2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-2782-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/EMBC48229.2022.9871227

CNN-based Two Step R Peak Detection Method: Combining Segmentation and Regression

Jaeseong Jang[1], Seongjae Park[1], Jin-Kook Kim[1], Junho An[1], and Sunghoon Jung[1*]

Abstract— **For semantic segmentation, U-Net provides an end-to-end trainable framework to detect multiple class objects from background. Due to its great achievements in computer vision tasks, U-Net has broadened its application to biomedical signal processing, especially, segmentation of waveforms in ECG signal. Despite its superior performance for QRS complex detection to other traditional signal processing methods, direct application of the U-Net to R peak detection has limitation since the U-Net structures tend to predict high probability around true peak. Such multiple detection results require additional process to determine a unique peak location in each QRS complex. In this study, we use a regression process to detect R peak instead of pixel-wise classification. Such regression process guarantees a unique peak location prediction. We collect data from resting ECG systems and wearable ECG devices as well as public ECG databases and the proposed model is trained on various combinations of the data sources. Especially, we investigate the robustness of the model for input data from the wearable devices when the model is trained by data from heterogeneous devices.**

I. INTRODUCTION

ECG recordings enable clinicians to obtain information of cardiac cycle events through graphs of electrical potential measured on the skin. The recordings are performed in various systems, such as resting ECG, portable ECG systems. Although resting ECG provides a simple way to evaluate the heart, extended recording period of the portable systems has more reliable clinical evidence [1], [2]. Especially, as recent progress of wearable devices provides longer recording period, laborious manual reading of ECG signal is a challenge for practical application of long period recordings of ECG for clinical decision. Accordingly, the need for robust and precise analysis methods has arisen for automatic reading process for ECG recordings.

In an ECG signal, cardiac cycle events are presented as characteristic waveforms, such as P wave, QRS complex, R peak, and T wave. Detection and localization of such waveforms are key steps for automatic analysis of ECG signal. Based on development of biomedical signal processing techniques, model-based methods have been suggested for the detection and localization tasks. However, for ECG signal acquired by wearable devices, inherent low signal-to-noise ratio (SNR) and patient dependancy of waveforms are inevitable so that it is hard to guarantee robust performance of such traditional methods [3]. Finding a universal method for the mentioned environments requires a highly non-linear process and decision thresholding, which is hard to be modelled.

Deep learning architectures have shown great achievements compared to model-based approaches, due to their huge capability to learn highly non-linear function. Especially, convolutional neural networks (CNNs) have become a powerful tool for image and signal processing as it shows superior peformance to traditional methods. By convolving features with accumulated layers, a CNN structures detect complex features which is useful in representing main features of target class or evaluating quantities for classification or regression. As a special structure of CNN, fully convolutional neural networks (FCNNs) were proposed to perform pixel-wise classification, i.e. semantic segmentation. Providing end-to-end trainable frameworks, FCNNs, especially U-Net, have broaden their application to biomedical signal processing. U-Net uses a series of encoding blocks and decoding blocks to produce semantic segmentation map [4]. There has been attempts to apply the U-Net structure for waveform detection tasks in ECG signal [5], [6]. Such U-Net however tends to produce false positive R peak around label position [5]. To localize a unique R peak in each QRS complex range, additional process steps are needed [7].

In this study, we proposed a CNN-based two-step method for R peak detection. Proposed method consists of QRS detection step and R peak regression step. In the QRS detection step, an U-Net structure is used to detect QRS complexes and their specific classes. In other words, the U-Net not only segments QRS complexes but also determines which types the detected QRS complexes are among normal sinus rhythm (NSR), ventricular premature contraction (VPC), and atrial premature contraction (APC), too. In addition to a vanila U-Net structure, we added residual blocks and squeeze-and-excitation layers to enhance training performance and emphasize relation between feature channels, respectively. After the QRS complex detection, the detected QRS complexes are resampled to a fixed length and each resampled QRS complex is input to R peak regression network. As metioned, to avoid multiple detection of R peak in one QRS detection, a regression network predicts a R peak location in the sense of normalized location in each QRS complex. The prediction result can be converted to physical sample location and it guarantees a unique R peak detection result in each QRS complex. The proposed model was evaluated on the dataset which consists of ECG signals from resting ECG and wearable ECG device. Without additional process, the proposed model successfully predicts a unique R peak location in every QRS complex.

