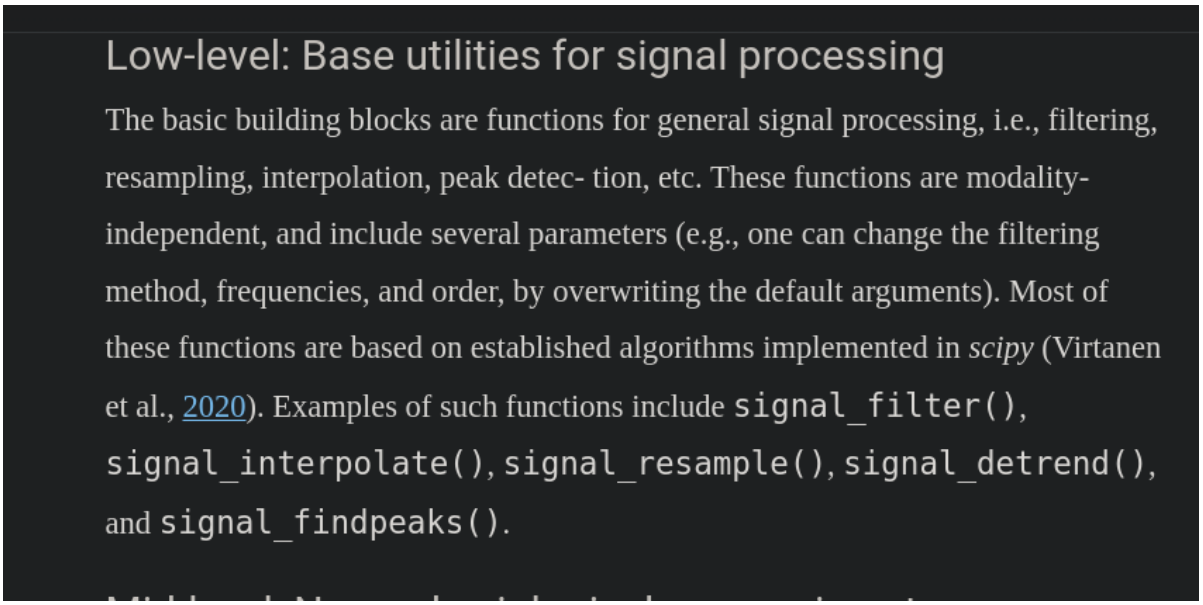
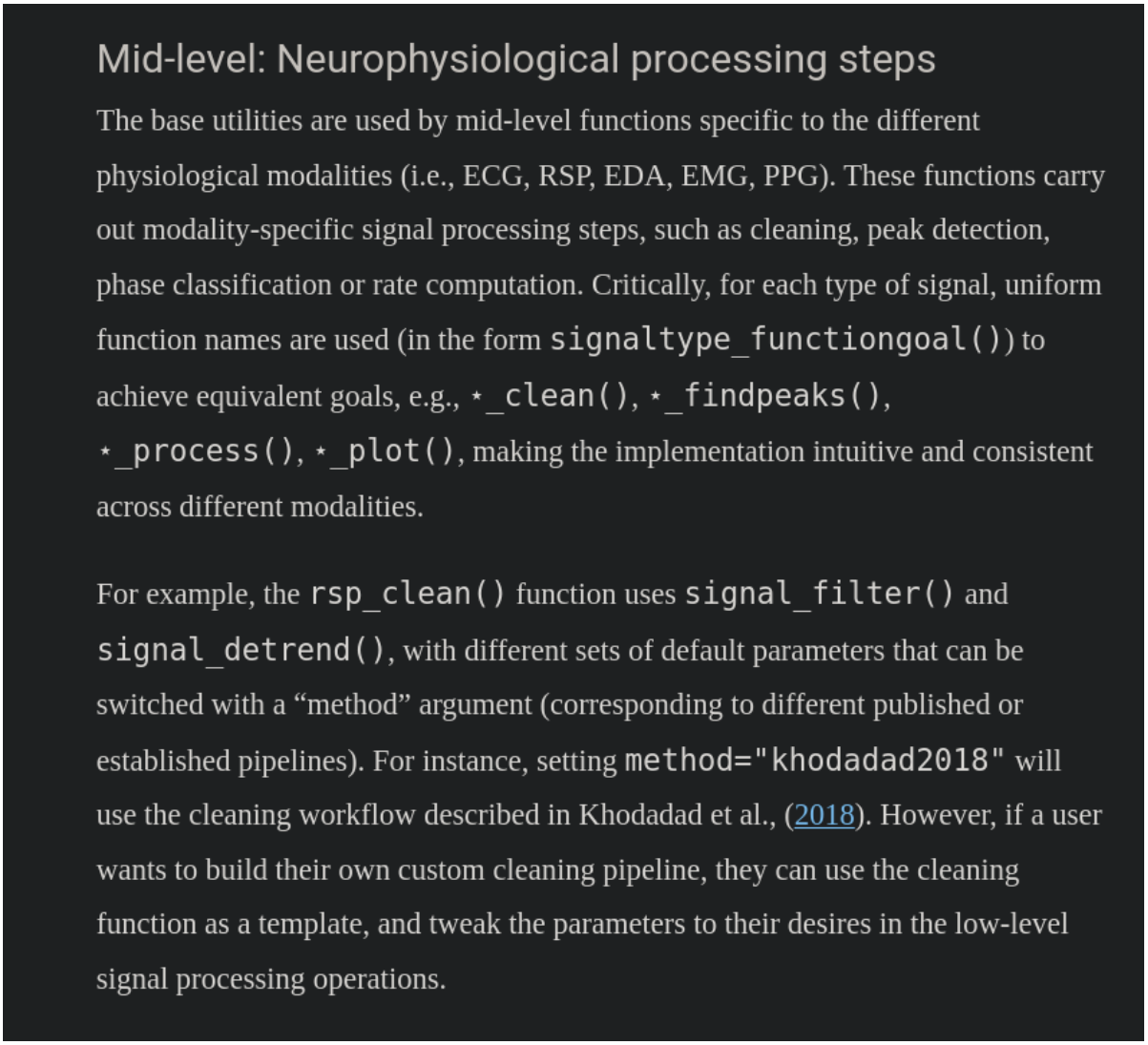
1.About nerokit2 library:

