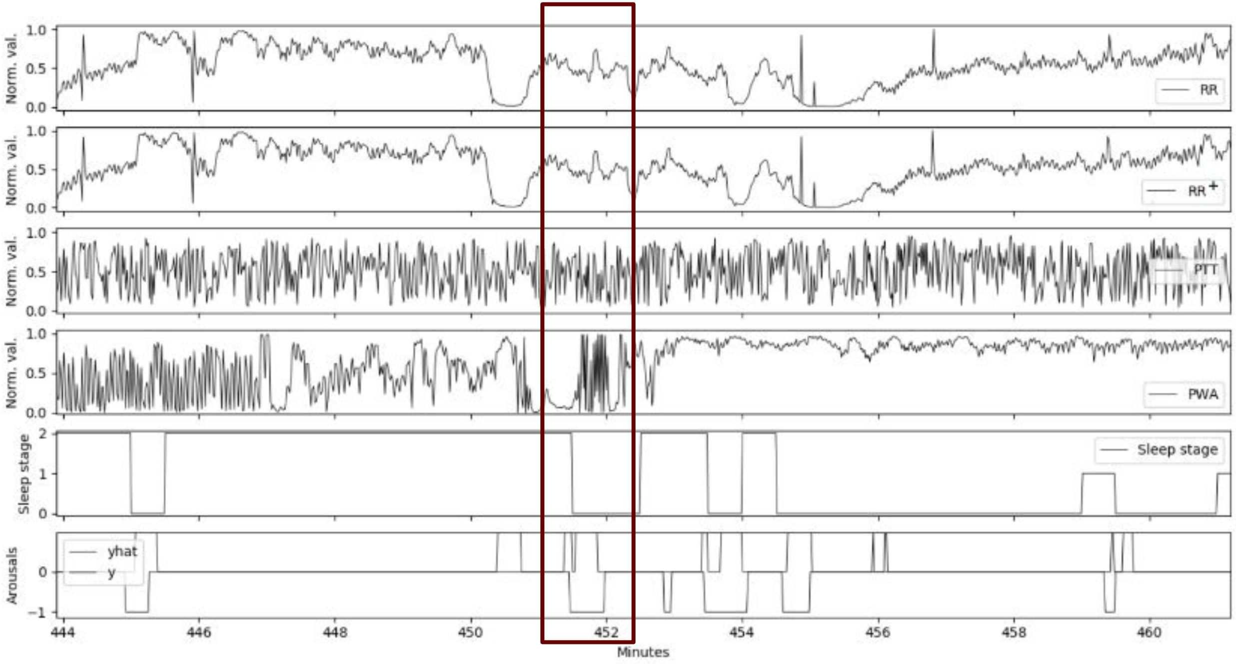
Abnormal heart rate (**Se 22%, P+ 84%**)



Regular heart rate (**Se 73%, P+ 84%**)