[1]HUINNO Co., Ltd., Seoul, The Republic of Korea jaeseongjang@huinno.com
* Corresponding author: Sunghoon Jung shjung@huinno.com

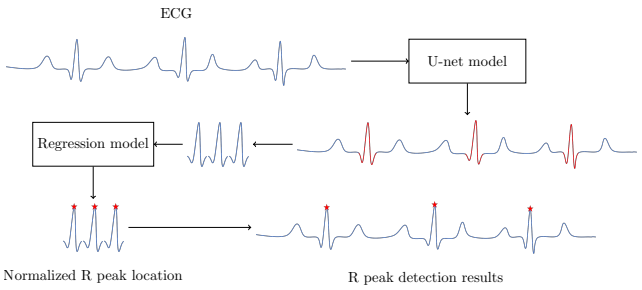

Fig. 1. Description of the detection method. From an ECG signal segment, the U-net model detects QRS compelx regions (red curves). The detected rgeions are extracted and resampled to a fixed length, and R peak is detected in each resampled QRS complex region by the regression model. The detected R peak in each QRS complex is converted to R peak location in the ECG signal segment.

II. METHODS

The basic structure of the proposed method is as follows:

1) QRS detection U-Net $\mathbf{F}_{\text{seg}}$
2) R peak regression CNN $\mathbf{F}_{\text{reg}}$.

For an input ECG signal, the U-Net $\mathbf{F}_{\text{seg}}$ predicts pixel-wise classes of QRS complex classes. Based on the produced segmentation, we detect QRS_{on} and QRS_{off} where the predicted class changes background into a QRS comlex class and a QRS complex into background, respectively. During QRS_{on} and QRS_{off} detection, QRS complexes shorter than 3 samples are rejected. From the detected pairs of QRS_{on} and QRS_{off}, we extract ECG signal in $[\text{QRS}_{\text{on}}, \text{QRS}_{\text{off}}]$, rescale the amplitude of the signal based on the minimum and maximum of the interval, and temporally resample so that all extracted signal has the same voltage range $[0, 1]$ and the fixed length L_{reg}. The resampled signal is then inserted to the regression model $\mathbf{F}_{\text{reg}}$ and the normalized R peak location is predicted as a value in $[0, 1]$. The output for each QRS complex is linearly converted to physical location between QRS_{on} and QRS_{off} so that 0 and 1 was converted to QRS_{on} and QRS_{off}, respectively. The detailed deep learning model structures are described below.

A. U-Net for Semantic Segmentation of QRS Complexes

Given an input signal $\mathbf{I} \in \mathbb{R}^L$, our U-Net model $\mathbf{F}_{\text{seg}}$ aims to produce a semantic segmentation map $\mathbf{S} \in \mathbb{R}^{L \times C}$ where L and C are the signal length and the number of classes, respectively. The proposed U-Net model consists of successive encoding blocks and successive decoding blocks. Each encoding block consists of one CBR (Convolution, Batch normalization, ReLU activation) block $\mathbf{F}_{\text{CBR}} : \mathbb{R}^{L \times C} \rightarrow \mathbb{R}^{L \times C'}$ and residual blocks $\mathbf{F}_{\text{Res}} : \mathbb{R}^{L \times C'} \rightarrow \mathbb{R}^{L \times C'}$. In each residual block, two successive CBR blocks $\mathbf{F}_{\text{CBR1}} : \mathbb{R}^{L \times C'} \rightarrow \mathbb{R}^{L \times C'}$, $\mathbf{F}_{\text{CBR2}} : \mathbb{R}^{L \times C'} \rightarrow \mathbb{R}^{L \times C'}$ are applied to input feature I and the feature $F \in \mathbb{R}^{L \times C'}$ is obtained. To enhance the feature, the global average pooling spatially squeezes the obtained feature

$$F_{\text{GAP}} = \frac{1}{L} \sum_{l=1}^{L} F(l, c)$$

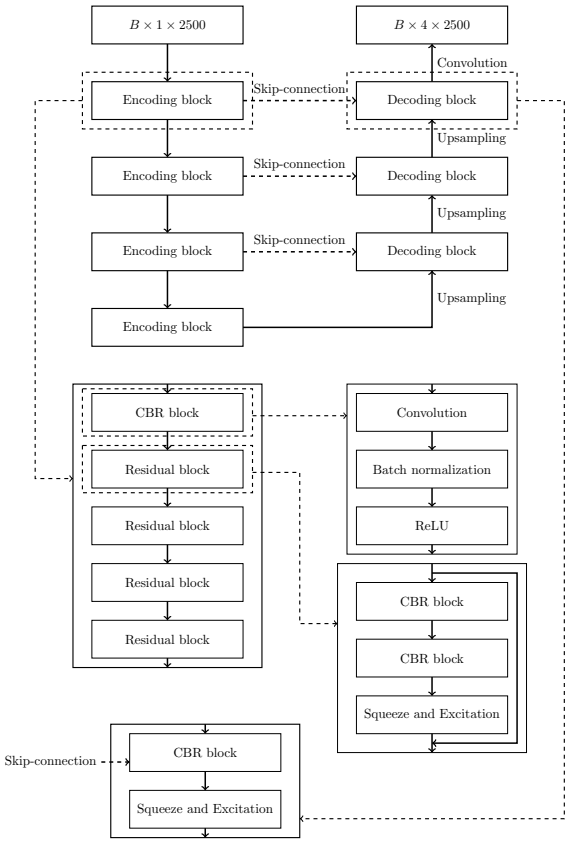

Fig. 2. Brief structure of the U-Net for QRS complex detection and detailed structures of each block. The CBR blocks in encoding blocks, residual blocks, and decoding blocks have the same structures. Since the proposed model predicts class for each sample, the output has a length of 2500 where B stands for batch size.

and the squeezed feature is compressed and expanded to obtain channel excitation z by two successive fully-connected layer, $\mathbf{FC}_1 : \mathbb{R}^{l \times C'} \rightarrow \mathbb{R}^{l \times C'/r}$ and $\mathbf{FC}_2 : \mathbb{R}^{l \times C'/r} \rightarrow \mathbb{R}^{l \times C'}$

$$z = \sigma(\mathbf{FC}_2(\delta(\mathbf{FC}_1(F_{\text{GAP}})))) \in \mathbb{R}^{1 \times C'}$$

where r is a parameter for the bottleneck in channel excitation, and σ and δ is the sigmoid and ReLU activation respectively. The channel excitation is applied to the feature F_{in} and residual operation outputs the feature F_{out} of the encoding block as follows:

$$F_{\text{out}} = I + z \star F_{\text{in}}$$

where $z \star F$ is channel-wise multiplication and $F_{\text{in}} \in \mathbb{R}^{L \times C'}$ is the input feature of the encoding block. The feature F_{out} from each encoding block is skip-connected to a corresponding decoding block, and the decoding block concatenates upsampled features from the previous block and the skip-connected features. The concatenated features in the decoding block are passed through a convolution layer, batch normalization, and the ReLU activation, then finally spatial and channel excitation are applied at the end of each decoding. After applying all the decoding blocks, a few convolution layers are followed to predict a class for each pixel.

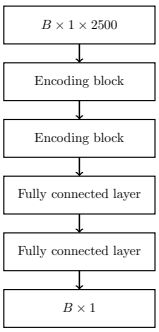

Fig. 3. Brief network structure for regression of normalized R peak location. A couple of encoding blocks are followed by two fully connected layer. B represents batch size. The structure of the encoding blocks is same as the one of the encoding block of U-Net in Fig. 2.

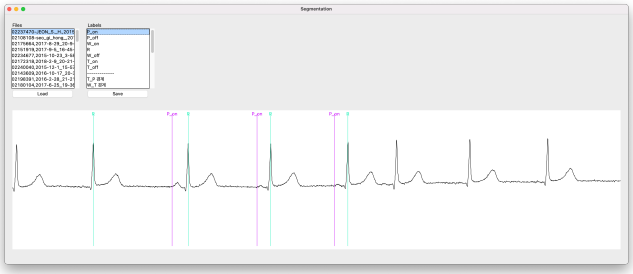

Fig. 4. A screenshot of the annotation GUI with an example ECG data loaded. For each heartbeat, P-wave, T-wave, QRS complex, and R peak are annotated with the GUI and save them to a labelled dataset.

