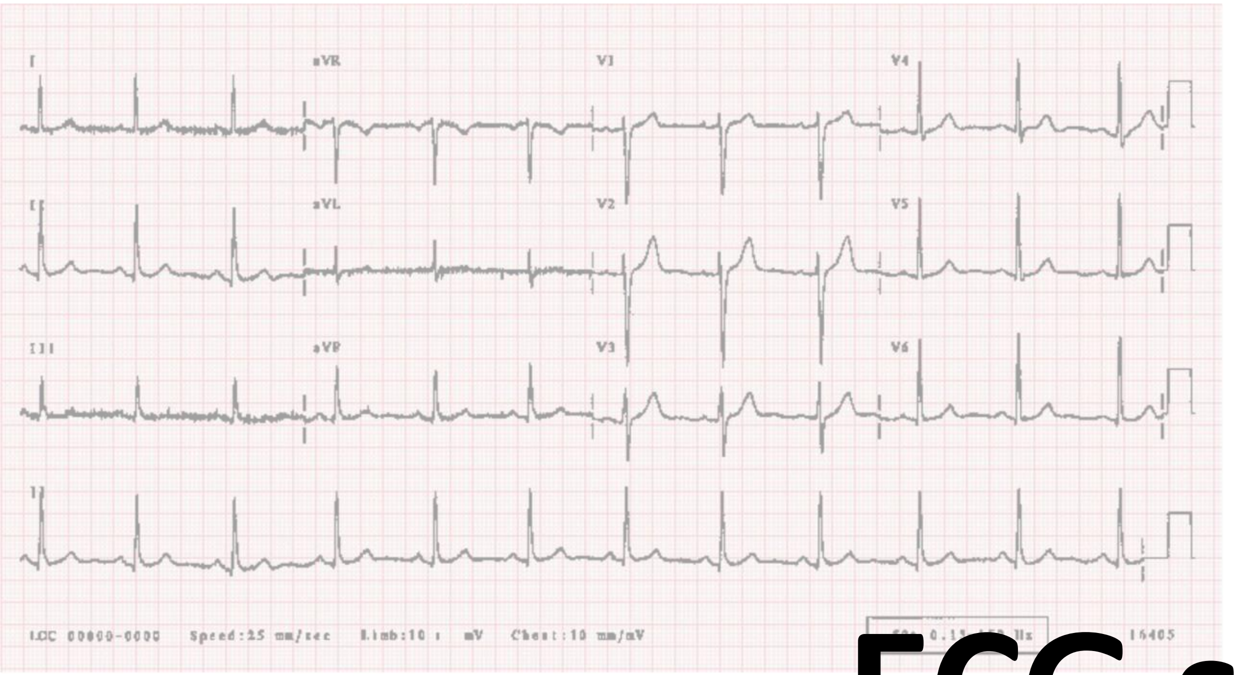
AML Project 2: Time-series classification

Yijiang Huang

yijiang.huang@inf.ethz.ch

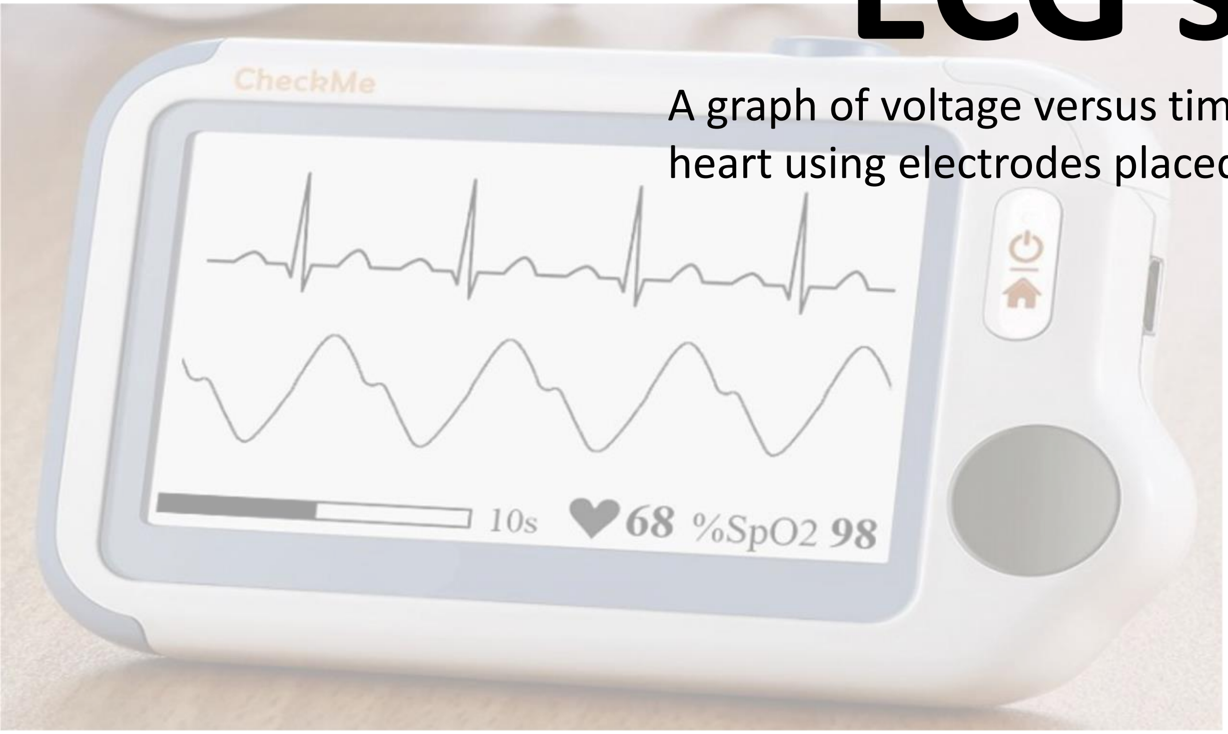
Nov. 11 – Dec. 2

Advanced Machine Learning, Autumn 2024

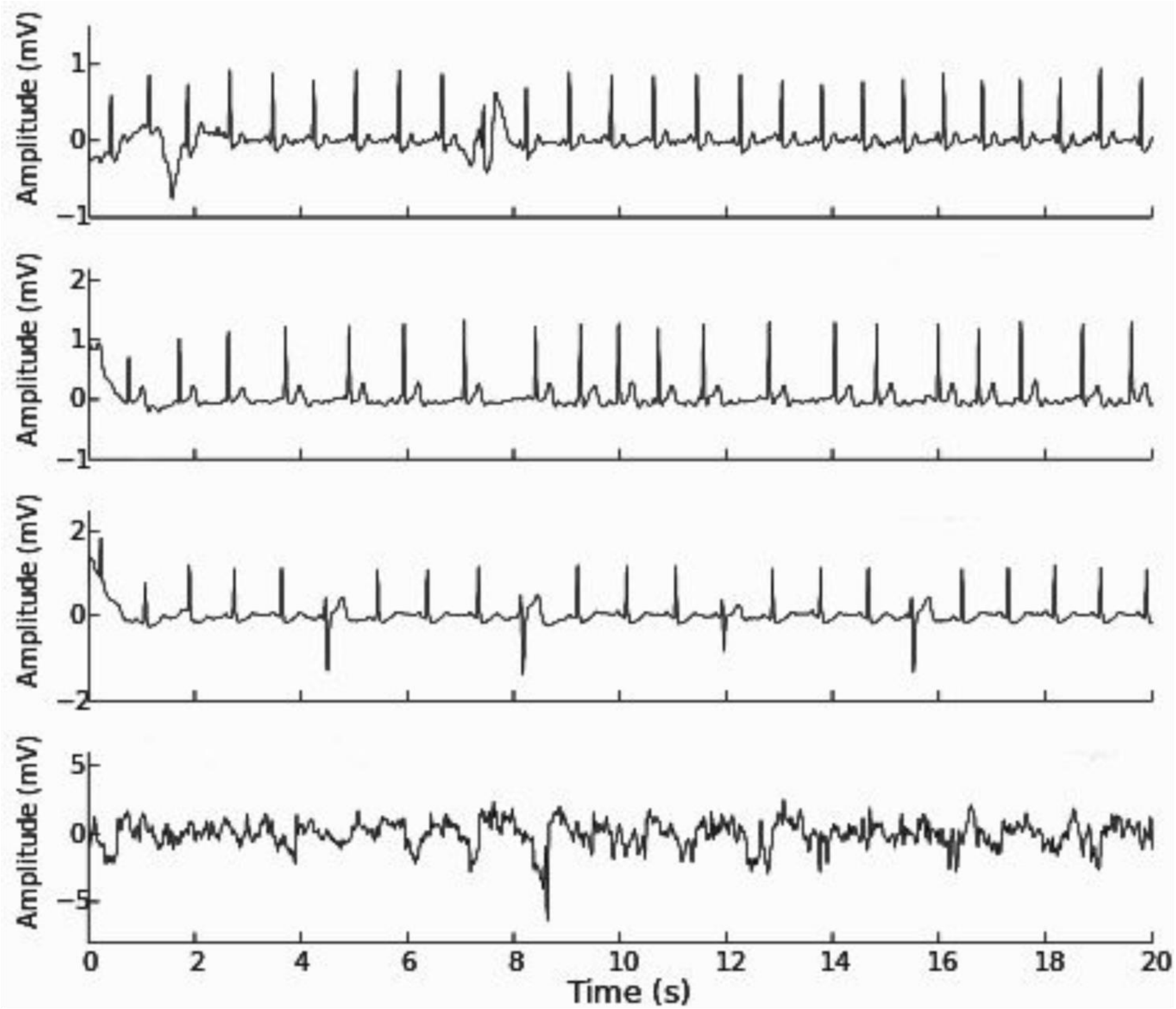


ECG signal

A graph of voltage versus time of the electrical activity of the heart using electrodes placed on the skin.



