bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.478609; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Electrocardiogram feature extraction and interval measurements using optimal representative cycles from persistent homology

Hunter Dlugas[1]

1 Department of Mathematics, Wayne State University, Detroit, MI, USA

¤Current Address: 1104 Faculty Administration Building, Wayne State University, Detroit, MI, USA
* fy7392@wayne.edu

Abstract

Cardiovascular diseases are among the leading causes of death, and their early detection and treatment is important for lowering their prevalence and mortality rate. Electrocardiograms (ECGs) record electrical activity of the heart to provide information used to diagnose and treat various cardiovascular diseases. Many approaches to computer-aided ECG analysis have been performed, including Fourier analysis, principal component analysis, analyzing morphological changes, and machine learning. Due to the high accuracy required of ECG-analysis software, there is no universally-agreed upon algorithm to identify P,Q,R,S, and T-waves and measure intervals of interest. Topological data analysis uses tools from algebraic topology to quantify hole-like shapes within data, and methods using persistence statistics and fractal dimension with machine learning have been applied to ECG signals in the context of detecting arrhythmias within recent years. To our knowledge, there does not exist a method of identifying P,Q,S, and T-waves and measuring intervals of interest which relies on topological features of the data, and we propose a novel topological method for performing these aspects of ECG analysis. Specifically, we establish criteria to identify cardinality-minimal and area-minimal 1-cycles with certain properties as P,Q,S, and T-waves. This yields a procedure for measuring the PR-interval, QT-interval, ST-segment, QRS-duration, P-wave duration, and T-wave duration in Lead II ECG data. We apply our procedure to 400 sets of simulated Lead II ECG signals and compare with the interval values set by the model. Additionally, the algorithm is used to identify cardinality-minimal and area-minimal 1-cycles as P,Q,S, and T-waves in two sets of 200 randomly sampled Lead II ECG signals of real patients with 11 common rhythms. Analysis of optimal 1-cycles identified as P,Q,S, and T-waves and comparison of interval measurements shows that 1-cycle reconstructions can provide useful information about the ECG signal and could hold utility in characterizing arrhythmias.

Author summary

Topological data analysis (TDA) has been a rapidly growing field within the past 15 years and has found applications across many fields. In the context of TDA, several algorithms primarily using persistence barcode statistics and machine learning have been applied to electrocardiogram (ECG) signals in recent years. We use a topological data-analytic method to identify subsets of an ECG signal which are representative of certain topological features in the ECG signal, and we propose that those subsets

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.478609; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

coincide with the P,Q,S, and T-waves in the ECG signal. We then use information about these subsets of the signal identified as P,Q,S, and T-waves to measure the PR-interval, QT-interval, ST-segment, QRS-duration, P-wave duration, and T-wave duration. We demonstrate our method on both simulated and real Lead II ECG data. These results show how identifying subsets of an ECG signal with certain topological properties could be used in analyzing the morphology of the signal over time and in arrhythmia-detection algorithms.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.478609; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Cardiovascular diseases are among the leading causes of death due to their high prevalence and mortality rate [1] [2] [3]. Electrocardiograms (ECGs) provide a non-invasive measure of the heart's electrical activity and are used in diagnosing and managing various cardiovascular diseases. Thus the analysis of ECGs is important for accurate diagnosis and proper treatment of cardiovascular diseases. Several approaches to automated ECG analysis have been performed, including machine learning [4] [5] [6] [7] [8] [9] [10] [11], wavelet transforms [12] [13] [14] [15] [16], and persistent homology [17] [18] [19] [20] [21] [22] [23]. Due to the high accuracy required of ECG-analysis software and the fact that the bulk of ECG analysis is carried out by healthcare providers, the development of algorithms that identifies P,Q,R,S, and T-waves, measures intervals of interest, and/or detects arrhythmias is an active area of research.

Topological data analysis (TDA) is concerned with the study of shapes constructed from a dataset which are invariant under continuous deformations such as stretching and twisting. Applications of TDA to ECG signals have used persistence statistics [20] [18], fractal dimension [18], and machine learning [17] [19]. Using cycle reconstructions has shown utility in various applications outside of ECG analysis such as analyzing structures on the atomic scale [24] and in structural engineering [25]. To our knowledge, nobody has performed an "inverse" analysis of using information from cycle reconstructions to analyze ECG signals. We focus our attention on the "rhythm lead", i.e. Lead II, and for the remainder of the paper, any reference to an ECG signal, whether simulated or real, refers to a Lead II ECG signal.

Suppose we are given an n-dimensional dataset, that is, a set of points in $\mathbb{R}^n$. There are n distinct types of topological features associated with that dataset: roughly, these "types" correspond to what one would intuitively regard as "holes" or "gaps" of differing dimensions in the dataset. To apply these methods to ECG data, we treat an ECG signal as two dimensional with one temporal dimension and one amplitude dimension rather than one dimensional with a specified sampling frequency. Consequently, from this perspective, there are two distinct types of topological features associated with ECG data: 0-dimensional homology features represent equivalence classes of connected components, while 1-dimensional homology features represent equivalence classes of non-contractible loops. The 0-dimensional homology features are useful in analyzing clustering phenomena and have recently shown utility as a metric of heart rate variability when applied to ECG signals [21] [22].

In this paper, our focus is instead on the 1-dimensional homology features, which we use in a novel way to analyze ECG data. Specifically, we identify subsets of an ECG signal as P,Q,S, and T-waves by considering these subsets to be representative cycles of 1-dimensional homology features of the signal with certain properties and then use these subsets identified as individual waves to measure the PR-interval, QT-interval, ST-segment, QRS-duration, P-wave duration, and T-wave duration. To illustrate the intuition behind 1-dimensional homology features and their representative cycles, we first present an example which demonstrates how the 1-dimensional homology features of a simple 2-dimensional dataset is related to the shape of the data in the "Example: a 1-dimensional topological feature in a simple dataset" section. This section describes some geometric intuition of 1-dimensional homology features without going into linear algebra details. These details are then given and the ideas from this section are formalized in the "Background on topological data analysis" section, where we define key terms in algebraic and computational topology to provide minimal background and establish terms that will be used in identifying features of ECG signals.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.478609; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Example: a 1-dimensional topological feature in a simple dataset

Consider the set of points in the Cartesian plane $\mathbb{R}^2$ shown in Fig 1A. Then consider a circle drawn around each datum, each with the same radius r. The geometric Čech complex of radius r is defined as the union of the interiors of these circles. It is a subset of $\mathbb{R}^2$, and as r increases, the geometric Čech complex of radius r is a larger and larger subset of $\mathbb{R}^2$. Fig 1B-E depicts the geometric Čech complexes of radius 0.5, $\frac{\sqrt{2}}{2}$, 2.5, and $\frac{5\sqrt{2}}{2}$ as the region shaded in blue. Notice that for $r < 0.5$, none of the circles overlap, whereas for $0.5 \leq r < \frac{\sqrt{2}}{2}$, the circles centered at the data comprising the smaller square overlap such that there is a "void" of non-overlapping space enclosed by their region of overlap. Hence for $0.5 \leq r < \frac{\sqrt{2}}{2}$, there exists a non-contractible loop within the geometric Čech complex of radius r. "Non-contractible" here means that the loop drawn around the void of non-overlapping space cannot be continuously deformed down to a single point without leaving the geometric Čech complex: the loop gets "stuck" on the void encircled by the geometric Čech complex, like trying to pull a rubber band off of a broomstick of infinite length. Notice that for $r < 0.5$, none of the circles of the example dataset in Fig 1A overlap, let alone overlap in such a way that they encircle some non-overlapping space. Consequently, we are unable to construct non-contractible loops within the geometric Čech complex of radius r for $r < 0.5$.

Fig 1. 1-dimensional homological features of example dataset in $\mathbb{R}^2$ A: Plot of example data. B,C,D,E: The blue region depicts the geometric Čech complex of radius $0.5, \frac{\sqrt{2}}{2}$, 2.5, $\frac{5\sqrt{2}}{2}$, respectively. F: Persistence diagram: each blue triangle represents a 1-dimensional homological feature, i.e. an equivalence class of non-contractible loops, with coordinates (birth radius, death radius); the y=x line is drawn to depict the persistence distribution of all 1-dimensional homological features over a set of r values by noting that the more persistent a 1-dimensional homological feature is, i.e. the larger the difference between its death radius and birth radius, the more 'above' the y=x line it will be.

Recall that if we can draw a non-contractible loop within the geometric Čech complex of radius r for some $r > 0$, then this loop must be "stuck" around some void encircled by the geometric Čech complex. This non-contractible loop can be continuously deformed to construct another non-contractible loop "stuck" around the same void. We say the two loops are homotopic. As an example, the red and green loops shown in Fig 1D are homotopic. The set of all possible non-contractible loops "stuck" around some void encircled by the geometric Čech complex forms an equivalence class of non-contractible loops, i.e. a set of non-contractible loops where any two non-contractible loops in the set are homotopic. In practice, rather than homotopy, we use a weaker but more technically-involved equivalence relation on loops called homology to utilize efficient algorithms such as Ripser [26] and GUDHI [27] in computing topological features. For a rigorous treatment of homotopy and homology, see [28].

Suppose we are given a dataset X and a positive real number r. We write $\check{C}_r(X)$ for the geometric Čech complex of X of radius r. For a given two-dimensional dataset such that there exists a non-contractible loop ℓ within its geometric Čech complex of radius r, we define the *birth radius* of ℓ as the smallest real number b such that some loop in $\check{C}_r(X)$ which is equivalent to ℓ and which is contained in the subset $\check{C}_b(X)$ of $\check{C}_r(X)$ exists. It follows immediately from this definition that $b \leq r$. Similarly, we define the *death radius* of ℓ as the smallest real number d such that $r \leq d$ and such that ℓ is contractible when regarded as a loop in $\check{C}_d(X)$. That is, the birth radius of a non-contractible loop is the smallest radius at which the equivalence class of that

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.478609; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

non-contractible loop forms, and the death radius is the smallest radius at which it vanishes (i.e., becomes contractible). For $r \in [\text{birth radius}, \text{death radius}]$, the equivalence class of non-contractible loops 'persists,' and this motivates the definition of the *persistence* of an equivalence class of non-contractible loops: the persistence is the difference between the death radius and the birth radius.

The data comprising the larger square in the example dataset were included to compare two 1-dimensional homology features of different persistence within a given 2-dimensional dataset. As previously noted, the birth radius and death radius of the 1-dimensional homology feature constructed from the data comprising the smaller square are 0.5 and $\frac{\sqrt{2}}{2} \approx 0.707$, respectively, so the persistence of this feature is approximately 0.707-0.5=0.207. The birth radius of the 1-dimensional homology feature corresponding to the larger square is 2.5, and the death radius is $\frac{5\sqrt{2}}{2} \approx 3.536$. