Atrial Fibrillation Detection

Bohdan Hlovatskyi

UCU. Signal Processing. December 2023

1 Introduction

Atrial fibrillation (AFib) is a common type of heart arrhythmia, or irregular heartbeat, that affects the upper chambers of the heart, known as the atria. In a normal heart rhythm, the atria contract in a coordinated and regular manner, pumping blood into the ventricles (the lower chambers of the heart). However, in atrial fibrillation, the electrical signals in the atria become chaotic, causing the atria to quiver or fibrillate instead of contracting effectively.

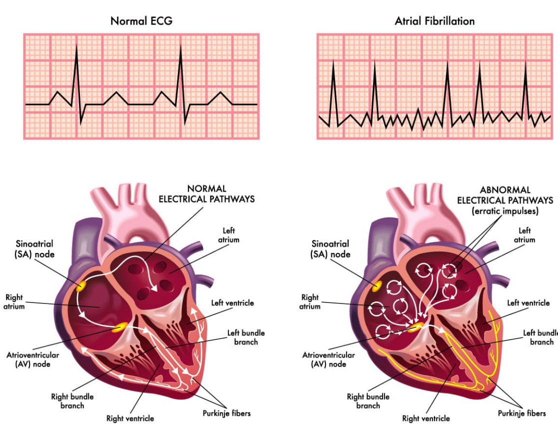


Figure 1: Atrial fibrillation

As a result of this irregular and rapid electrical activity, the atria may not effectively pump blood into the ventricles, leading to an irregular heartbeat. This can cause symptoms such as palpitations (feeling of rapid, fluttering, or pounding heartbeats), fatigue, dizziness, and shortness of breath. Additionally, AF increases the risk of blood clots forming in the atria, which can then travel to the brain and cause a stroke.

Arrhythmia is usually detected when patients seek medical attention due to the symptoms mentioned above or during regular check-ups via a stethoscope and pulse palpation, and are confirmed with a 24-hour ECG recording (Holter ECG). Analyzing this recording might be challenging to do manually, thus a more automatic approach is desirable. Ideally, it would be to perform this analysis on-device, in other words, being able to accurately detect AF from short ECG recordings.

2 Preliminaries

2.1 QRS complex

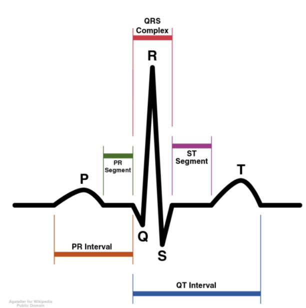
The essential elements of an ECG include the P wave, QRS complex, and T wave. Qrs complex is shown in Figure 2a. The most widely used feature that can be extracted from ECG is the RR interval, also called the inter-beat interval, which is the time elapsed between two consecutive R waves that can be directly used to calculate the average heart rate HR.

2.2 Two sample Kolmogorov-Smirnov test

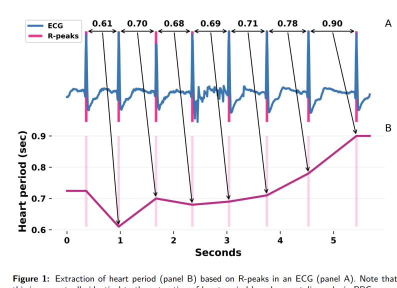
A probabilistic test to check whether two underlying one-dimensional probability distributions differ. The Kolmogorov-Smirnov statistic is:

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)|,$$

where $F_{1,n}$, $F_{2,m}$ are the empirical distribution functions of the first and the second sample respectively.



(a) QRS complex



(b) Heart period extraction

3 EDA

The training set, taken from [2] and provided by [1] contains 8,528 single lead ECG recordings lasting from 9s to just over 60s. ECG stands for Electrocardiogram. It is a diagnostic test that records the electrical activity of the heart over some time. The ECG provides important information about the heart's rhythm and electrical conduction, helping healthcare professionals assess the heart's health and function. A total of 12,186 ECG recordings were generously donated for this Challenge by AliveCor. Each recording was taken by an individual who had purchased one of three generations of AliveCor's single-channel ECG device". ECG recordings were sampled as 300Hz and they have been bandpass filtered by the AliveCor device. All data are provided in MATLAB V4 WFDB-compliant format (each including a .mat file containing the ECG and a .hea file containing the waveform information).