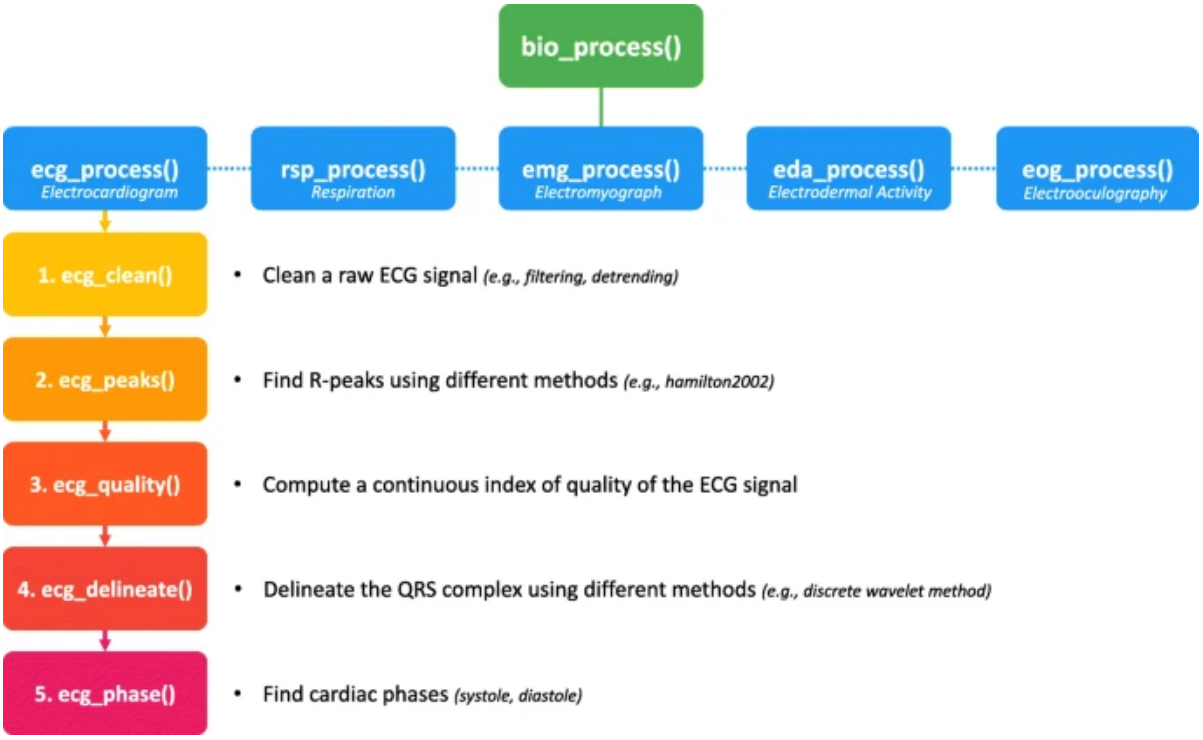
NeuroKit2 paper :
https://link.springer.com/article/10.3758/s13428-020-01516-y#additional-information