Input (time series) **Output** (classification)



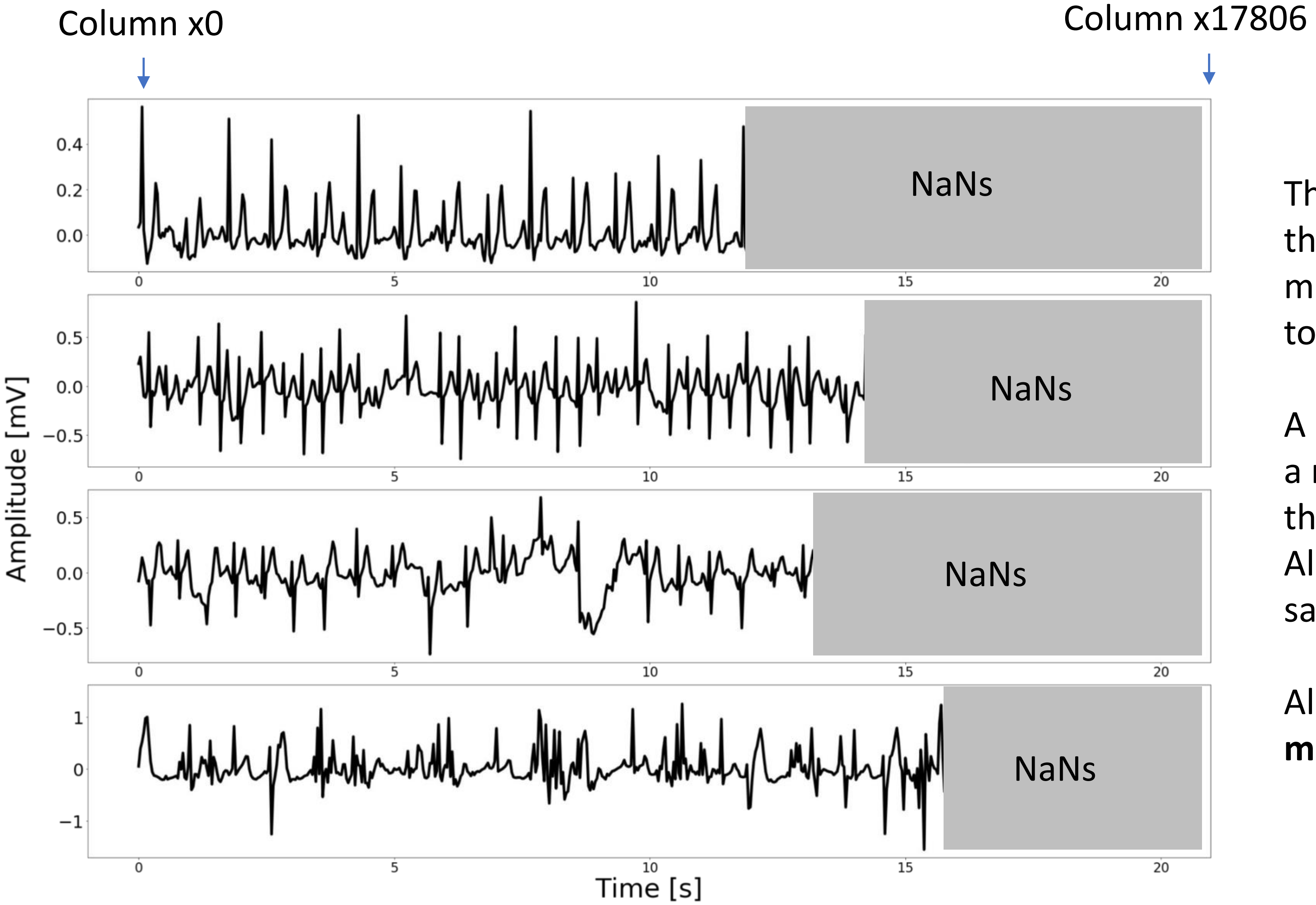
Class 0

Class 1

Class 2

Class 3

⚠ Variable signal length



The ECG data is a raw signal that corresponds to measurements at time stamps 0 to 17806.