Type	# recording	Time length (s)				
		Mean	SD	Max	Median	Min
Normal	5154	31.9	10.0	61.0	30	9.0
AF	771	31.6	12.5	60	30	10.0
Other rhythm	2557	34.1	11.8	60.9	30	9.1
Noisy	46	27.1	9.0	60	30	10.2
Total	**8528**	**32.5**	**10.9**	**61.0**	**30**	**9.0**

Figure 3: Training dataset statistics

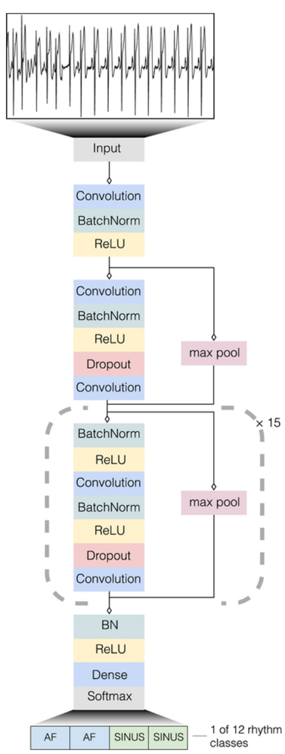
4 Literature review

Collecting ECG data is not a trivial task. Comprehensive analysis and diagnostics require long-time recordings to be processed and analyzed, as arrhythmia can occur sporadically. Moreover, it might be beneficial to analyze longer sequences as there might be some long-time correlation that cannot be detected from short-time recordings. Historically, there are two main benchmarks regarding arrhythmia detection - MIT-BIH Arrhythmia Database [3] and The PhysioNet in Cardiology Challenge 2017 [1]. In this work, the focus is made on the second one as being able to accurately detect arrhythmia from short-time recordings is beneficial for on-device monitoring as it does not require a lot of memory to store long sequences and thus can detect arrhythmia on the device without any need of a mainframe computer.

Classical methods are commonly based on a set of hand-crafted features that are used to train a simple machine-learning model, such as logistic regression, tree ensembles, etc. The foundational work by Tateno et. al [4] in this field utilizes such an approach of hand-crafted features. To detect AF each beat is analyzed in the window of 100 beats centered at the one being analyzed. The density histogram of RR is then compared to the standard density histogram. The similarity of those two is evaluated with the Kolmogorov-Smirnow test.

With slight modifications to this idea, they were able to achieve a sensitivity of 93.2% and a specificity of 96.7% on the MIT-BIH dataset [3]. Effectively extending this idea, one may utilize other hand-crafted features, such as QRS correlations, which can show an absence of P waves, which is considered to be a symptom of atrial fibrillation, PRT segments analysis, - some arrhythmias are defined in terms of prolonged segments, etc. One more important measurement is heart rate variability, which describes the variation between heartbeats.

While classical methods show good results, they might be hard to effectively tune to each patient thus being not particularly robust in real-world applications. With the advent of deep learning, arrhythmia detection has come to cardiologist-level accuracy. The most cited paper published in Nature by Hannum et al. [5], was published in Nature. The method proposed showed a performance hardly different from the one produced by a committee of cardiologists. Nevertheless, the architecture proposed was a simple CNN-based architecture, which did not utilize sequential information to its fullest (as it might be done with RNNs or Transformers). Looking at the newer methods for AF detection, one may consider the winners of corresponding challenges. Generally speaking, such methods do not utilize hand-crafted features much and rely on the original signal or its spectral properties directly. One important example of such architecture, an architecture being of the best currently at MIT-BIH dataset is a work by He et. all [6].



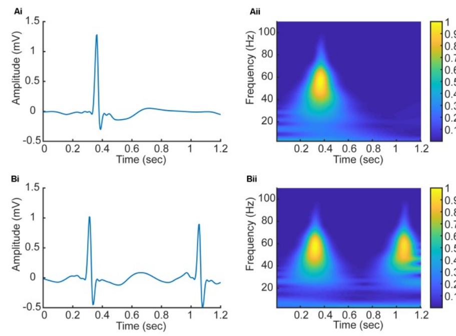
In the heart of their approach lies CNN which operates on the spectrogram images of individual beats. With such an approach, they get a stunning result of 99+

State-of-the-art method [7] follows this approach to some extent, additionally fusing an LSTM-based attention network to process the ECG sequence both in time and spectral domains. ECGNET allows not only to make the amount of input signal reach but also to process the whole sequence altogether, being able to find long-term connections between heartbeats. With those improvements, they can score 99.5% on the MIT-BIH dataset.