B. R Peak Regression

As mentioned above, detected QRS complex is rescaled and resampled to a fixed length L_{reg}. The regression network $\mathbf{F}_{\text{reg}}$ predicts *normalized* location of R peak R_{norm} in the QRS complex:

$$\mathbf{F}_{\text{reg}} : \mathbb{R}^{L_{\text{qrs}}} \rightarrow [0, 1].$$

The normalized output R_{norm} is converted into a sample number R_{samp}:

$$R_{\text{samp}} = R_{\text{norm}} \times (\text{QRS}_{\text{off}} - \text{QRS}_{\text{on}}) + \text{QRS}_{\text{on}}$$

where QRS_{on} and QRS_{off} are the start and end sample numbers of the input QRS complex respectively.

For encoding process of the regression model, the same structure of encoding block is used as the U-Net. A few successive encoding blocks are followed by fully-connected (FC) layers for regression. Except for the last FC layer, ReLU function is used for activation. For the last FC layer which predicts regression result, sigmoid function is used to obtain normalize output. The detailed structure is decribed in Fig. 3.

III. EXPERIMENTAL SETUP

A. Dataset

For experiments, we combined QT Database (QT DB) [8], Lobachevsky University Electrocardiography Database (LU DB) [9], the in-house resting ECG data, and 1-channel ECG data acquired by a wearable device, MEMO Patch™(https://www.huinno.com). QT DB provides 105 2-lead ECG recordings of 15 minutes with a sampling frequency of 250 Hz, whose onset, peak, and end of P, T waves, and QRS complexes are marked. For each patient, if a MLII record is available for the patient, we extracted one MLII ECG record of the length of 10 seconds since overfitting to a patient-specific waveform morphology might be caused if we used all available 10-second ECG records from a patient. For data whose manual annotation is available, we selected the interval which is overlapped with the manual annotation interval with the manual annotation. For data whose only automatic annotation is available, automatic annotation was imported. LU DB contains 200 12-lead ECG data with a sampling frequency of 250 Hz whose P, T wave and QRS complex boundaries are annotated by cardiologists. After data collection from the public databases, visual inspection was performed and annotations were corrected using a self-developed GUI based on Python 4. In contrast to the public databases whose pre-annotations were provided, the in-house data was annotated by clinical experts. In addition to the public databases, the in-house resting ECG data was collected. The data has the length of 10 seconds and were digitized at 500 Hz. Without any cropping along temporal samples, the data was resampled with sampling rate with 250 Hz. Combining the three dataset, QT DB, LU DB, and the in-house data, we divded the combined data into the ratio of 8:2 for train-validation and test data. The train-validation data was divided again into train and validation set with the ratio of 8:2. QRS complex regions were annotated with 3 classes (NSR, VPC, APC) and a unique R peak was annotated in each QRS complex. For application to wearable ECG systems, data from MEMO Patch was used for an additional test. MEMO Patch provides 1-channel ECG records (up to 14 days) with 24 dB resolution and 250 Hz sampling frequency. Similar to the in-house data, the MEMO Patch data was annotated by clinical experts. Table I shows the number of 10-second ECG data for each class.

TABLE I
THE NUMBER OF 10-SEC ECG RECORDINGS IN DATASETS.

	NSR	VPC	APC
In-house Data	249	465	555
MEMO Patch Data	58	40	68

B. Model Architecture and Learning

The U-Net structure consists of 4 encoding blocks and 3 decoding blocks. At the end of the decoding blocks, 2 fully-connected layers are followed to produce semantic segmentation map. Each encoding block consists 1 CBR and 4 residual blocks. We increase the number of channel of the 4 encoding blocks by multiples of 16. Specifically, as the 4 encoding blocks get deeper, 16, 32, 48, and 64 channels of features are applied to the encoding blocks. For every convolution layer of encoding blocks, weight size is set to 7 with padding size of 3 then ReLU activation

is followed. To decrease feature size, the first CBR block of each encoding block convolves feature with stride of 2. For each decoding block, a CBR block is applied to combine skip-connected features and upsampled feature from the previous decoding block. For training, the cross-entropy loss with softmax output was used and weight decay was set with $\lambda = 0.0001$. The inital learning rate was set to 0.01 and learning rate decreased by the ratio 0.8 if no further validation loss improved. Minibatch size was 16 and Adam method was used for optimzation.

For the regression model, we resampled each QRS complex as the legnth of 32. Weight size of convolution layers and the number of channel were set to the same as one of the U-Net. We use however just 2 encoding blocks, and 2 FC layers are followed for regression. For training, the mean squared error loss was used and weight decay was set with $\lambda = 0.0001$. The inital learning rate was set to 0.0001 and learning rate decreased by the ratio 0.8 if no further validation loss improved. All weights were selected when validation loss converged. All theses experiments were performed on an NVIDIA RTX 3090 GPU with 24 GB RAM.

IV. RESULTS

In this section, quantitative and qualitative results of our experiments are presented on the mentioned dataset. The trained model were investigated in the following two points of view:

1) Wheter the model detects labelled QRS complex region and correctly predicts its class
2) How precisely R peak location is detected

As described in the previous section, the trained model was tested on two types of the test set from the combined dataset and our wearable devices.

A. QRS Detection

For evalueation of QRS detection performance, F_1 score, precision, and recall were measured. F_1 score is the harmonic mean of precision and recall:

$$F_1 = \frac{2}{\text{Recall}^{-1} + \text{Precision}^{-1}}$$

where

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}}$$
$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}.$$

We present results for QRS detection on the two dataset from the combined data (test data) and measurable devices (MEMO Patch data) in Tab. II. Our model shows performance decrement 8.3 %p, 8.2 %p and 7.3 %p in F_1 score, precision, and recall, respectively in MEMO Patch data test.

TABLE II

IoU AND F_1 SCORE FOR THE TEST DATA SETS.

	F_1	Precision	Recall
Test data	94.01 %	94.28 %	93.74 %
MEMO Patch data	85.73 %	86.06 %	86.48 %

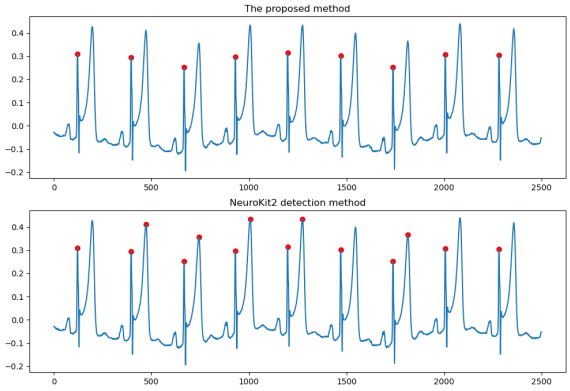

Fig. 5. Comparision results of R peak detection using the proposed model (Top) and NeuroKit2 (Bottom). The proposed model detects R peaks of while the method of NeuroKit2 detect false R peaks on T-waves. The horizontal and vertical axes of the both plots represent the number of sample with the sampling rate of 250 Hz and voltage in mV, respectively.

B. R peak detection

For evaluation of R peak detection performance, the root mean squared error (RMSE) was measured and the number of test samples whose error is greater than or equal to 5 sample intervals (5+ Error) was counted. Here, the our criteria of 5 sample intervals, 0.02 seconds in time, corresponds to the half of the 1 mm square in ECG paper. We present results for R peak detection for the test data and MEMO Patch data in Tab. III.

TABLE III

R PEAK DETECTION RESULTS FOR THE TEST DATA SETS. IN THE THIRD COLUMN, THE RATIO OF THE NUMBER OF 5+ ERROR TO THE TOTAL NUMBER OF BEATS IS EVALUATED

	RMSE	5+ Error
Test data	0.56	0.36 % (12/3308)
MEMO Patch data	0.52	0.33 % (7/2097)

In the value of RMSE, R peak is detected within error than 1 sample interval. Moreover, on the both test data sets, there is no noticeable decrement of R peak detection performance.