High-level wrappers for processing and analysis

The mid-level functions are assembled in high-level wrappers, that are convenient entry points for new users. For instance, the `ecg_process()` function internally chains the mid-level functions `ecg_clean()`, `ecg_peaks()`, `ecg_quality()`, `ecg_delineate()`, and `ecg_phase()`, as shown in Fig. 1. A specific processing pipeline can be selected with the `method` argument that is then propagated throughout the internal functions. Easily switching between processing pipelines allows for the comparison of different methods, and streamlines critical but time-consuming steps in reproducible research, such as the validation of data preparation and quality control (Quintana et al., 2016). Finally, the package includes convenience-functions (e.g., `bio_process`) that enable the combined processing of multiple types of signals at once (e.g., `bio_process(ecg=ecg_signal, eda=eda_signal)`).



2.ecg_peaks()

ecg_peaks(*ecg_cleaned*, *sampling_rate=1000*, *method='neurokit'*, *correct_artifacts=False*, ***kwargs*)
[source]
Find R-peaks in an ECG signal

Find R-peaks in an ECG signal using the specified method. The method accepts unfiltered ECG signals as input, although it is expected that a filtered (cleaned) ECG will result in better results.

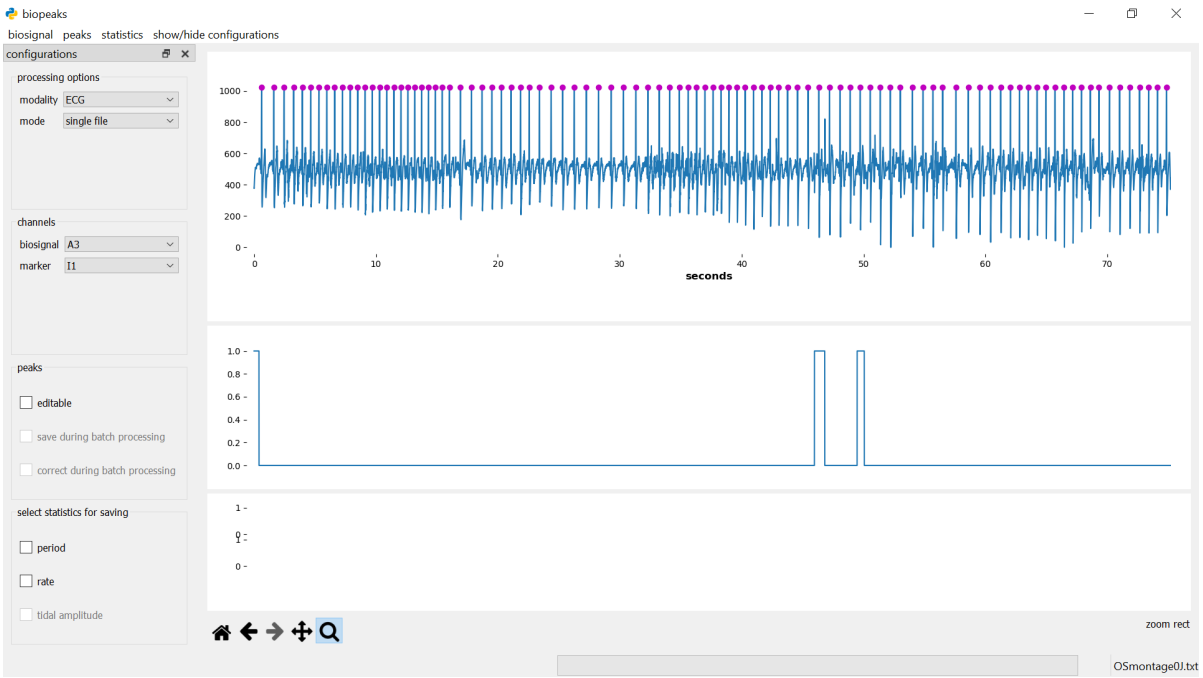
```
Different algorithms for peak-detection include:

  1. neurokit (default): QRS complexes are detected based on
     the steepness of the absolute gradient of the ECG signal.
     Subsequently, R-peaks are detected as local maxima in the
     QRS complexes. Unpublished, but see
     neuropsychology/NeuroKit#476

  2. pantompkins1985: Algorithm by Pan & Tompkins (1985).


  3. hamilton2002: Algorithm by Hamilton (2002).

  4. zong2003: Algorithm by Zong et al. (2003).

  5. martinez2004: Algorithm by Martinez et al (2004).

  6. christov2004: Algorithm by Christov (2004).

  7. gamboa2008: Algorithm by Gamboa (2008).

  8. elgendi2010: Algorithm by Elgendi et al. (2010).

  9. engzeemod2012: Original algorithm by Engelse & Zeelenberg
     (1979) modified by Lourenço et al. (2012).

  10.  kalidas2017: Algorithm by Kalidas et al. (2017).

  11.  nabian2018: Algorithm by Nabian et al. (2018) based on
     the Pan-Tompkins algorithm.

  12.  rodrigues2021: Adaptation of the work by Sadhukhan &
     Mitra (2012) and Gutiérrez-Rivas et al. (2015) by
     Rodrigues et al. (2021).

  13.  koka2022: Algorithm by Koka et al. (2022) based on the
     visibility graphs.
```

14. promac: ProMAC combines the result of several R-peak detectors in a probabilistic way. For a given peak detector, the binary signal representing the peak locations is convolved with a Gaussian distribution, resulting in a probabilistic representation of each peak location. This procedure is repeated for all selected methods and the resulting signals are accumulated. Finally, a threshold is used to accept or reject the peak locations. See this discussion for more information on the origins of the method: neuropsychology/NeuroKit#222

2.1 More about neurokit** (default):

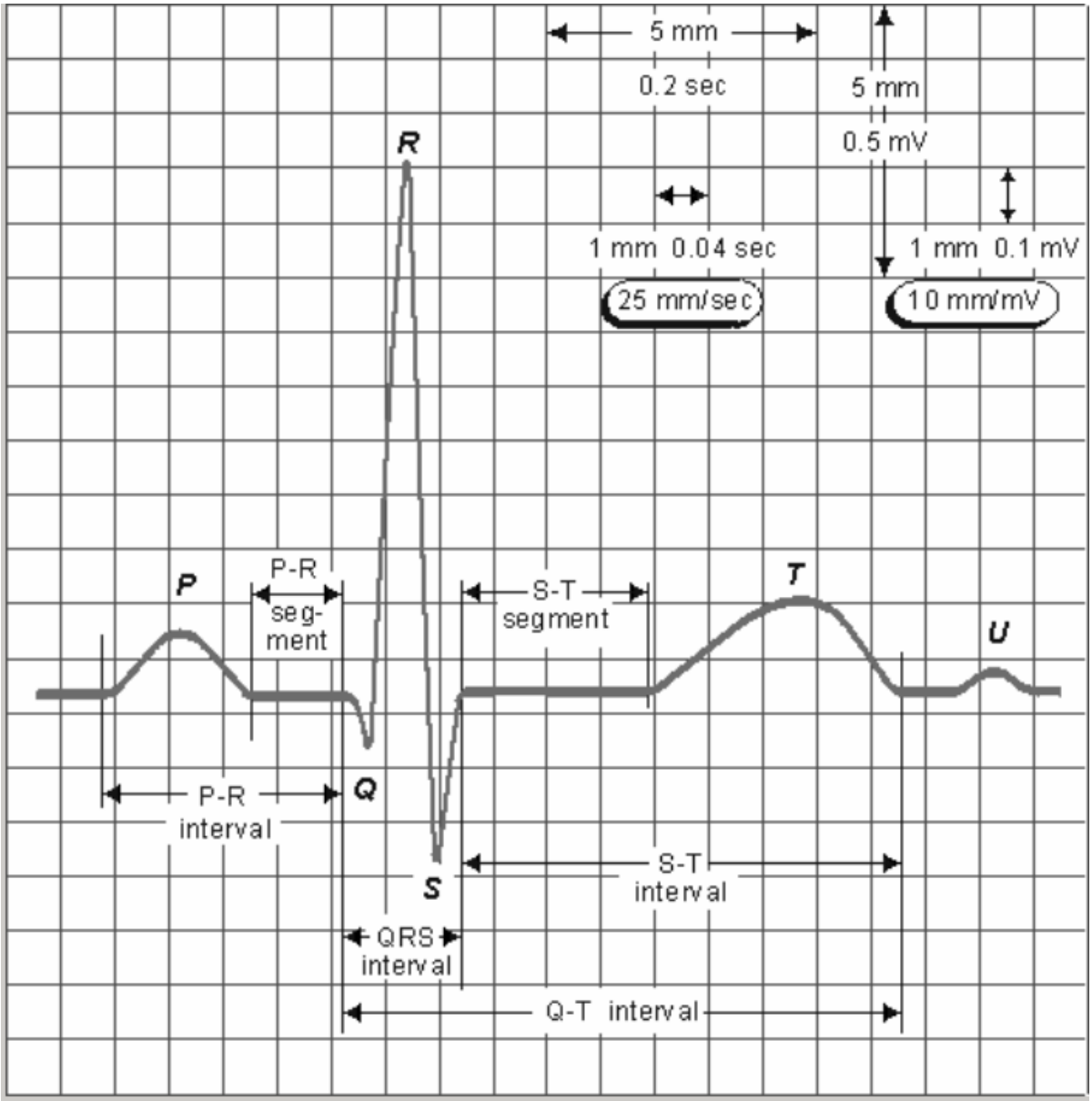
* **neurokit** (default): QRS complexes are detected based on the steepness of the absolute
gradient of the ECG signal. Subsequently, R-peaks are detected as local maxima in
the QRS complexes. Unpublished, but see https://github.com/neuropsychology/NeuroKit/issues/476