A NaN shows up not because of a missing value but because of the variable length of the signal. All signals are sampled in the same frequency of 300Hz.

All columns **after the last measurement** have NaN values.

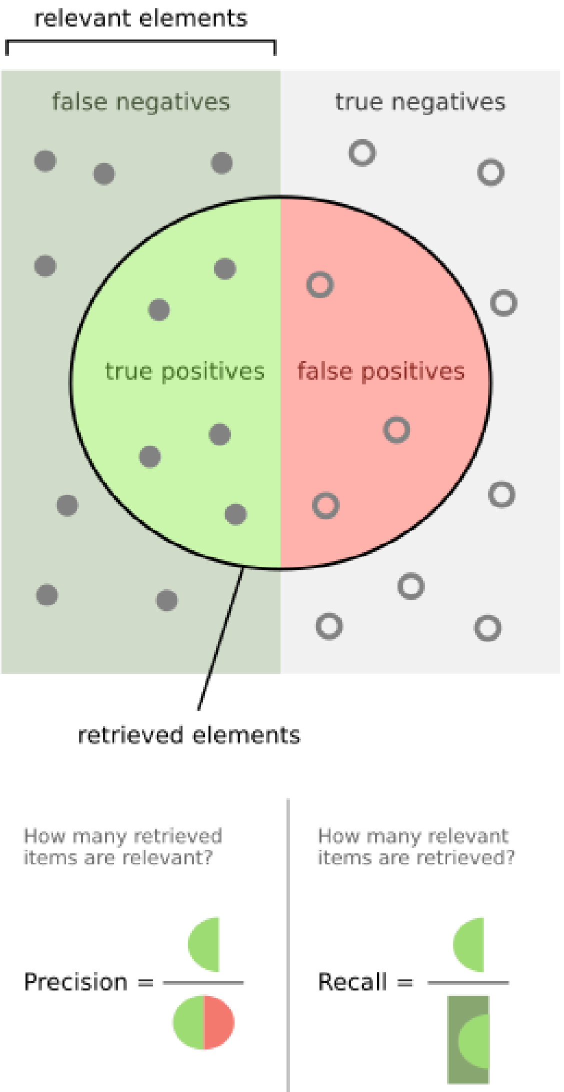
Evaluation Metrics

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

In python:

```
from sklearn.metrics import f1_score

F1 = f1_score(y_true, y_pred, average='micro')
```

