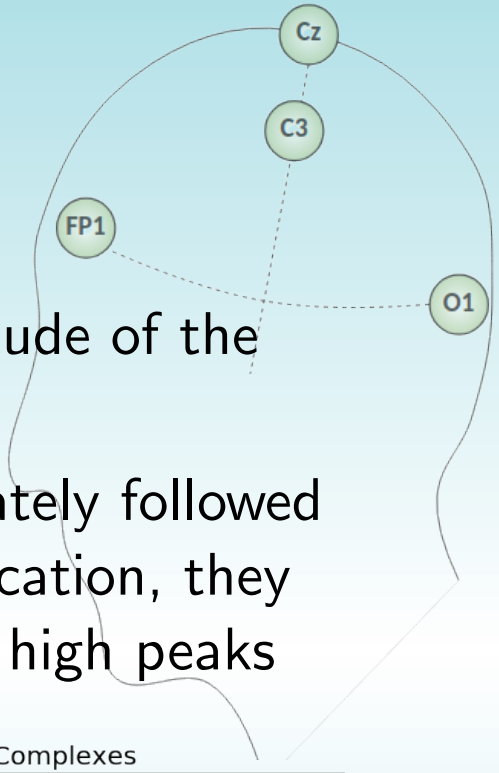
Classification of EEG Signals using Genetic Programming for Feature Construction

Humies

Icaro Miranda, Claus Aranha, Marcelo Ladeira

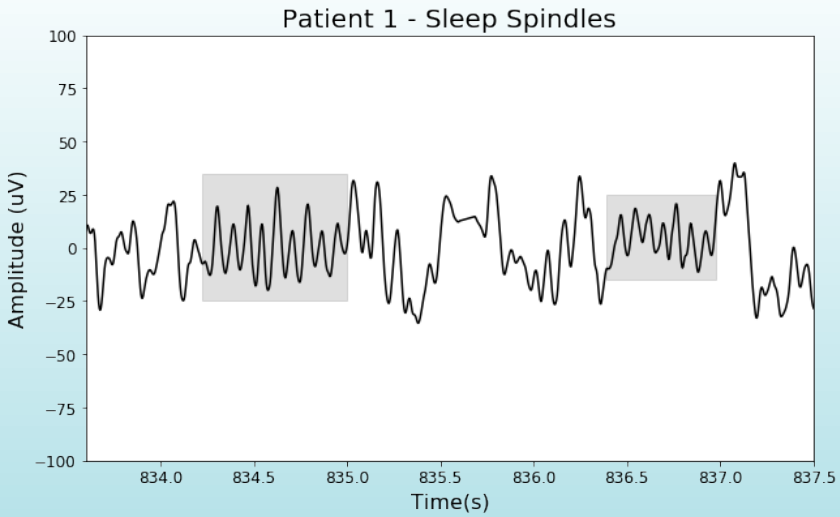
June 2019

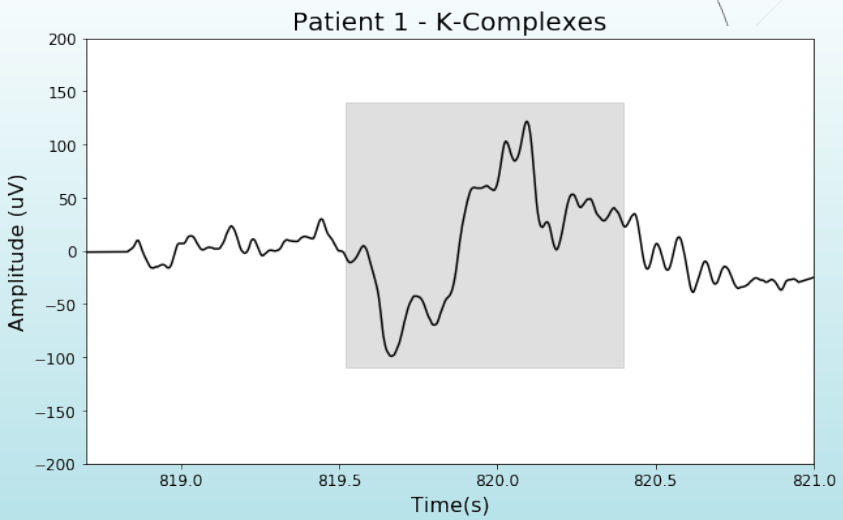
Sleep Phenomena Identification



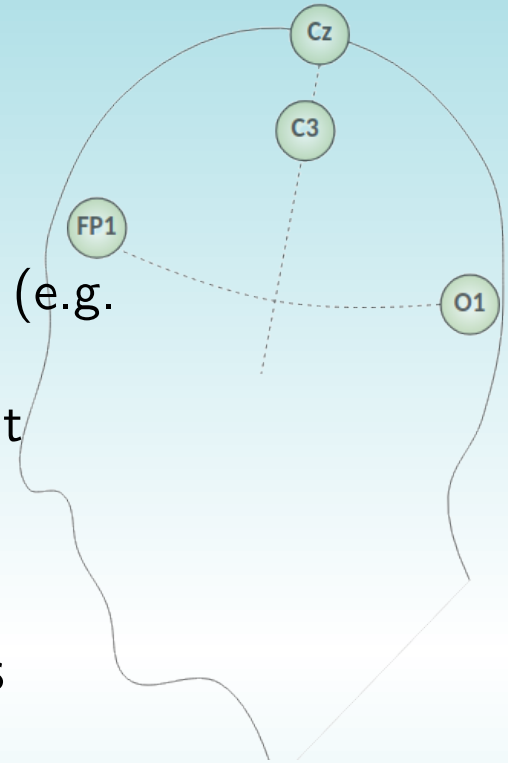
Identified **VISUALLY** by human specialists

- **Sleep Spindles:** are bursts in EEG. The amplitude of the signal increases and decreases progressively
- **K-Complexes:** A negative acute wave immediately followed by a positive component in EEG. In the identification, they can be easily confused with any waveform with high peaks



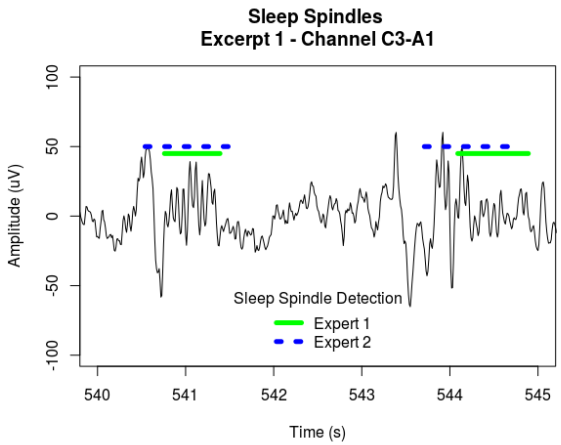


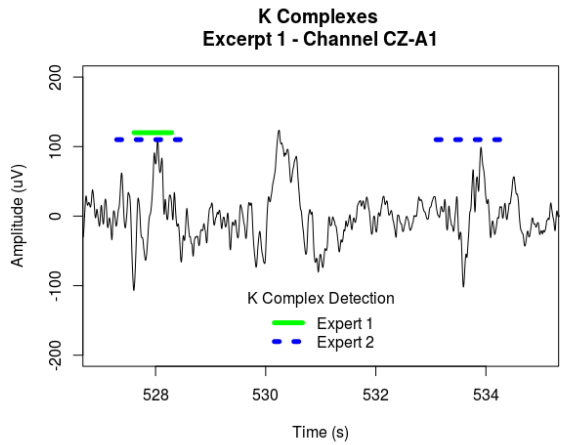
Importance



- Important biomarker related to **brain health**
  - Early detection of neurodegenerative disorders (e.g. Alzheimer's disease)
  - Assessment of children's cognitive development
  - Helps to understand schizophrenia and stress
- Sleep Staging
- Identification of multiple sleep-related disorders

Specialists can disagree!