-> leads us to "biopeaks: a graphical user interface for feature extraction from heart- and breathing biosignals" paper



3.ECG GRID standard

[source]



The standard ECG paper with time intervals and segments.

The ECG is a plot of voltage on the vertical axis against time on the horizontal axis.

The ECG paper speed is ordinarily **25 mm/sec.** As a result, each **1 mm (small)** horizontal box corresponds to **0.04 sec (40 ms)**, with heavier lines forming larger boxes that include **five small boxes** and hence represent **0.20 sec (200 ms)** intervals.

horizontally :
On occasion, the paper speed is increased to **50 mm/sec** to better define waveforms. In this situation, there are only **six leads** per sheet of paper. Each **large box** is therefore only **0.10 sec** and each **small box** is only **0.02 sec**.

In addition, **the heart rate appears to be one-half of what is recorded at 25 mm/sec paper speed,** and all of the ECG intervals are twice as long as normal. Other paper speeds are occasionally used.

Vertically:
The ECG graph measures the **height (amplitude)** of a given wave or **deflection**. The standard calibration is **10 mm (10 small boxes), equal to 1 mV**.

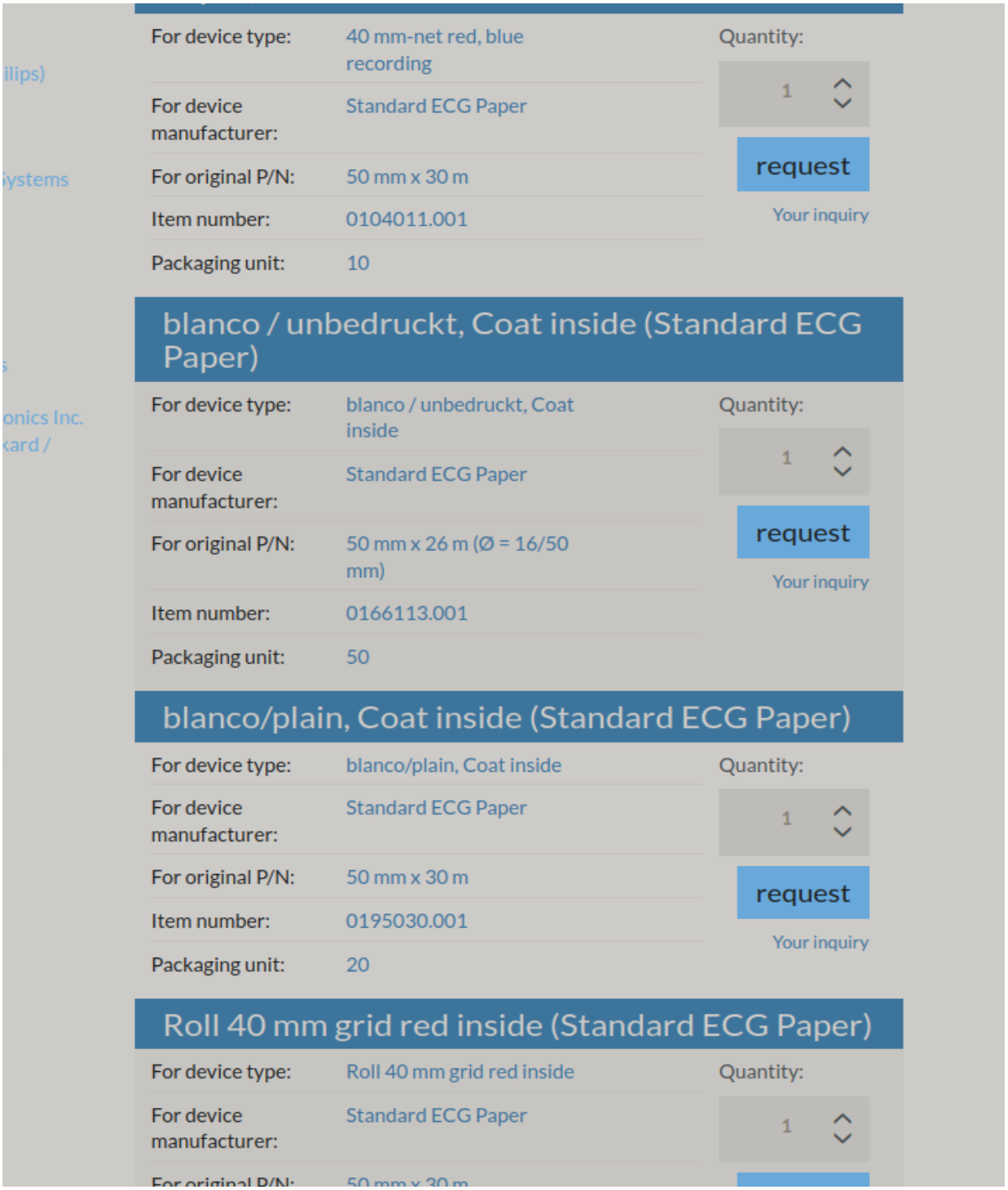
On occasion, particularly **when the waveforms are small**, **double standard is used (20 mm equals 1 mv)**.