Model learns that **abnormal heart rate ≠ arousal**, but has low sensitivity

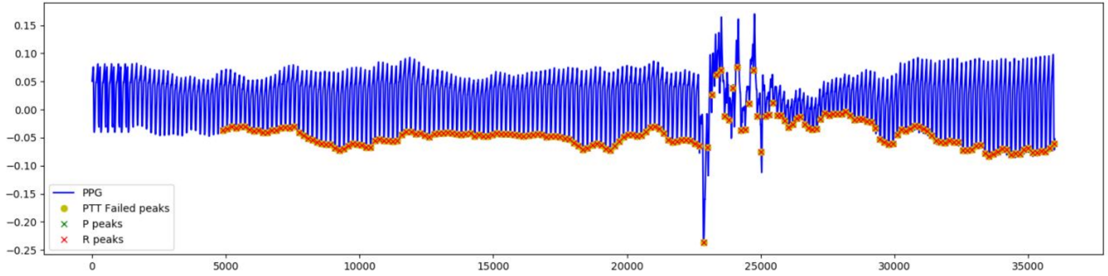
Results: sleep stage impact



Model learns that **short wake annotation = arousal likely**

Results: PPG - high error trimming

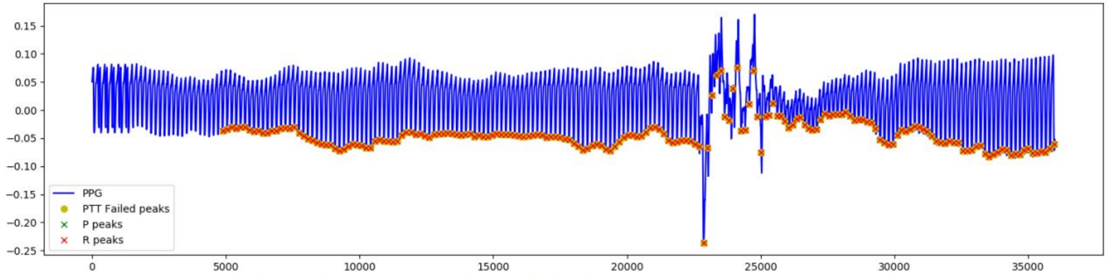
Most files in trimming are missed entirely



(a) No peaks detected, and PTT errors are at every peak.

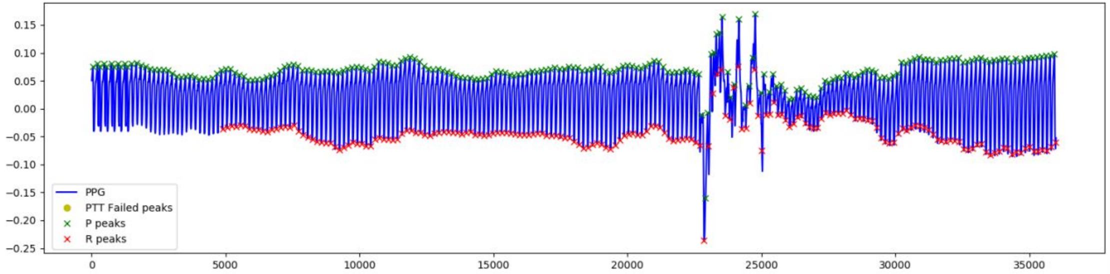
Results: PPG - high error trimming

Most files in trimming are missed entirely



(a) No peaks detected, and PTT errors are at every peak.

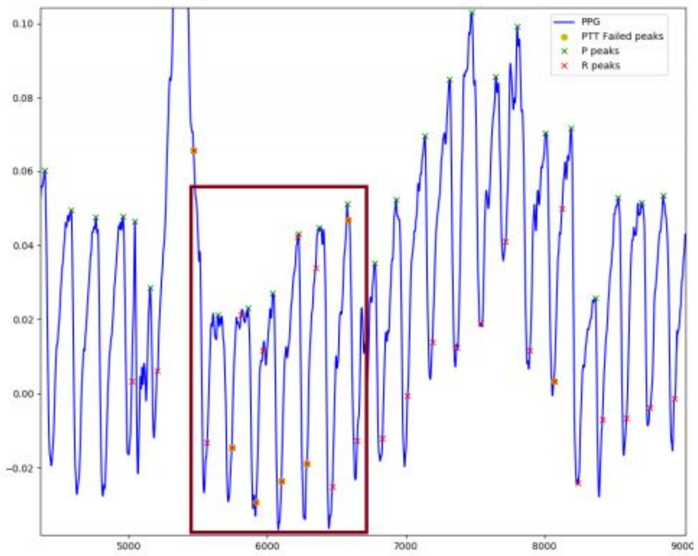
But when capping outlier peaks, these files are saved



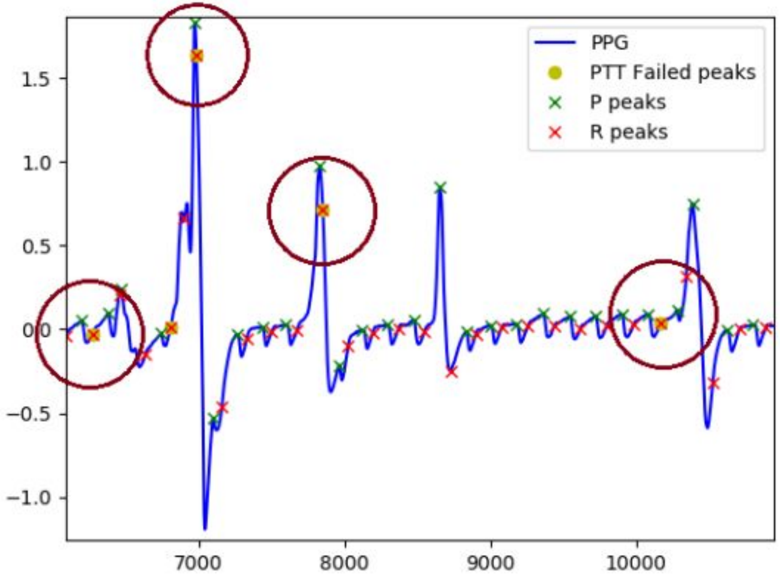
(b) All peaks detected and no errors are found.

