Unsupervised non-parametric change point detection in electrocardiography

Nikolay Shvetsov  
Nikolay.Shvetsov@skoltech.ru  
Skolkovo Institute of Science and Technology  
Moscow, Russia

Nazar Buzun  
n.buzun@skoltech.ru  
Skolkovo Institute of Science and Technology  
Moscow, Russia

Dmitry V. Dylov  
d.dylov@skoltech.ru  
Skolkovo Institute of Science and Technology  
Moscow, Russia

ABSTRACT

We propose a new unsupervised and non-parametric method to detect change points in electrocardiography. The detection relies on optimal transport theory combined with topological analysis and the bootstrap procedure. The algorithm is designed to detect changes in virtually any harmonic or a partially harmonic signal and is verified on ECG data streams. We successfully find abnormal or irregular cardiac cycles in the waveforms for the six of the most frequent types of clinical arrhythmias using a single algorithm. Our unsupervised approach reaches the level of performance of the supervised state-of-the-art techniques. We provide conceptual justification for the efficiency of the method.

CCS CONCEPTS
• Computing methodologies → Anomaly detection; • Theory of computation → Gaussian processes; • Unsupervised learning and clustering; • Models of learning.

KEYWORDS
Data streams, Anomaly detection, Arrhythmia detection, Periodic and quasi-periodic signals, Optimal transport, Topological data analysis, Wasserstein distance, Bootstrap, Unsupervised learning.

1 INTRODUCTION

Cardiovascular diseases are the top cause of mortality and the major life threat in adults worldwide [21]. Electrocardiography (ECG) has become the most frequently used clinical modality, attracting a multidisciplinary effort to detect conceivable markers of the heart problems in the recordings of the electrical function of the heart. Typical single-channel ECG signal is a one-dimensional time series measurement, close to a periodic one, with each period consisting of three main parts: P wave, QRS complex, and T wave.

Among many things, the ECG modality allows detecting disruptions in the cardiac rhythm, being the major proxy for the doctors to diagnose heart arrhythmias [11].

Mathematically, cardiac arrhythmias correspond to some degree of broken periodicity in the ECG data stream [21]. The six of the most frequent types of clinical arrhythmias are atrial flutter, atrial fibrillation, supraventricular tachycardia, premature atrial contraction, and ventricular rhythms [13]. Each of these conditions is defined by a different set of morphologic and temporal characteristics in the PQRST complex in the ECG signals.

In this work, we were motivated to develop a single, model-agnostic, unsupervised algorithm to detect all of such arrhythmias in a binary classification scenario, focusing on high detection sensitivity and specificity. Thanks to the non-parametric construction of the proposed change point statistic, we ensure zero modeling bias and applicability to a wide range of quasi-periodic signals.

1.1 Problem statement

The formal problem statement is the following. Let $X_t$ be the quasi-periodic signal with a period $T$. One has to test the hypotheses

$$
\mathcal{H}_0 : \{ X_t \sim P_{f_0(t) / T}, \forall t \in [0, n] \},
\mathcal{H}_1 : \{ \exists \tau^* : X_t \sim P_{f_{\tau^*}(t) / T} : X_t \sim P_{f(t)} \}.
$$

In the notation above, $P$ represents a probability distribution, $n$ is the dataset size (duration of observation), $\tau^*$ is the change point time, $f_0(t) / T$ and $f_{\tau^*}(t)$ are the functions parametrizing the corresponding distributions for $t \in [0, \tau^*]$ and $t \in [\tau^*, n]$ respectively. Note that the time in the latter is not divided by $T$, implying the occurrence of the anomaly in the preceding rhythm.

![Figure 1: Pipeline of the proposed algorithm. Here, $\tau$ is the center of the second sliding window, $2h$ is the size of the second sliding window, $W_p^P$ – Wasserstein distance, $\{\mu_{\tau}^P(t), \mu_{\tau + h}^P(t)\}$ are the bootstrap measures in the left and the right parts of the second sliding window.](image-url)
2 METHODOLOGY

The major difficulty in the statistical study of the problem (1) is twofold: the dependent data and the lack of a suitable parametric model for an intricate signal, such as ECG. To address these challenges, we propose a new pipeline shown in Figure 1. In the proposed algorithm, one can leverage the optimal transport (OT) approach, allowing to build a non-parametric change point statistic to test the hypotheses (1).