When the waveforms are **very large**, **half standard may be used (5 mm equals 1 mv)**. Paper speed and voltage are usually printed on the bottom of the ECG for reference.

Examples:

1. [source]

> - **Waveform data**: 12-lead ECG signals recorded at 500 Hz for 10 s. The header file contains general information about the signal, such as sampling rate and units, including the column name of the signal, and the data file contains 12 signal information converted to 16 bits.



4.wave standard

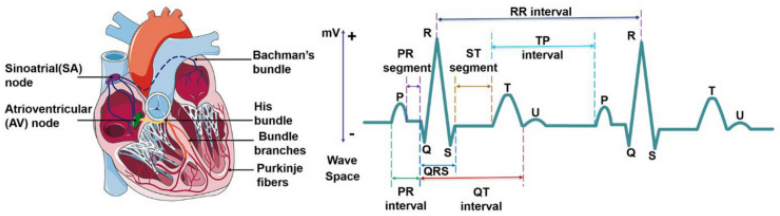


Table 1 ECG features and the normal values for a healthy adult.

Features	Description	Amplitude	Duration	Disease Diagnosis	References
R-R interval	The interval between two successive R-waves of the QRS complex ventricular rate		0.6–1.2 s	Paroxysmal atrial fibrillation Congestive heart failure	[45,46]
P wave	Atrial depolarization	0.25 mV	0.08–0.11 s	Atrial fibrillation Atrial hypertrophy	[47]
P-R interval	The time between the onset of atrial depolarization and the onset of ventricular depolarization		0.12–0.2 s	Stroke	[41]
QRS complex	Ventricular depolarization	1.60 mV for R peak	0.06–0.1 s	Ventricular enlargement Heart failure Tachycardia Acute Coronary Syndrome	[48,49,50]
ST-segment	The interval between ventricular depolarization and repolarization		0.05–0.155 s	Myocardial ischemia or infarction	[51]
T wave	Ventricular repolarization	0.1–0.8 mV	0.05–0.25 s	Myocardial infarction Pulmonary embolism	[46,52,53]
U wave	The last phase of ventricular repolarization	May not be observed because of its small size	Unknown	Unknown	[44]

QT interval	The time is taken for ventricular depolarisation and repolarisation		0.35–0.44 s	Hypokalemia ventricular arrhythmias	[54]

Open in a separate window

5.Data sets and feature extraction:

5.1 paper
[source]

Feature extraction: biosppy
Data set:

5.2 PhysioNet Challenge 2021 Data set:
[link]
2.1 The first source is the China Physiological Signal Challenge in 2018 (CPSC 2018):
10,330 ECGs.

Each recording is between 6 and 144 seconds long with a sampling frequency of 500 Hz.

2.2 The second source is the St Petersburg INCART 12-lead Arrhythmia Database:

contains 74 annotated ECGs.

extracted from 32 Holter monitor recordings. Each recording is 30 minutes long with a sampling frequency of 257 Hz.

2.3 The third source is the Physikalisch-Technische Bundesanstalt (PTB):

22,353 ECGs.

between 10 and 120 seconds long with a sampling frequency of either 500 or 1,000 Hz.

2.4 The fourth source is a Georgia database:

between 5 and 10 seconds long with a sampling frequency of 500 Hz.

2.5 The sixth source is the Chapman University, Shaoxing People's Hospital (Chapman-Shaoxing) and Ningbo First Hospital (Ningbo) database:

contains 45,152 ECGS.