Importance

- Undiagnosed sleep disturbances generate **economic losses**
- In the United States, in a study by Kappur et. al (Sleep, 1999), the losses with undiagnosed obstructive sleep apnea were approximately $ 3 billion per year.

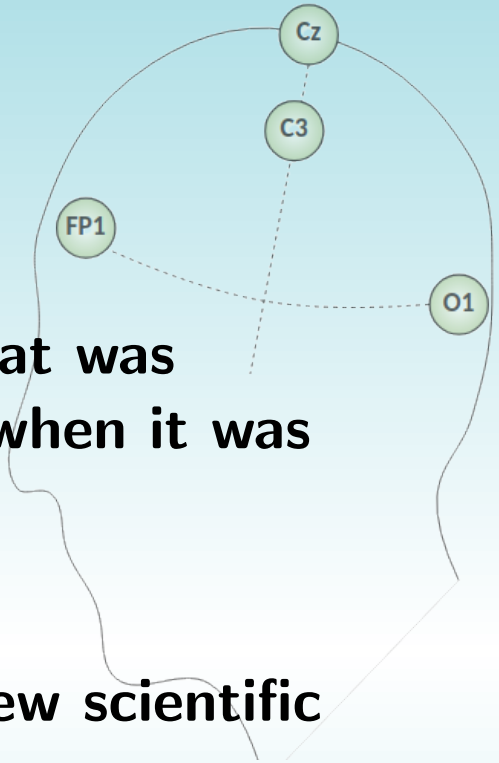
Does this examination reflect usual sleep conditions?

Many wires!

https://medsonogoiania.com.br/sem-categoria/como-e-avaliacao-medica-de-transtornos-do-sono/

"I can't sleep during PSG examination."

Why is it human competitive?



Criterion B

The result is equal to or better than a result that was accepted as a new scientific result at the time when it was published in a peer-reviewed scientific journal.

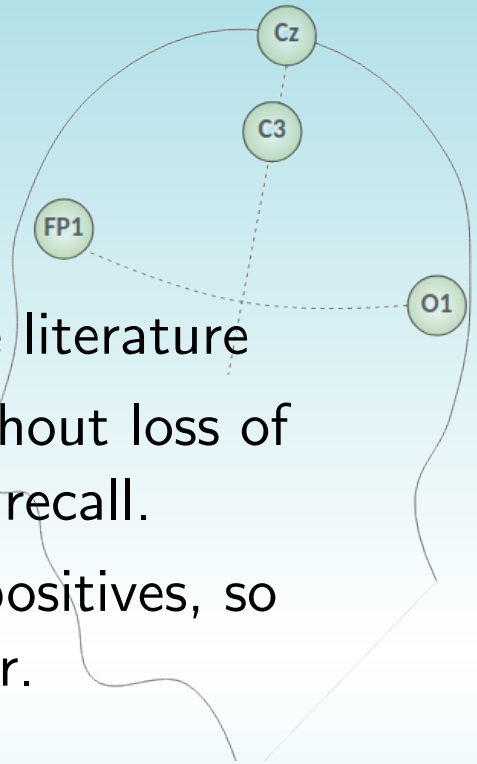
Criterion D

The result is publishable in its own right as a new scientific result independent of the fact that the result was mechanically created.

Criterion G

The result solves a problem of indisputable difficulty in its field.

Why is it human competitive?

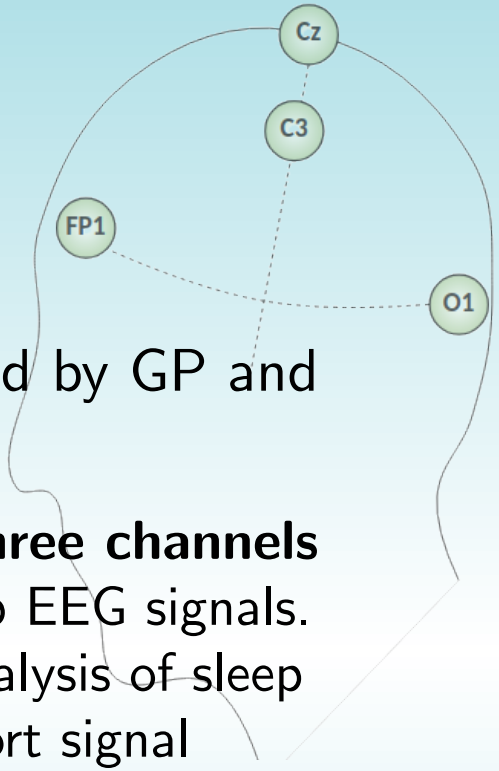


Criterion B - Performance Comparison

- Our model outperforms current models from the literature
- Better trade-off between precision and recall without loss of specificity, and generates classifiers with greater recall.
- False negatives are a bigger concern than false positives, so classifiers with great recall is a desirable behavior.

Reference	Recall	Specif.	Prec.	F_1
Lachner-Piza et al., 2018	0.65	0.98	0.38	0.48
Tsanas and Clifford, 2015	**0.76**	0.92	0.33	0.46
Zhuang et al., 2016	0.51	**0.99**	**0.70**	**0.59**
Genetic Programming	**0.75**	**0.98**	0.35	0.48

Why is it human competitive?



Criterion D - New Scientific Results

- Two properties of the EEG data that were found by GP and are of interest in their own right.
  - The GP has identified that only **one out of three channels** used is necessary for the task of model of sleep EEG signals.
  - The accuracy of model identification in the analysis of sleep EEG signals can be improved by the use of short signal samples (**2 second samples**).