Results: PPG - relation to ECG

Two R-peaks within natural P-rhythm

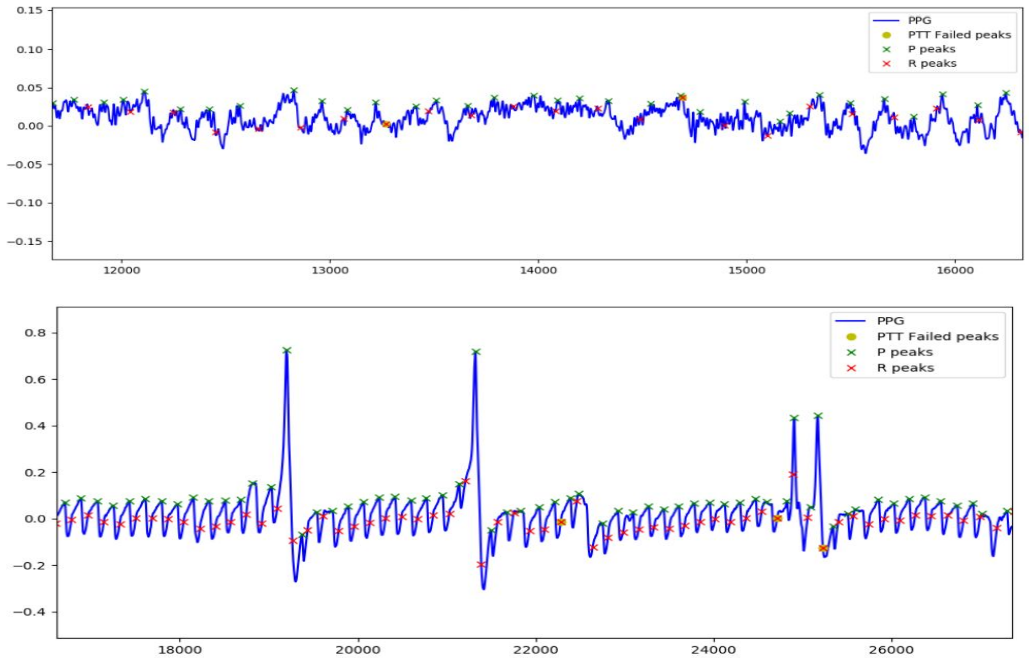


R-peaks stacked with same indices



Results: PPG - peak detection performance

- Performs well on visual inspection
- No reasonable way of quantifying this claim



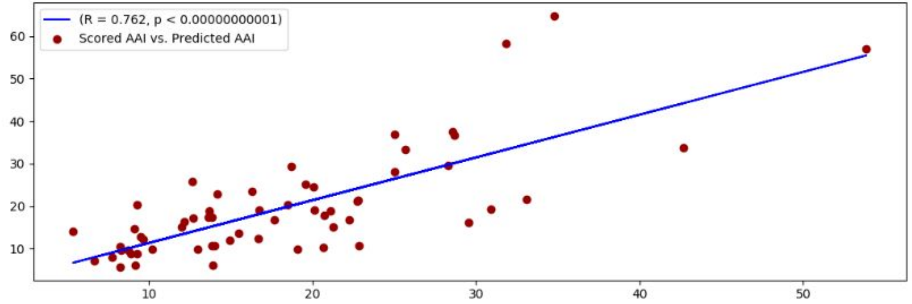
Results: comparison to other studies

G. Pillar et al. (2002)

“Autonomic Arousal Index: an Automated Detection Based on Peripheral Arterial Tonometry”

Their correlation: $(R = 0.87, p < 0.0001)$

Our correlation: $(R = 0.76, p < 0.00000000001)$



- Different main goal: Arousal index correlation
- Precision and sensitivity for arousal detection omitted

Results: comparison to other studies

M. Basner et al. (2007)

"An ECG-based Algorithm for the Automatic Identification of Autonomic Activations Associated with Cortical Arousal"

Their Metrics:	Se: 68.2%	Sp: 95.2%
Ours compared:	Se: 58.53%	Sp: 99.38%.