In Fig. 5, we plotted the results of a selective example of MEMO Patch data that the proposed model shows more robust results for general ECG waveforms compared to NeuroKit2 algorithm [10]. In the example, the result acquried by NeuroKit2 algorithm detects R peaks around the offsets of T waves due to the sharp waveform which is patient-specific. In contrast to the NeuroKit2 results, our method

stablly captures QRS complex range and detects the R peak in each QRS complex.

V. CONCLUSION AND DISCUSSION

In this paper, we proposed a CNN-based two-step method for R peak detection. Replacing R peak detection by regression process guarantees a unique R peak detection in each QRS complex. If the uniquess and existence of a target are guaranteed in an input data, regression could be a more convenient method since its structure forces a unique estimation result. Since the R peak regression is based on the CNN-based QRS region detection, the model shows more robust and stable R peak detection results in spite of low SNR and patient-dependent QRS waveform. In diagnosis, such stable detection methods enable precise heart rate estimation for long-time ECG data. Moreover, in QRS complex segmentation step, we extract class of QRS complexes as well as their location and duration. Such precise class information alleviate laborious works of clinical staffs to inspect long-time ECG data.

Our work has room for imporvement in many ways. Firstly, in addition to APC and VPC, more various abnormal beat classes can be added. In this work, APC and VPC are considered, but there still remain clinically meaningful abnormal beat classes. Moreover, P wave and T wave are additional waveform classes as well as QRS complexes [6]. Based on the QRS complex detection, detection of the waves expands clinical information, such as ST segment analysis. For the regression process, resampling of detected QRS complexes was required for a fixed input length. By using adaptive pooling layers which produce outputs of a fixed size, we can get rid of the resampling step or integrate the regression process into the U-net model so that the model produces R peak detection result as well as semantic segmentation results.

ACKNOWLEDGEMENTS

This work was supported by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project Number: 1711138361, KMDF_PR_20200901_0174 and 1711139106, KMDF_PR_20210527_0004)

REFERENCES

[1] E. B. Bass, E. I. Curtiss, V. C. Arena, B. H. Hanusa, A. Cecchetti, M. Karpf, and W. N. Kapoor, "The duration of holter monitoring in patients with syncope: is 24 hours enough?" *Archives of Internal Medicine*, vol. 150, no. 5, pp. 1073–1078, 1990.

[2] P. M. Barrett, R. Komatireddy, S. Haaser, S. Topol, J. Sheard, J. Encinas, A. J. Fought, and E. J. Topol, "Comparison of 24-hour holter monitoring with 14-day novel adhesive patch electrocardiographic monitoring," *The American journal of medicine*, vol. 127, no. 1, pp. 95–e11, 2014.

[3] M. Elgendi, B. Eskofier, S. Dokos, and D. Abbott, "Revisiting qrs detection methodologies for portable, wearable, battery-operated, and wireless ecg systems," *PloS one*, vol. 9, no. 1, p. e84018, 2014.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[5] G. Jimenez-Perez, A. Alcaine, and O. Camara, "U-net architecture for the automatic detection and delineation of the electrocardiogram," in *2019 Computing in Cardiology (CinC)*. IEEE, 2019, pp. Page–1.

[6] A. Peimankar and S. Puthusserypady, "Dens-ecg: A deep learning approach for ecg signal delineation," *Expert Systems with Applications*, vol. 165, p. 113911, 2021.

[7] M. Šarlija, F. Jurišić, and S. Popović, "A convolutional neural network based approach to qrs detection," in *Proceedings of the 10th international symposium on image and signal processing and analysis*. IEEE, 2017, pp. 121–125.

[8] P. Laguna, R. G. Mark, A. Goldberg, and G. B. Moody, "A database for evaluation of algorithms for measurement of qt and other waveform intervals in the ecg," in *Computers in cardiology 1997*. IEEE, 1997, pp. 673–676.

[9] A. Kalyakulina, I. Yusipov, V. Moskalenko, A. Nikolskiy, K. Kosonogov, N. Zolotykh, and M. Ivanchenko, "Lobachevsky university electrocardiography database."

[10] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. H. A. Chen, "NeuroKit2: A python toolbox for neurophysiological signal processing," *Behavior Research Methods*, vol. 53, no. 4, pp. 1689–1696, feb 2021. [Online]. Available: https://doi.org/10.3758%2Fs13428-020-01516-y