Each recording is 10 seconds long with a sampling frequency of 500 Hz.

2.6 The seventh source is UMich Database from the University of Michigan:

19,642 ECGs .

Each recording is 10 seconds long with a sampling frequency of either 250 Hz or 500 Hz.

For example, a header file A0001.hea may have the following conte

```
A0001 12 500 7500 05-Feb-2020 11:39:16
A0001.mat 16+24 1000/mV 16 0 28 -1716 0 I
A0001.mat 16+24 1000/mV 16 0 7 2029 0 II
A0001.mat 16+24 1000/mV 16 0 -21 3745 0 III
A0001.mat 16+24 1000/mV 16 0 -17 3680 0 aVR
A0001.mat 16+24 1000/mV 16 0 24 -2664 0 aVL
A0001.mat 16+24 1000/mV 16 0 -7 -1499 0 aVF
A0001.mat 16+24 1000/mV 16 0 -290 390 0 V1
A0001.mat 16+24 1000/mV 16 0 -204 157 0 V2
A0001.mat 16+24 1000/mV 16 0 -96 -2555 0 V3
A0001.mat 16+24 1000/mV 16 0 -112 49 0 V4
A0001.mat 16+24 1000/mV 16 0 -596 -321 0 V5
A0001.mat 16+24 1000/mV 16 0 -16 -3112 0 V6
#Age: 74
#Sex: Male
#Dx: 426783006
#Rx: Unknown
#Hx: Unknown
#Sx: Unknown
```

5.3 PTB-XL, a large publicly available electrocardiography dataset
[paper] [dataset]

The PTB-XL ECG dataset is a large dataset of 21799 clinical 12-lead ECGs from 18869 patients of 10 second length. The raw waveform data was annotated by up to two cardiologists, who assigned potentially multiple ECG statements to each record. The in total 71 different ECG statements conform to the SCP-ECG standard and cover diagnostic, form, and rhythm statements. To ensure comparability of machine learning algorithms trained on the dataset, we provide recommended splits into training and test sets. In combination with the extensive annotation, this turns the dataset into a rich resource for the training and the evaluation of automatic ECG interpretation algorithms. The dataset is complemented by extensive metadata on demographics, infarction characteristics, likelihoods for diagnostic ECG statements as well as annotated signal properties.

5.4 PTB Diagnostic ECG Database

[source]
Duration: 10 seconds

The PTB Diagnostic ECG Database

- Number of Samples: 14552
- Number of Categories: 2
- Sampling Frequency: 125Hz
- Data Source: Physionet's PTB Diagnostic Database

6 Peak detection

6.1 R-Peak Detection Using Chaos Analysis in Standard and Real Time ECG Databases:

[source]