- Higher sensitivity, but much lower specificity in respect to imbalance in data
- Precision for arousal detection omitted (even though it is used in the study)
- Due to outliers in the paper, comparison of arousal index correlation is omitted

Results: comparison to other studies

M. Olsen et al. (2018)

"Automatic Detection of Autonomic Activations Associated with Cortical Arousals in Sleep"

Their Metrics:	Se: 63%	P+: 72%	F1: 0.6720	# ANN features: 25
Our Metrics:	Se: 58.53%	P+: 71.68%	F1: 0.6444	# ANN features: 4

- Higher sensitivity
- Similar precision
- Score achieved through removing predicted arousals of short durations
- Different ANN architecture does not compare fairly

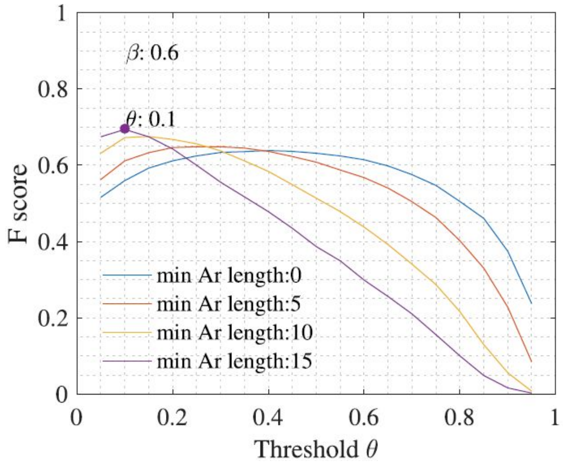


Image Source: M. Olsen et al. (2018), Sleep, zsy006

User interface

User target group:

- Researchers and otherwise affiliates of sleep studies

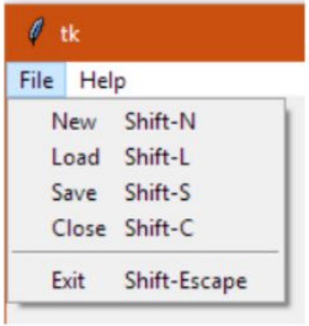
User Stories:

- User wants to perform new PSG analysis
- User wants to save analysis for later inspection
- User wants to load already performed analysis
- User wants to review analysis visually

Use case testing

User interface: functions - new analysis

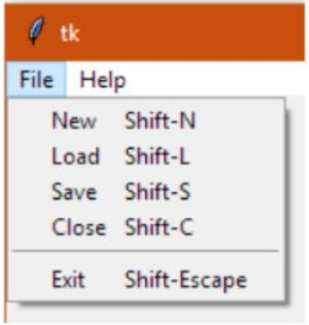
File menu:

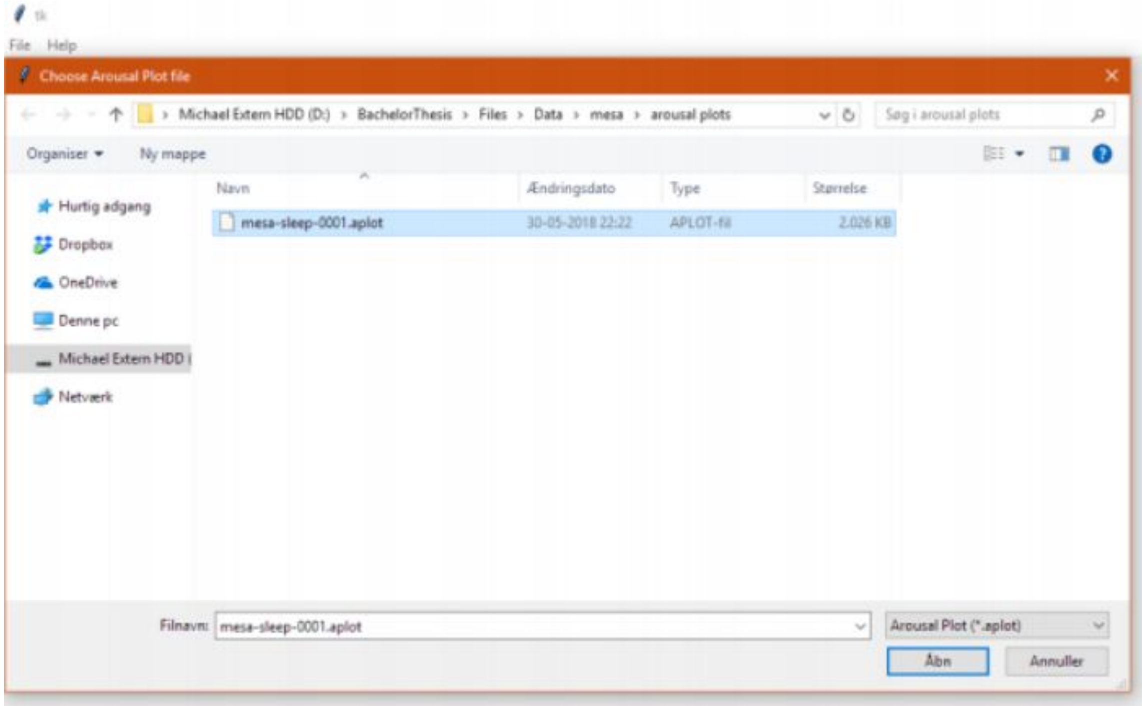


[A very short video showing steps in performing new analysis]
[Preprocessing and analysis takes a long time to display]

User interface: functions - load

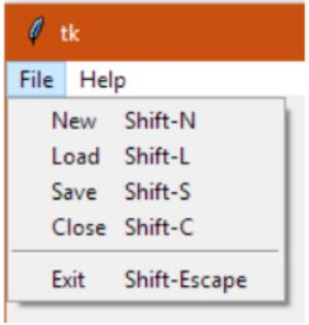
File menu:

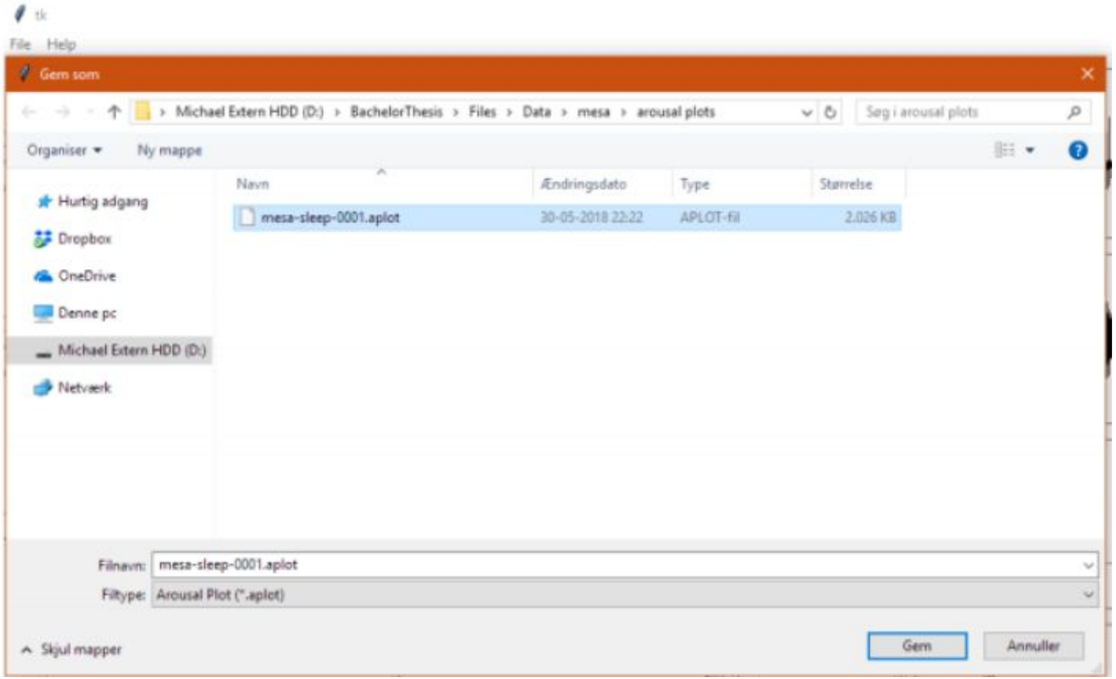




User interface: functions - save

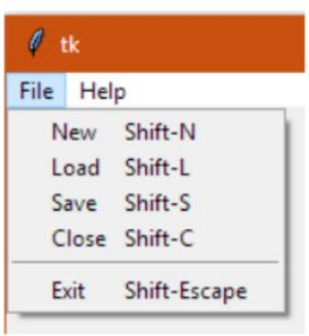
File menu:

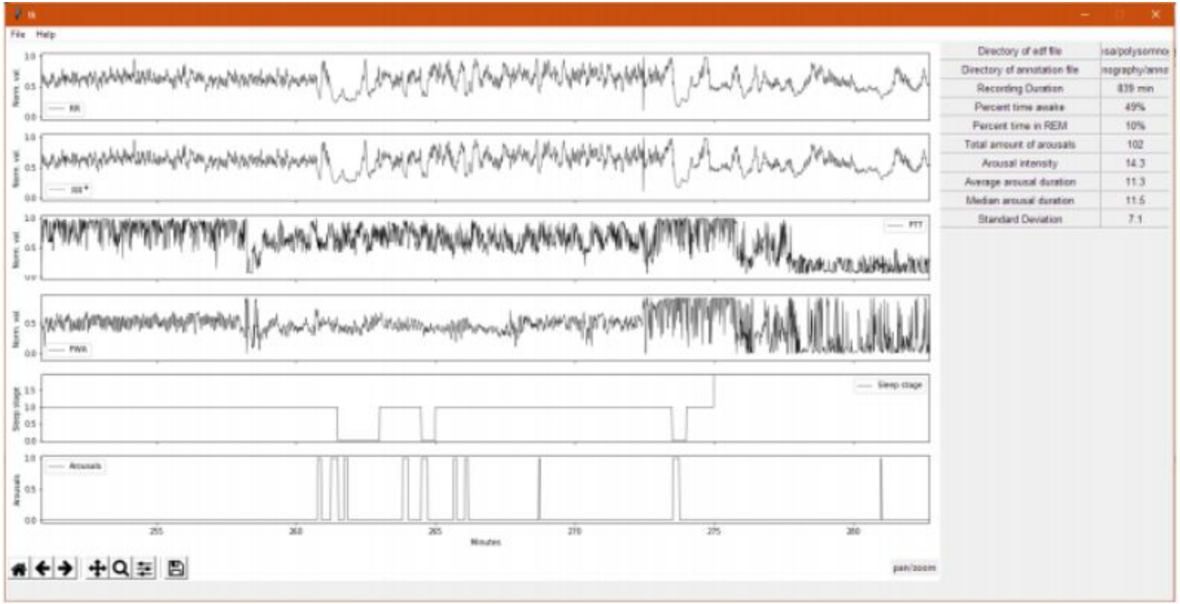




User interface: functions - review analysis visually

File menu:





Conclusion

Best model: 2-layer bidirectional RNN (GRU cells)

Best feature set: { RR+, sleep annotation }

Generalisation performance: (Se: 58.53%, P+: 71.68%, F1: 0.6444)

Findings:

- **RR-interval** is prone to abnormal heart rates
- **Multimodality** increases precision, but PPG has insignificant positive impact overall
- **RR template matching** improves performance ➜ similar tendency expected for PPG

Updated results

Changes to error correction…

{ RR+ }

Se: 62%, P+: 67%

+3 -1

{ PTT, PWA }

Se: 60%, P+: 66%

+30 -14

{ RR, RR+, PTT, PWA }

Se: 62%, P+: 71%

+4 +3

(MESA tests)

Future works

- **Additional modalities ➜** increased performance?
- **Upsampling + template matching on PPG ➜** increased PTT and PWA impact?
- **Post-processing ➜** increased performance?
- **Additional features ➜** overall more robust model?
- **Convolutional layer ➜** automatic feature extraction on input signals

Questions