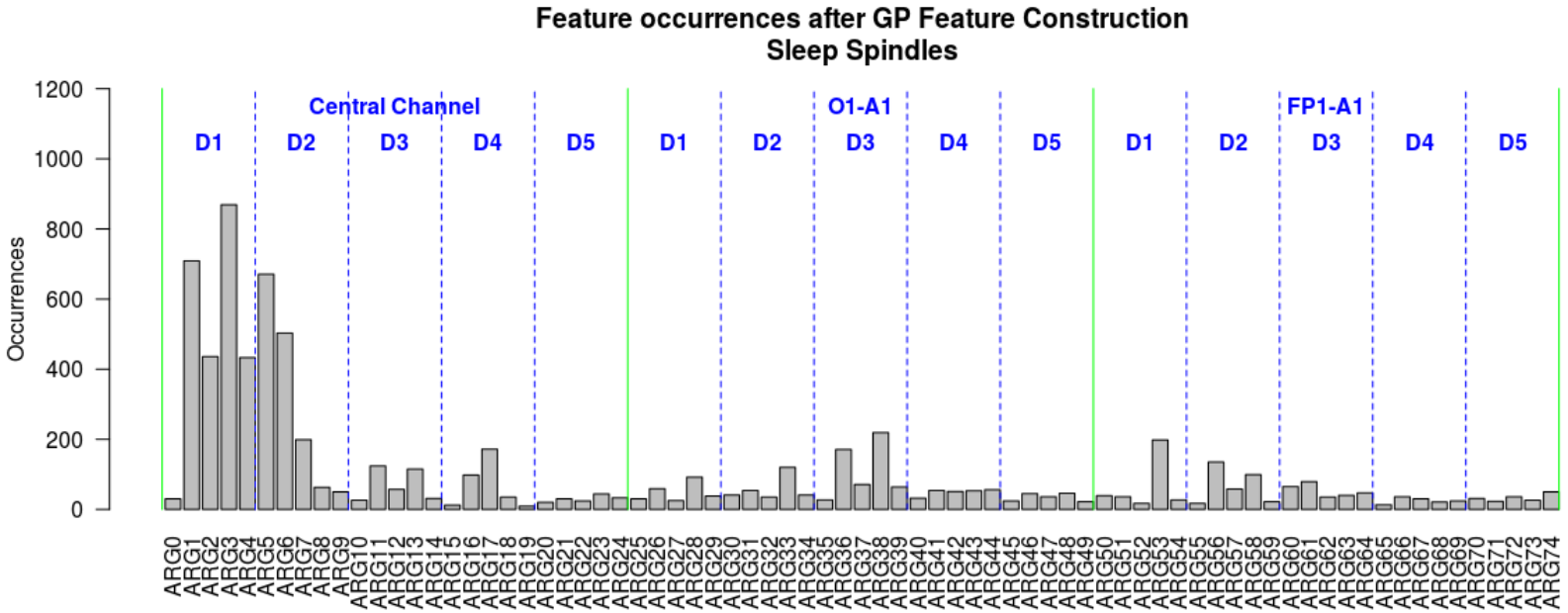
Raw input features

No normalization, No filtering, No hypothesis, No assumptions $\rightarrow$ More interpretability

Why is it human competitive?

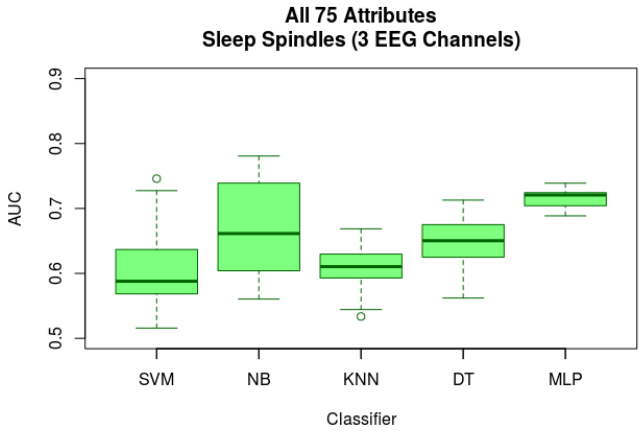
Channel Reduction

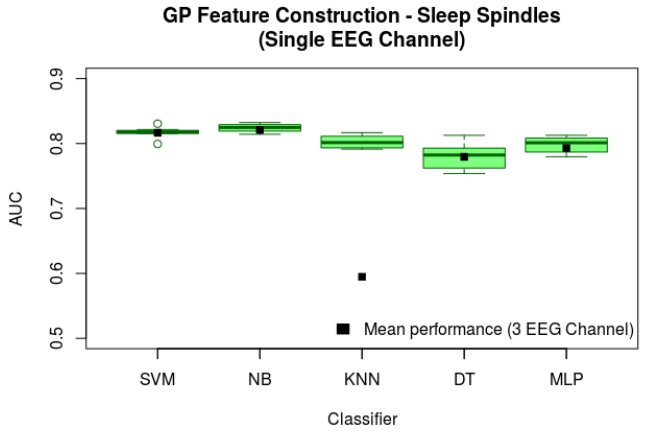
- Models' favorites features indicates that the central channel is more relevant.