In what follows, we propose to apply the TDA/OT approach not to the original signal, but to a projection of the quasi-periodic function into a closed curves space (the point cloud), allowing both the periodic and the morphologic components of the original signal’s waveform to be considered. Eventually, we estimate quantiles of the change point statistic with the bootstrap procedure in order to set a threshold under the null hypotheses assumption.

2.1 The first window: calculate point clouds

The first step of our approach is to map the original time series into the point cloud. We use the method based on the sliding windows with 1-dimensional persistence scoring described in [23]. The main idea is to present the original periodical signal as a closed curve, which will help us to apply the optimal transport formalism to the quasi-periodic data. Let’s define the sliding window

\[ SW(t) = \{ X_t, X_{t+1}, \ldots, X_{t+M} \}, \]

which makes an embedding of the signal \( X_t \) at point \( t \) into \( \mathbb{R}^{M+1} \). Iterating through different values of \( t \) with a step \( \Delta t \), one gets a collection of points called sliding-window point cloud of \( X_t \). A critical parameter for this embedding is the first window size \( M \).

A natural choice is to make it equal to the duration of a single period of the signal (e.g., one PQRST cycle in the normal beat pattern). After calculating the array of points (2), we apply Principal Component Analysis (PCA) for the obtained cloud in order to increase robustness, reduce effect of experimental fluctuations in the signal (i.e., measurement noise), and to have a possibility to visualize the rhythm disturbance. Example of the point clouds are shown in Figure 3 for the case of data streams arriving from an ECG monitor. Notably, for all types of physiological signals considered herein, we found that the rhythm disturbance could be easily distinguished by a naked eye regardless of whether two or three main PCA components were plotted.

2.2 The second window: Wasserstein distances

In order to find structural changes in the point cloud corresponding to the structural changes in the original time series, we elaborate the method described in [4]. The main idea is that at each time step, the procedure extracts a data slice from the point cloud, splits it in two equal-size parts, and computes Wasserstein distance between them. The size of the sliding window could be equal to several curve loops in the point cloud. The method avoids the rise of values of the Wasserstein distances due to fluctuations and neglects noise-driven changes in the curves, effectively tracing only the meaningful structural changes in the signal.

Wasserstein distance is defined on probability distribution pairs on some metric space [28]. By definition, the Wasserstein distance of degree \( p \) between the probability measures \( \mu \) and \( \nu \) is

\[ W^p_p(\mu, \nu) = \left( \inf_{\gamma \in \Pi(\mu, \nu)} \int \| x - y \|^p d\gamma(x, y) \right)^{1/p}. \]

Let’s introduce a change point statistic (the maximum distance over the window positions):

\[ T(2h) = \max_{\tau} W^p_p(\mu(\tau), \mu(\tau)), \]

where \( \delta_{X_t} \) is the Dirac function at position \( X_t \) (i.e., a unit mass concentrated at location \( X_t \)), \( \tau \) is the central point of the sliding window of length \( 2h \), implying that the data series within the sliding window is \( [X_{\tau-h}, \ldots, X_{\tau+h}] \).

We calculate the Wasserstein distance for each position of the second sliding window and create a new time series to be used for showing how the curves differ inside of the window. In practice, one can calculate Wasserstein distances via the Sinkhorn algorithm using the Optimal Transport Library [4, 9].

![Figure 2: Values of Wasserstein distances (top) help detect abnormal rhythm in the original time series (bottom) in an unsupervised manner. The inset shows original distribution \( W^p_p \) (blue) and its approximation \( W^p_p \) (orange).](attachment:figure2.png)

2.3 Moving blocks bootstrap for rhythm analysis and arrhythmia detection

In this step, we compute Wasserstein distance and execute the bootstrap procedure:

\[ T^b(2h) = \max_{\tau} \inf_{\gamma \in \Pi(\mu(\tau), \mu(\tau)))} \int \| x - y \|^p d\gamma(x, y) \]

where the set \( k(t) \) is generated by the Moving Block Bootstrap (MBB), and where the data is split and shuffled into \( n \) blocks randomly. Naturally, we assume that the points located in the peaks...
of the plot of the Wasserstein distances correspond to the arrhythmia points on the original periodic signal. MBB was formulated in separate works by Künsch [18] and Lahiri [19] as a new scheme to create pseudo-samples. The usual bootstrap forms new samples, taking only random observations from the initial sample, whereas, the MBB performs this procedure only within a row of the formed blocks. We use a weighted block structure of the MBB, which generates random weights for each block and, importantly, preserves the structure of the original time series.