Looking back at the Atrial Fibrillation Detection on PhysioNet Challenge 2017, one might find both feature-based and DL-based approaches in the leaderboard. The best final score is 0.83, which a the true positive rate. One may argue that for a single dataset, it is easier to come up with a nice set of features, and for relatively simple tasks, it might be enough to score good results, thus the presence of such handcrafted feature-based methods is not a big surprise for this task.

Identifying Normal, AF, and other Abnormal ECG Rhythms using a Cascaded Binary Classifier [9]. It uses a set of hand-crafted features from PQRST detections that are fed to AdaBoost for final classification.

ENCASE: an ENsemble ClASsifiEr for ECG Classification Using Expert Features and Deep Neural Networks [8] uses both deep learning based and PQRST-based features. XGBoost is trained on the combination of them.



5 Method description

Given an ECG recording, typically 30s long, one needs to classify whether it is a normal heart rhythm or it has some atrial fibrillation features. Three methods of solving the task will be presented - a classical method featuring the Kolmogorov-Smirnov test, classical machine learning on top of hand-crafted features, and a deep learning based LSTM network. A brief explanation of how methods work will be presented in those subsections, analysis of results and possible shortcomings will be presented in the Results section.

5.1 Experiment setup & Evaluation

The classical ML approach of hold-out validation was chosen. As the test set for the dataset is not available, the train set has been split into train and validation subsets, in proportion 4:1. Main metric for evaluation was chosen to be F1 score with macro averaging.

5.2 Prepossessing

As a signal is already cleaned up in the dataset, no additional filtering is required. As proof of this, the amplitude spectral characteristic of an example signal is provided in Figure 4. It is worth noting, that if the signal hasn't been processed beforehand, a set of low-pass and high-pass filters could be designed easily from this spectral characteristic. One such filter would be a low pass filter that would cut off all the frequencies lower than 0.5 Hz to get rid of a drift effect while recording the signal. Another possibility would be to apply a stop band filter at 50 Hz to get rid of power line interference.

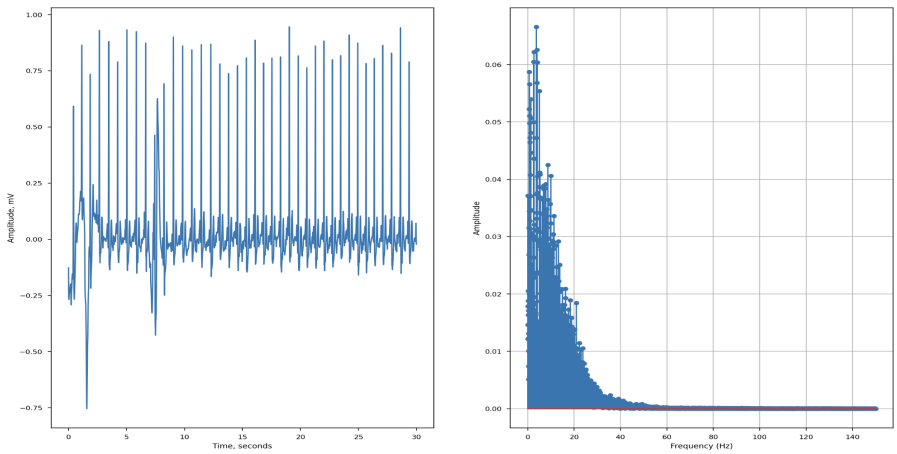


Figure 4: An ECG sample and its amplitude spectral characteristics.

5.3 Peak extraction

A vital part of ECG preprocessing that is often used in downstream applications is R-peaks and QRS complex extraction. There are a lot of algorithms aimed at R-peaks extraction, though the majority of them are based on those simple steps:

- filter a signal
- differentiate a signal to highlight sharp slopes of QRS complex
- apply a moving average
- determines a threshold, can be done adaptively or hardcoded
- detect QRS peak, if a value is greater than the threshold
- additional post-processing or hand-crafted rules to determine peaks.