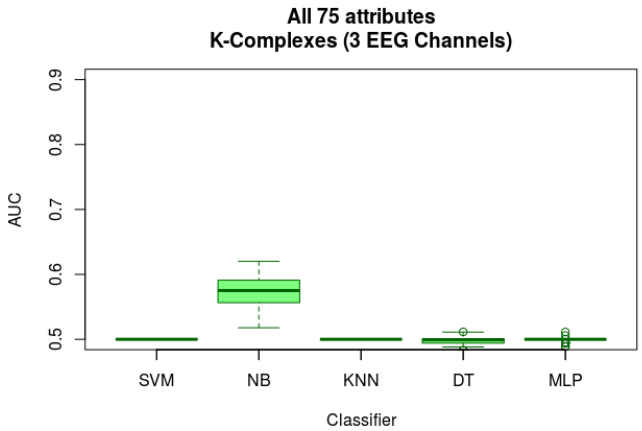
Significance of the Results

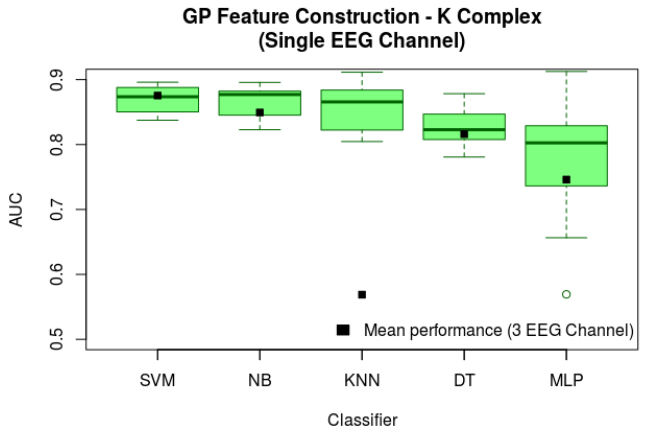
Sleep Spindles





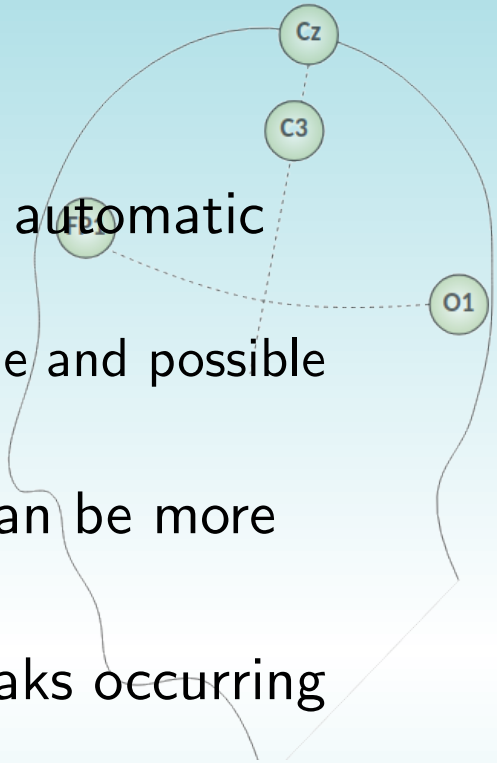
K Complexes



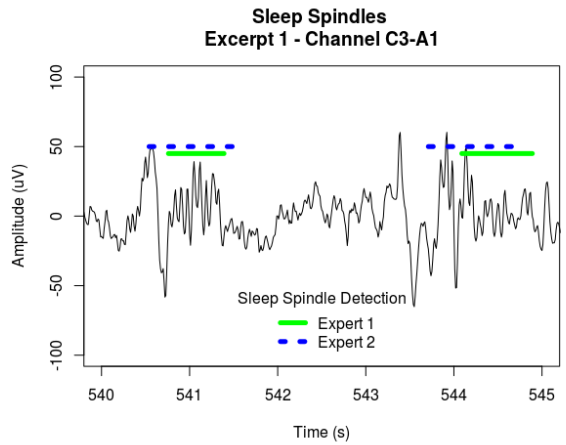


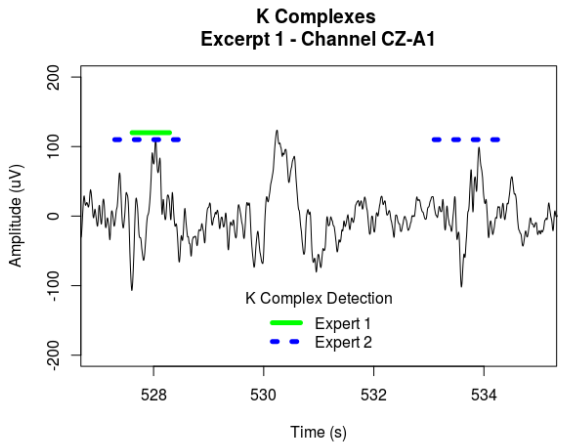
Why is it human competitive?