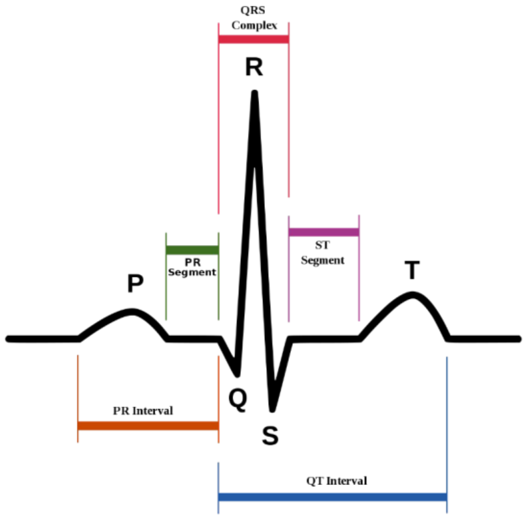


https://en.wikipedia.org/wiki/F-score

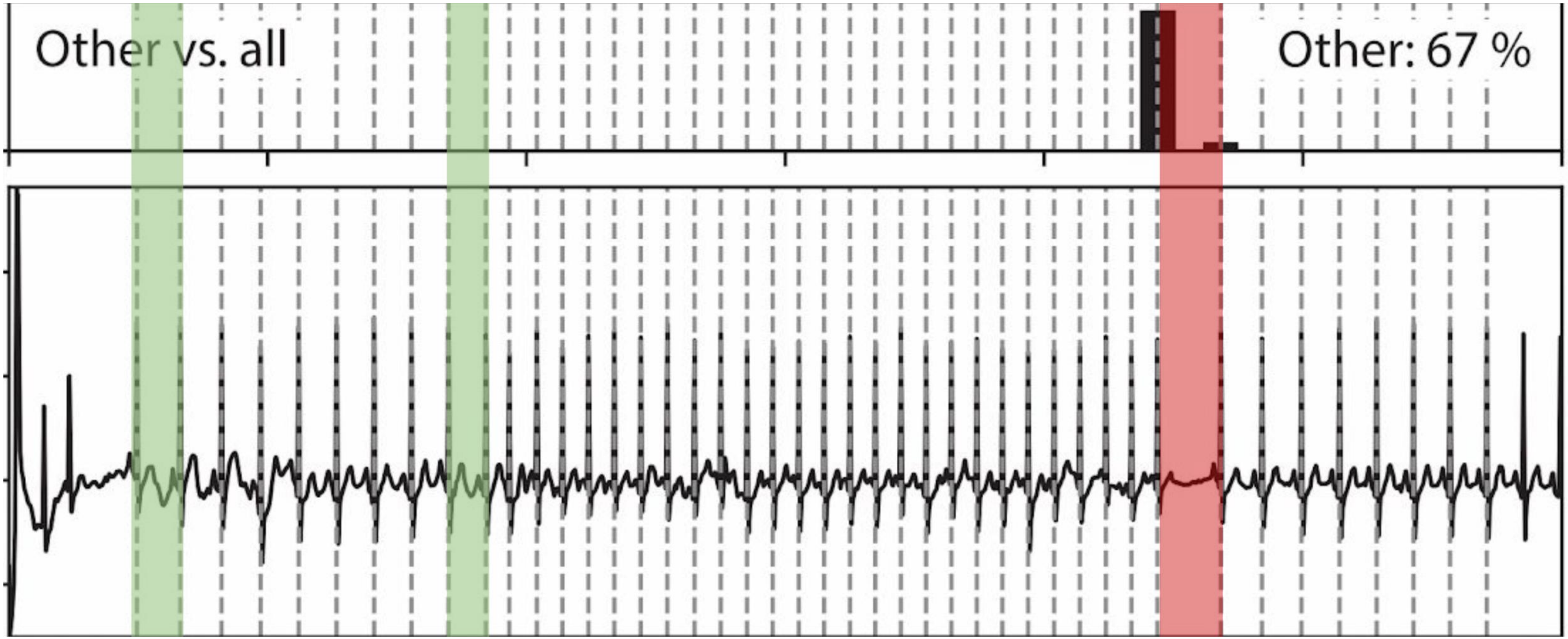
ECG is a sequence…

… of Data points

… of Heartbeats

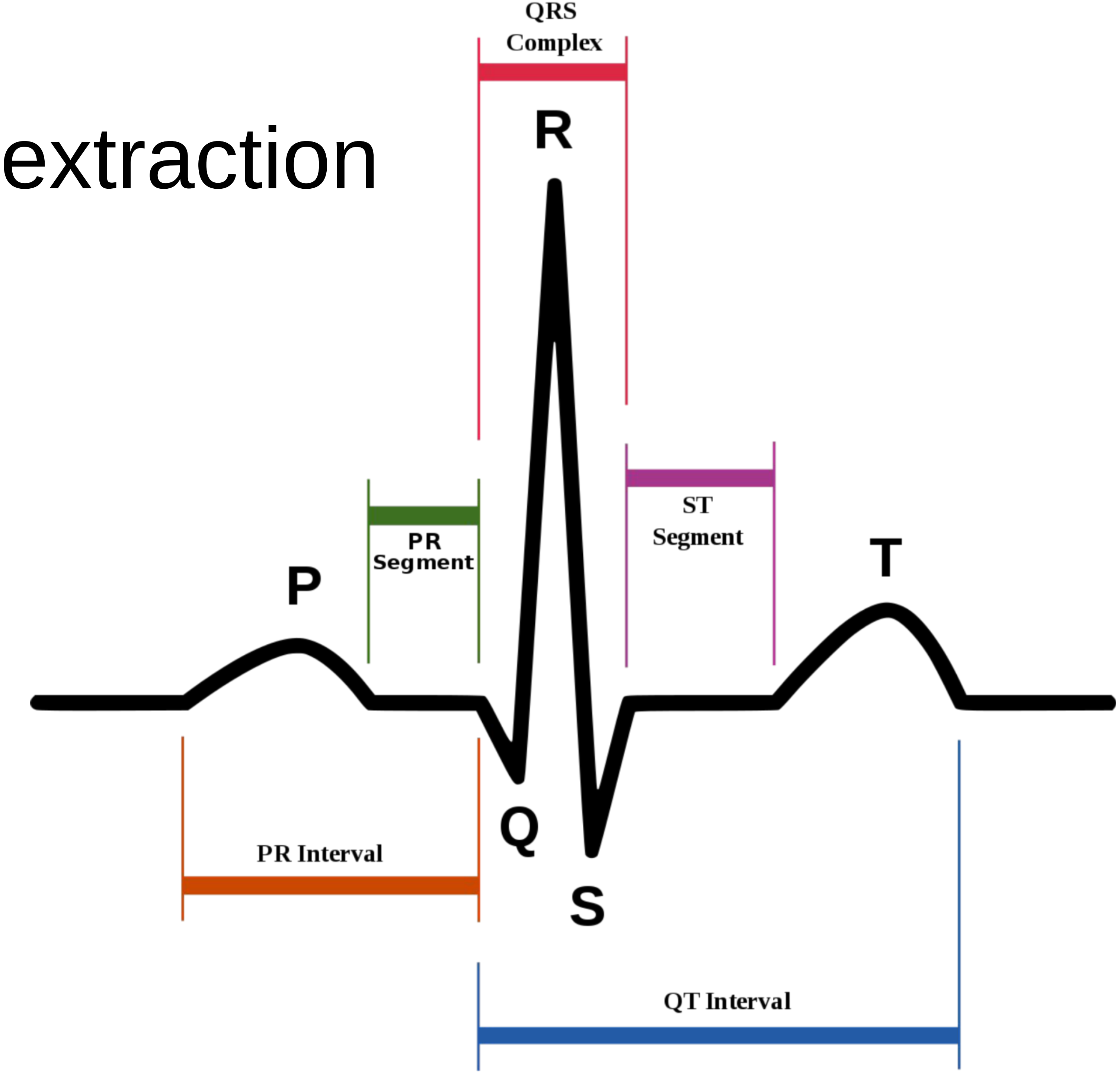


Splitting into Heartbeats

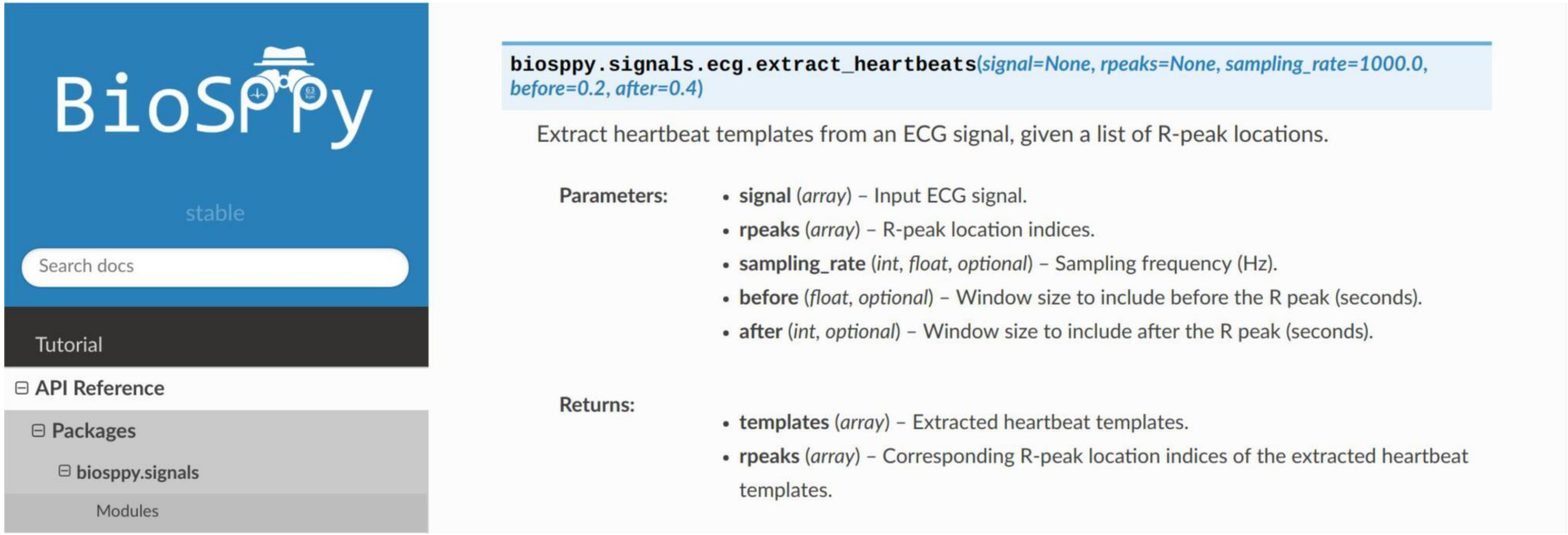


Manual feature extraction

- **RR** interval
- **R** amplitude
- **Q** amplitude
- **QRS** duration
- Heart rate variability
- Frequency domain



import biosppy.signals.ecg as ecg



biosppy.signals.ecg.extract_heartbeats(*signal=None, rpeaks=None, sampling_rate=1000.0, before=0.2, after=0.4*)

Extract heartbeat templates from an ECG signal, given a list of R-peak locations.