After the MBB resampling, we create a list of change point statistic values $T^h(2h)$ and set the threshold with a confidence level corresponding to the border between the normal points and the points of arrhythmia (see Figure 2). It is assumed that quantiles of $T^h(2h)$ are close to the quantiles of $T(2h)$ (the property called bootstrap consistency [27]).

## 3 EXPERIMENTS

In this section, we will demonstrate the use of the pipeline proposed in Figure 1 on real physiological data streams. The real ECG signals are at the main focus of this paper. But we also experimented with other physiological signals (wearable accelerometer, worn on a limb of a patient with Parkinson’s disease, simplified model of pathological brain activity [17]), and tested artificial ECG signals with common synthetically imposed arrhythmias. These additional experiments are described in [27].

### 3.1 Data

We used the MIT-BIH arrhythmia dataset from the PhysioNet [22] as the main source of recordings with expert annotations (so that we could compare our unsupervised detection pipeline to the ground truth). The dataset contains 48 half-hour excerpts of two-channel ambulatory ECG waveforms.

To compare against the benchmarks, 23 of such recordings were chosen at random from a set of 4000 24-hour ambulatory ECG recordings to include most common arrhythmia types, replicating the data selection of the other works. The remaining 25 recordings, considered herein, include less common (but still clinically significant) arrhythmias which we described in the introduction. Each of these recordings contains two 30-min ECG lead signals (mostly MLII lead and lead V1/V2/V4/V5 [11]) sampled at the frequency of 360 Hz. Our algorithm proved to work without any data preprocessing, filtering, or noise reduction, and detected all types of existing arrhythmias without supervision.

### 3.2 Comparison with state-of-the-art

#### 3.2.1 Data preparation

Each ECG series was split to parts of different size (40,000, 80,000, and 120,000 points). If we take the indexes of the points, where the values are above the separation line calculated in the bootstrap procedure, these indexes/points in the original ECG will be the points with the arrhythmia. The PhysioNet datasets [22] have the annotations accompanying the data; therefore, it is possible to compare the predicted labels of the points with the ground truth. To calculate the performance metrics defined below, we used a hold-out test set comprising the ECG signals with the normal rhythms (160 parts) and the ECG with various confirmed arrhythmias (192 parts).

The parameters of the first sliding window have the following values: $M s = 450$, $s = 1$, $Δ t = 2$ ($Δ t$ is the step of the moving window), corresponding to the typical ECG sampling parameters, such as those in the MIT-BIH dataset [22]. The size of the second sliding window is equal to 4 curve loops, meaning that the window separates the series into 2 parts with 2 curve loops in each part. We also chose the confidence level to be equal to $α = 5%$.

#### 3.2.2 Results

Complete caparison of our method with the SOTA algorithms is shown in Table 1. We tested the method on the short-episode arrhythmia in the long-term clinical monitoring data stream (see Figure 3). On the hold-out set, our method yields the specificity of 89.8% ± 6%, and the sensitivity of 94% ± 4.0%. Similarly to other machine learning problems, the optimal choice of prediction threshold and the sizes of the sliding windows can define the trade-off between the high recall and the low false positive rate.

![Figure 3: Algorithm’s performance on a long-term monitoring data of the clinical ECG data stream. The detected arrhythmia is visible in the Wasserstein graph and in the 3D point cloud. The ground-truth annotations are provided by an expert. Notice the trajectory of the red cloud in an orthogonal plane, which emphasizes the onset of arrhythmia.](image)

## 4 DISCUSSION AND CONCLUSIONS

We presented a new unsupervised and non-parametric learning algorithm for detection of rhythm anomalies in the raw data of quasi-periodic signals, validating it on ECG data-streams. The detection relies on optimal transport theory combined with topological analysis and the bootstrap procedure. Mathematically, it is not obvious that such a bootstrap converges, effectively requiring one to prove the convergence of the bootstrap procedure theoretically [27]. The resulting method is the first to apply the bootstrap procedure to a popular measure of similarity (Wasserstein distance) for periodic signals, with the simple pipeline providing a robust statistical approach to predict abnormal rhythms in an unsupervised manner and with high computational efficiency.
Table 1: Comparison of the proposed approach (shown in bold) with the state-of-the-art.

| Method                                | Sensitivity, % | Specificity, % | AUC, % | Accuracy, % | Supervision |
|----------------------------------------|----------------|----------------|--------|-------------|-------------|
| Bootstrap on ECG real data             | 94.0 ± 4.0     | 89.8 ± 6.0     | 96.0   | 94.2        | □           |
| Ruptures (PELT) on Wasserstein distance data [16, 29] | 91.6            | 77.0            | 85.3   | 85.2        | □           |
| Ruptures (PELT) on Euclidean distance data [16, 29] | 88.9            | 84.1            | 86.5   | 85.4        | □           |
| BOCP on Wasserstein distance data [1]   | 92.0            | 80.6            | 86.8   | 87.3        | □           |
| BOCP on Euclidean distance data [1]     | 83.8            | 88.9            | 89.4   | 89.0        | □           |
| SVM+PCA [12]                           | 70.0            | 98.0            | 88.0   | 90.0        | △           |
| 2D CNN [14]                            | 99.6            | 97.8            | 98.6   | 98.9        | □           |
| Echo State Network [2]                 | 84.4            | 99.7            | 90.0   | 97.7        | □           |
| LD QRs-time interval-based feature [24] | 75.9            | 77.7            | 77.0   | -           | -           |
| Noninvasive Scoring Method + Logistic Regression [15] | 97.0            | 63.0            | 81.0   | -           | -           |
| DT+Heart rate features [10]            | 98.1            | 85.0            | 91.0   | -           | -           |

* Supervision: □ Unsupervised, △ Semi-supervised, ○ Supervised

Remarkably, this completely unsupervised and non-parametric method reached the performance level of the state-of-the-art supervised benchmarks in the centuries-old problem of arrhythmia detection in electrocardiography (ECG). One may naturally expect it to work even better if it is combined with the deep learning methods [11], especially if set up in a recurrent neural network configuration [8, 20] or with pre-trained models [26]. Another line of the future work can entail the extension of the algorithm for the multi-task classification [25] also using the unsupervised bootstrap method on the Wasserstein distances. Overall, the suggested pipeline holds potential to become a new accepted way of setting arrhythmia alarms in various monitoring modalities.

We acknowledge support from IoT Center of Excellence (NTI) and Russian Foundation for Basic Research (grant #19-29-01240).

REFERENCES

[1] Ryan Prescott Adams and David J. C. MacKay. 2007. Bayesian Online Changepoint Detection. arXiv:0710.3742 (2007).

[2] Miquel Alfaro, Miguel C. Soriano, and Silvia Ortin. 2019. A Fast Machine Learning Model for ECG-Based Heartbeat Classification and Arrhythmia Detection. Frontiers in Physics 7 (2019), 103. https://doi.org/10.3389/fphy.2019.00163

[3] Nazar Buzun. 2019. Gaussian approximation for empirical barycenters. arXiv:1904.08891 (2019).

[4] Nazar Buzun and Valeriy Avanesov. 2017. Bootstrap for change point detection. arXiv:1710.07285 (2017).

[5] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. 2013. Comparison and anti-concentration bounds for maxima of Gaussian random vectors. arXiv:1301.4807 (2013).

[6] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. 2014. Anti-concentration and honest, adaptive confidence bands. The Annals of Statistics 42, 5 (Oct 2014), 1787–1818. https://doi.org/10.1214/14-aos1235

[7] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. 2014. Gaussian approximation of suprema of empirical processes. The Annals of Statistics 42, 4 (Aug 2014), 1564–1597. https://doi.org/10.1214/14-aos1230

[8] Aritra Chowdhury, Dmitry V Dylov, Qing Li, Michael MacDonald, Dan E Meyer, Michael Marinz, and Alberto Santamaria-Pang. 2017. Blood vessel characterization using virtual 3D models and convolutional neural networks in fluorescence microscopy. In IEEE ISBI 2017. IEEE, 629–632.

[9] Marco Cuturi. 2013. Sinkhorn distances: Lightspeed computation of optimal transport. In Advances in neural information processing systems. 2292–2300.

[10] J Faganeli and F Jager. 2010. Automatic classification of transient ischaemic and transient non-ischaemic heart-rate related ST segment deviation episodes in ambulatory ECG records. Physiological Measurement 31, 3 (Feb 2010), 323–337. https://doi.org/10.1088/0967-3334/31/3/004

[11] Giulia Guidi and Manas Karandikar. 2014. Classification of Arrhythmia using ECG data. https://doi.org/10.20447/2014

[12] Jing Hu, Hua Zhang, Zhichong Lin, Yiu Xu, and Fumin Guo. 2018. Direct arrhythmia classification from compressive ECG signals in wearable health monitoring system. Journal of Circuits, Systems and Computers 27, 6 (2018), 1850088.