Conclusion

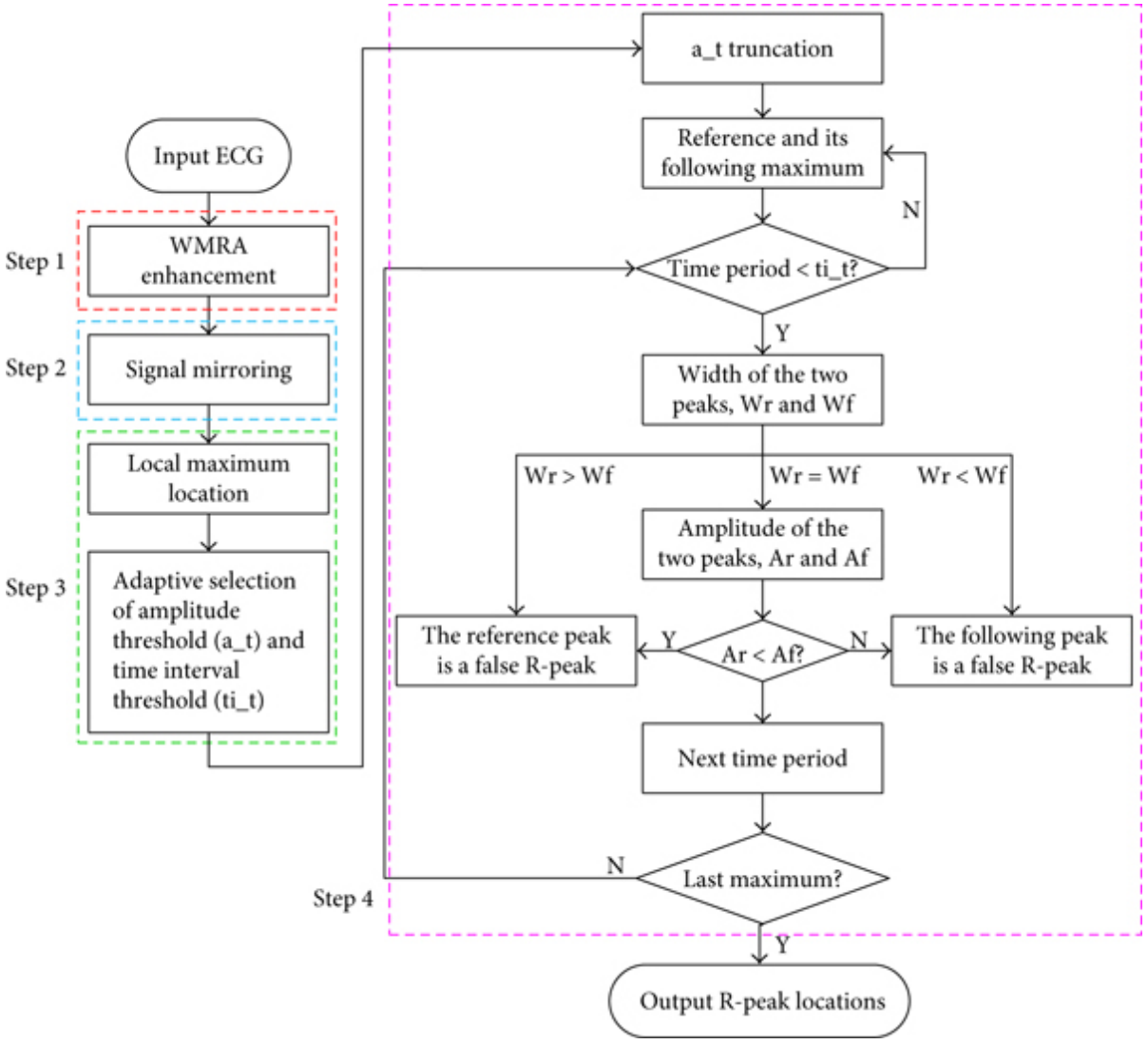
In this paper two techniques-(i) chaos analysis with PCA (without pre-processing) and (ii) IPCA, chaos analysis with PCA (with pre-processing) have been proposed and implemented successfully. Without pre-processing the proposed R-peak detection has yielded Se of 99.91%, PP of 99.93%, & DER of 0.163% for PN DB and Se of 99.77%, PP of 99.83%, & DER of 0.387% for RT DB, whereas with pre-processing the proposed R-peak detection has yielded Se of 99.95%, PP of 99.96%, & DER of 0.093% for PN DB and…

6.2

[source] Algorithm:

6. Conclusions

In this study, an adaptive and time-efficient methodology has been developed for automatic ECG R-peak detection. It is an adaptive method integrating WMRA, signal mirroring, local maximum detection, and amplitude and time interval thresholding. The accuracy performances were tested by using ECG records from MITDB and QTDB. Experimental results indicate that the proposed algorithm achieves average SEN, +P, and ACC of 99.39%, 99.49%, and 98.89% for MITDB, and 99.83%, 99.90%, and 99.73% for QTDB, respectively. In addition, time consumption and time complexity of the algorithm are computed to prove its time efficiency. By processing one ECG record, the average time cost is 0.872 s for MITDB and 0.763 s for QTDB, achieving 30.6% and 32.9%, respectively, of time reduction compared to the Pan-Tompkins method. Experiments on time complexity demonstrate that the proposed method is provided with linear time complexity; both our method and the Pan-Tompkins method are less sensitive to data increase.



7. Anomali detection

7.1 ECG-NET: A deep LSTM autoencoder for detecting anomalous ECG [source]

The aforementioned model has been applied on publicly available *ECG5000* dataset. From the experimental results it is observed that the proposed model achieved more than 98% accuracy having precision, recall and F1 values more than 0.94, 0.97, 0.96 respectively. The performance of the proposed method is also found to be superior in most of the cases as compared to the results of seven other recent counter-part methods reported in the literature.

Conclusions and future work

This article presents a LSTM based autoencoder model called *ECG-NET* for classification of normal and anomaly ECGs from time sequence heartbeat data. Encoder part of the proposed *ECG-NET* model is being trained only using normal ECG samples to learn the behavior of normal heartbeat from its encoded latent space representation, whereas the decoder part tries to reconstruct the original ECG signal form its encoded latent space representation. If the reconstruction loss value is low (less than a...

https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&as_vis=1&q=heart+arrhythmia+anomaly+detection+using+machine+learning+from+ecg&btnG=

https://ieeexplore.ieee.org/abstract/document/9110549/figures#figures

https://journals.plos.org/plosone/search?q=ecg&page=1&utm_content=a&utm_campaign=ENG-467