Parameters:

- **signal** (*array*) – Input ECG signal.
- **rpeaks** (*array*) – R-peak location indices.
- **sampling_rate** (*int, float, optional*) – Sampling frequency (Hz).
- **before** (*float, optional*) – Window size to include before the R peak (seconds).
- **after** (*int, optional*) – Window size to include after the R peak (seconds).

Returns:

- **templates** (*array*) – Extracted heartbeat templates.
- **rpeaks** (*array*) – Corresponding R-peak location indices of the extracted heartbeat templates.

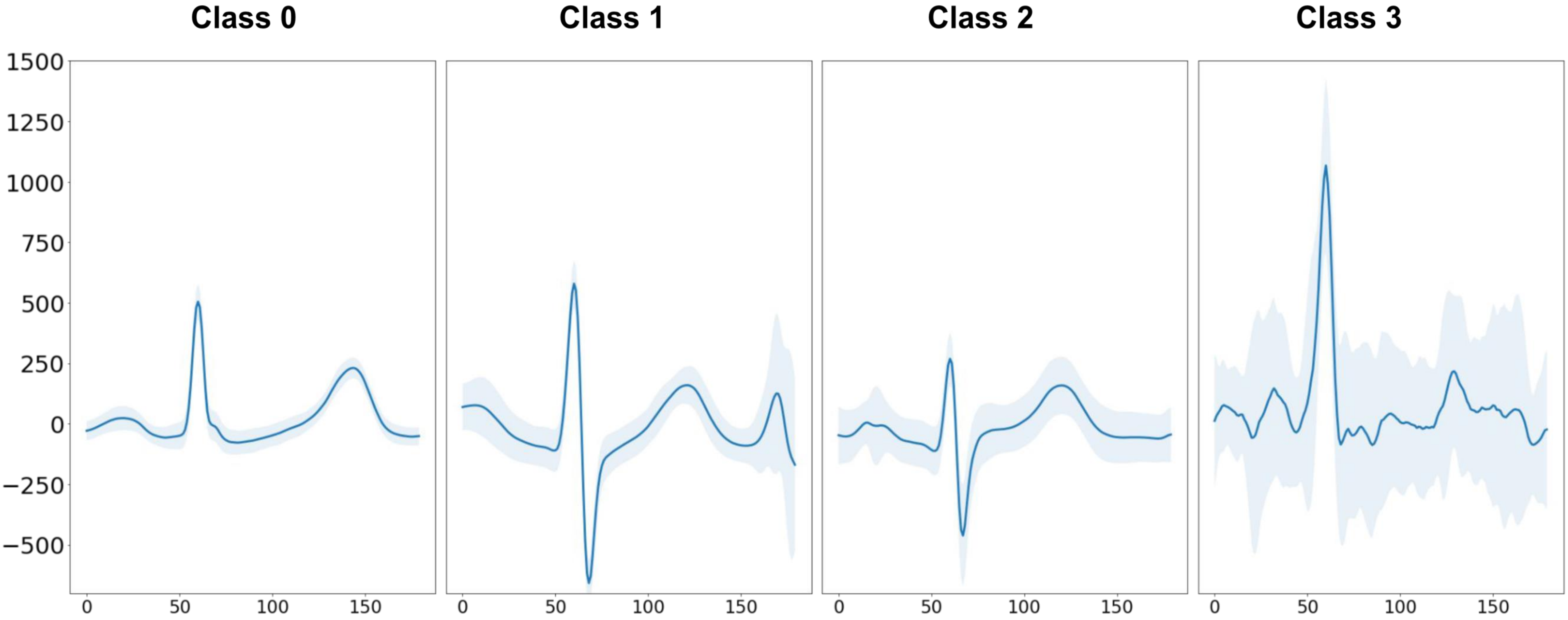
https://biosppy.readthedocs.io/en/stable/biosppy.signals.html#biosppy.signals.ecg.extract_heartbeats

We provide you an example Jupyter notebook using biosppy!