[13] Jane Huff. 2006. ECG workout: Exercises in arrhythmia interpretation. Lippincott Williams & Wilkins.

[14] Tae Joon Jun, Hoang Minh Nguyen, Daeyeon Kang, Dobyrun Kim, Daeyeong Kim, and Young-Hak Kim. 2018. ECG arrhythmia classification using a 2-D convolutional neural network. arXiv:1804.06812 (2018).

[15] Hiroshi Kawazoe, Yukiko Nakano, Hidenori Ochi, Masahiko Takagi, Yuuseki Hayashi, Yuko Uchumura, Takehito Tokenyama, Yoshikazu Watanabe, Hiroya Matsumura, Shunsuke Tomomori, Akinori Sairaku, Karayoshi Suenari, Akinori Awazu, Youku Miwa, Kyoko Soejima, Kazuaki Chayama, and Yasuki Kihara. 2016. Risk stratification of ventricular fibrillation in Brugada syndrome using noninvasive scoring methods. Heart Rhythm 13, 10 (2016), 1947 – 1954. https://doi.org/10.1016/j.hrthm.2016.07.009 Focus Issue: Sudden Death.

[16] Rebecca Killick, Paul Fearnhead, and Idris A Eckley. 2012. Optimal detection of changepoints with a linear computational cost. J. Amer. Statist. Assoc. 107, 500 (2012), 1590–1598.

[17] Dmitriy Keylov, Dmitry V Dylov, and Michael Rosenshan. 2020. Reinforcement learning for suppression of collective activity in oscillatory ensembles. Chaos 30, 3 (2020). https://doi.org/10.1063/1.5128909 arXiv:1909.12154

[18] Hans R Kunsch. 1989. The jackknife and the bootstrap for general stationary observations. The annals of Statistics (1989), 1217–1241.

[19] Soumendra Nath Lahiri. 2013. Resampling methods for dependent data. Springer Science & Business Media.

[20] Vadim Liventsev, Irina Fedulova, and Dmitry Dylov. 2019. Deep Text Prior: Weakly Supervised Learning for Assertion Classification. In International Conference on Artificial Neural Networks. Springer, 243–257.

[21] Nicos Maglaveras, Telemachos Stamkopoulou, Konstantinos Diamantaras, Costas Pappas, and Michael Stratou. 1998. ECG pattern recognition and classification using non-linear transformations and neural networks: A review. International Journal of Medical Informatics 52, 1 (1998), 191 – 208. https://doi.org/10.1016/S1386-5056(99)00138-5

[22] George B Moody and Roger G Mark. 2001. The impact of the MIT-BIH arrhythmia database. IEEE Engineering in Medicine and Biology Magazine 20, 1 (2001), 45–50.

[23] Jose Pereira and John Harer. 2013. Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis. arXiv:1307.6188 (2013).

[24] Philip de Chazal, M. O'Dwyer, and R. B. Reilly. 2004. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. IEEE Transactions on Biomedical Engineering 51, 7 (July 2004), 1196–1206. https://doi.org/10.1109/TBME.2004.827559

[25] Ivan Rodin, Irina Fedulova, Artem Shelmanov, and Dmitry V Dylov. 2019. Multi-task and Multimodal Neural Network Model for Interpretable Analysis of X-ray Images. In 2019 IEEE BIBM. IEEE, 1601–1604.

[26] Artem Shelmanov, Vadim Liventsev, Danil Kireev, Nikita Khromov, Alexander Panchenko, Irina Fedulova, and Dmitry V Dylov. 2019. Active Learning with Deep Pre-trained Models for Sequence Tagging of Clinical and Biomedical Texts. In 2019 IEEE BIBM. IEEE, 482–489.

[27] Nikolay Shvetsov, Nazar Buzun, and Dmitry V. Dylov. 2020. Unsupervised non-parametric change point detection in quasi-periodic signals arXiv:2002.02717 [cs.LG]

[28] Max Sommerfeld and Axel Munk. 2016. Inference for Empirical Wasserstein Distances on Finite Spaces. arXiv:1610.03827 (2016).

[29] Charles Truong, Laurent Oudre, and Nicolas Vayatis. 2018. Selective review of offline change point detection methods. Signal Processing 167 (2018).