Inspired by neurokit implementation of R-peaks detection, a custom R-peaks detection mechanism has been implemented:

Listing 1: Core of the R-peaks detection algorithm

```
grad = np.gradient(ecg_signal)
abs_grad = np.abs(grad)
moving_average = lambda (signal, window_size): \
    np.convolve(signal, np.ones((window_size,)) / window_size, mode="same")

smoothgrad = moving_average(abs_grad, int(np.rint(0.1 * SAMPLING_RATE)))

# adaptive threshold detection
```

```
avggrad = moving_average(smoothgrad, int(np.rint(0.75 * SAMPLING_RATE)))
qrs = smoothgrad > 1.5 * avggrad

# then R-peaks can be found as a local maximum in red regions,
# displayed in Figure 4.
```

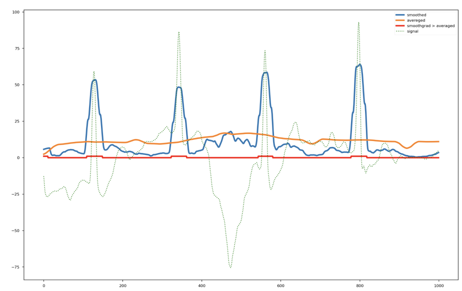


Figure 5: A process of peaks extraction from an ECG sample

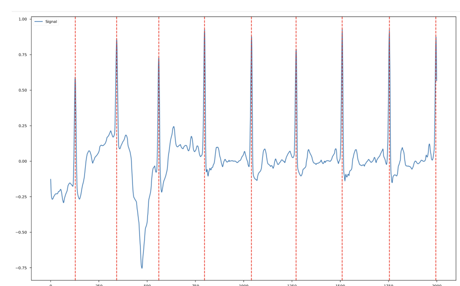


Figure 6: Example of extracted peaks from an ECG sample

5.4 Classical method

A method inspired by [4] was implemented. Given an ECG recording and the corresponding label, R-peaks of a signal are detected. Afterward, the time of RR intervals is found and stored in one of two lists determined by a label that stores RR intervals that come from a recording that is classified as a normal one and a recording that is classified as having signs of atrial fibrillation. Each of the RR intervals within one list might be seen as an independent identically distributed random variable. Thus, having collected enough amount of those, we can approximate the cumulative distribution function of a corresponding random variable (basically we have two random variables that describe RR intervals with two underlying distributions - ”normal” and ”atrial fibrillation”).

One might notice in Figure 7 that the histogram for RR-intervals that are recorded from ECGs classified as containing signs of atrial fibrillation has some

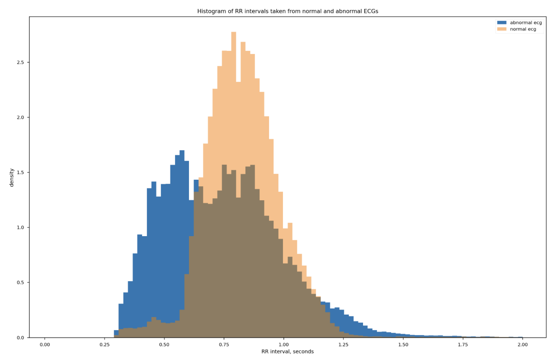


Figure 7: Histograms of RR-interval time distributions

kind of bimodal distribution with one of the mods being aligned with a Gaussian that can be aligned with the histogram of RR-s coming from normal ECGs.

Having collected all the necessary statistics, the two-sample Kolmogorov-Smirnov test can be applied to test for each new ECG to which of the two underlying distributions it is more likely to belong.

5.5 Machine learning

As a baseline Logistic Regression was chosen. Given an ECG signal, we flatten it into a single vector of dimensionality 9000, which is an optimal length in terms of cutting/padding a signal. Then simple end-to-end training is performed. To improve on top of it, hand-crafted features were created via the neurokit2 library [10], including all the peaks from the QRS complex, their on- and offsets, etc. Alongside the original signal, those are fed to the same Logistic Regression Model. As a final step, XGBoost is trained on the same set of features.