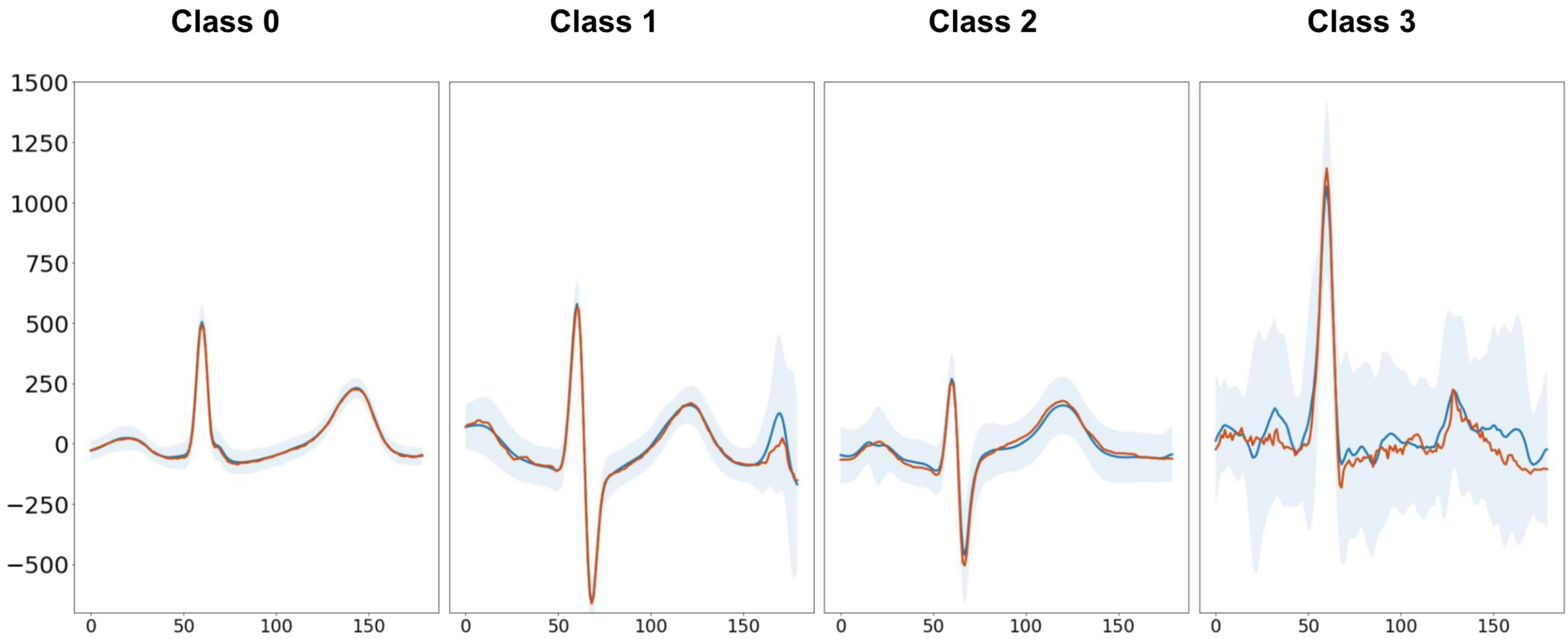
(you will download it with the data from Kaggle)

But feel free to experiment with other libraries of your choice:
Some other Python libraries: `neurokit`, `pyhrv`, `hrv`, `heartpy`, etc…
Wave extraction: ecg-kit (Matlab/Octave), ecgpuwave (Fortran, partially ported to Matlab/Octave)

Mean heartbeat with variance



Mean heartbeat with variance and median



Submission to Kaggle

- Kaggle link (you must log in Kaggle with your ETH email)
- Team names must be alphanumeric (A-Z, a-z, 0-9).
- **Each** member of a team must fill out the Project 2 form in Moodle **separately**.
  - Name of the team in Kaggle, legis, and a description of how you solved the task.
- You only submit a test prediction file to Kaggle, not code. But make sure your results are reproducible. We might ask for your code after the deadline.
- Project period: Nov 11th 3pm – Dec 2rd 2pm
- Public/private leaderboard:
  - Public leaderboard is available during the competition
  - Private leaderboard will open from Dec 2rd 3pm
- Public baseline for passing the project: 0.7
- Check the “Competition Rules” section on the Kaggle competition for more FAQs.

Moodle project form submission

To obtain points for this project, you have to describe the following in your Moodle form submission:

- You need to select one of your group's submissions for grading. You will only be graded based on this submission.
- You have to write a short description of the approach that you have used. Each student has to write their own description individually, and you **are not allowed to share the description** with your other group members.

If your submission does not include the points above, you will receive zero points for the project.

Considerations

- Do not use AutoML packages. if in doubt, always ask!
- Be aware of overfitting on the public test set
- Be careful about the submission time
  - Your group has a joint total **10** submissions per day.
- Describe what you did when you hand in the project
- Keep your implementation reproducible for potential review
- Do not wait until the last day to submit something
  - Servers usually get overloaded and crash causing long waiting times

Questions?