Criterion G - Difficult Problem

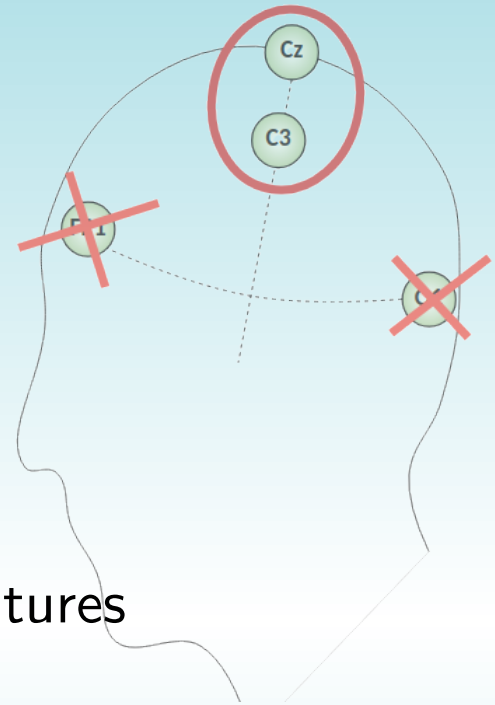


- It is challenging for both human specialists and automatic classifiers
  - For human specialists, is a tiresome, error-prone and possible biased process
- Sleep spindles in patients with sleep disorders can be more difficult to identify
- K complexes can be mistaken with any high peaks occurring in the EEG.



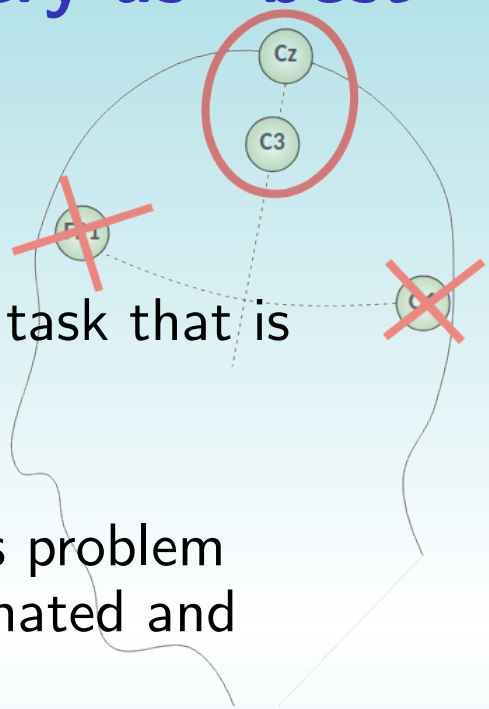


Significance of the Results



- Improvement of classifier's performance
- Dimensionality reduction: 75 features $\rightarrow$ 12 features
- Channel Reduction
  - Less EEG channels
  - Less discomfort examination

Why the judges should consider the entry as "best"



- Sleep phenomena identification is an important task that is still done by human experts.
  - The identification is not always the same!
- Our system finds new important features of this problem which will improve future results for both automated and human analysts.
  - The human specialist can exclude only the false positives, since our model avoids False Negatives
- We believe it is able to generalize and find similar results in other similar domains.
  - Can be easily applied on any event identification problem on time-series

Thank you!

Our code is available at GitHub