5.6 Deep learning

A custom variant of the CNNLSTM model has been trained. It features convolution modules at the beginning to reduce the dimensionality of a signal, bidirectional LSTM is then applied with two fully connected layers on top of it to perform classification. Occasional batch norms are added to stabilize the training.

Listing 2: CNNLSTM custom model

```
ECGLSTM(
  (conv1): Conv1d(1, 3, kernel_size=(3,), stride=(3,))
  (conv2): Conv1d(3, 9, kernel_size=(3,), stride=(2,))
  (bn1): BatchNorm1d(9, eps=1e-05, momentum=0.1, affine=True)
  (conv3): Conv1d(9, 9, kernel_size=(3,), stride=(2,))
  (conv4): Conv1d(9, 27, kernel_size=(3,), stride=(2,))
  (conv5): Conv1d(27, 27, kernel_size=(3,), stride=(2,))
```

```
  (conv6): Conv1d(27, 81, kernel_size=(3,), stride=(2,))
  (bn2): BatchNorm1d(81, eps=1e-05, momentum=0.1, affine=True)
  (lstm): LSTM(81, 256, num_layers=2, batch_first=True, bidirectional=True)
  (fc1): Linear(in_features=1024, out_features=256, bias=True)
  (bn3): BatchNorm1d(256, eps=1e-05, momentum=0.1, affine=True)
  (fc2): Linear(in_features=256, out_features=2, bias=True)
)
```

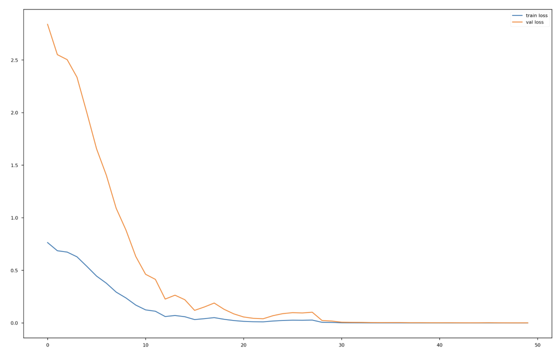


Figure 8: Train and validation losses for CNNLSTM

5.7 Worth mentioning

An interesting approach explored is the possibility of training a CNN on top of individual heartbeats. It has not been completed due to the absence of labels for each individual heartbeat, though in practice it would be possible to train such a model that would predict the label for each separate heartbeat given its discrete wavelet transform.

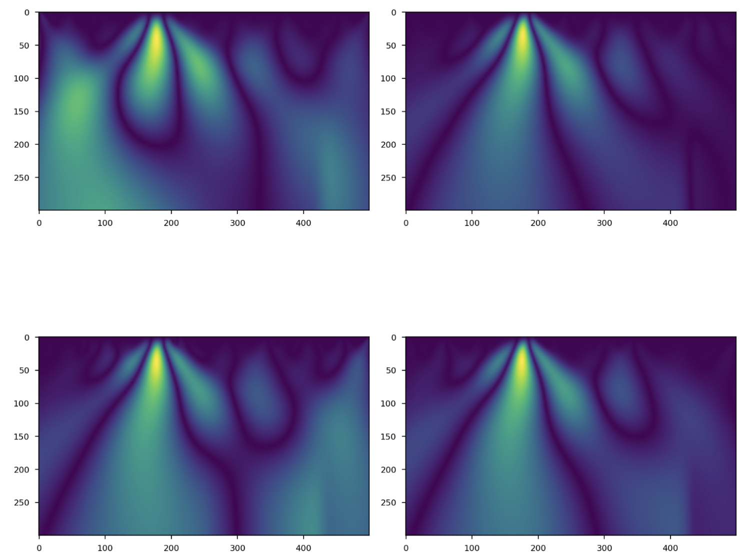


Figure 9: Example of extracted dwt transforms for individual hearth beats. X-axis - index of a sample, Y-axis - frequency, Hz

6 Results

Surprisingly enough, the classical method with the KS Test showed a strong baseline with an F1 value of 0.6. As bimodal distribution arises from RR intervals taken from ECGs that are classified as having arrhythmia, it is relatively easy to math samples containing short-time RR intervals to it and thus make a confident prediction regarding the presence of an arrhythmia. Logistic regression has shown to be inferior to this approach, as it does not imply any high-level hand-crafted information such as distribution of RR intervals and thus learns from the data solely. As the dimensionality of data is quite high, SGD is used for optimization, which makes it hard to find the optimal solution. At the same time, leveraging a more advanced model in the face of XGBoost gave a strong improvement over the baseline almost by 0.1 F1 score. Deep Learning methods are shown to be the most advanced with a top value of 0.75 F1. It is worth noting, that it reached almost zero loss both at training and validation, suggesting that some ECG recordings and noisy or mislabeled.

Method	**F1 Result**
KS Test	**0.6**
Logistic Regression	0.48
Logistic Regression with add. feat.	0.56
XGBoost	**0.69**
CNNLSTM	**0.75**

Table 1: F1 Results for different Methods

7 Future work

Several possible directions might be further explored to boost the performance even more. First of all, the signal's spectral properties were hardly used. This might be combined with deep learning model training on several datasets to boost its generalization abilities. As was shown in the 5.7, one might train a CNN model at discrete wavelet transform or spectrogram of individual heartbeats and use those deep features as additional input to the classification model. Regarding the classical methods, more advanced feature engineering can be used. It would require, though, deeper knowledge about the domain.

8 Conclusion

Atrial fibrillation is one of the most common types of heart arrhythmia. It is dangerous as it increases the risk of a stroke. Approximately five percent of the population is diagnosed with heart arrhythmia. As it would be ideal to be able to conduct all-day monitoring on-device, efficient and accurate methods are needed for arrhythmia detection from short ECG recordings, typically thirty seconds long. In this work, several approaches to Atrial Fibrillation detection were implemented and discussed. They were compared on a dataset provided by PhysioNet in Cardiology Challenge 2017. While methods with hand-crafted features give good performance, they require a lot of domain knowledge, thus making deep learning approaches superior in terms of accuracy for a beginner.

References

[1] Clifford GD, Liu C, Moody B, Li-wei HL, Silva I, Li Q, Johnson AE, Mark RG. AF classification from a short single lead ECG recording: The PhysioNet/computing in cardiology challenge 2017. In 2017 Computing in Cardiology (CinC) 2017 Sep 24 (pp. 1-4). IEEE. https://doi.org/10.22489/CinC.2017.065-469.

[2] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet:

Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220.

[3] Moody GB, Mark RG. The impact of the MIT-BIH Arrhythmia Database. IEEE Eng in Med and Biol 20(3):45-50 (May-June 2001). (PMID: 11446209).

[4] Tateno, K., amp; Glass, L. (n.d.). A method for detection of atrial fibrillation using RR Intervals. Computers in Cardiology 2000. Vol.27 (Cat. 00CH37163). https://doi.org/10.1109/cic.2000.898539

[5] Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., amp; Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine, 25(1), 65–69. https://doi.org/10.1038/s41591-018-0268-3

[6] He, R., Wang, K., Zhao, N., Liu, Y., Yuan, Y., Li, Q., amp; Zhang, H. (2018). Automatic detection of atrial fibrillation based on continuous wavelet transform and 2d Convolutional Neural Networks. Frontiers in Physiology, 9. https://doi.org/10.3389/fphys.2018.01206

[7] Mousavi, S. (2019). ECGNET: Learning Where to Attend for Detection of Atrial Fibrillation (AF) with Deep Visual Attention. https://doi.org/10.31224/osf.io/m9teb

[8] Hong, S., Wu, M., Zhou, Y., Wang, Q., Shang, J., Li, H., amp; Xie, J. (2017). ENCASE: An ensemble classifier for ECG classification using expert features and Deep Neural Networks. Computing in Cardiology Conference (CinC). https://doi.org/10.22489/cinc.2017.178-245

[9] S. Datta et al., "Identifying normal, AF and other abnormal ECG rhythms using a cascaded binary classifier," 2017 Computing in Cardiology (CinC), Rennes, France, 2017, pp. 1-4, doi: 10.22489/CinC.2017.173-154.

[10] Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., Chen, S. A. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. Behavior Research Methods, 53(4), 1689-1696. https://doi.org/10.3758/s13428-020-01516-y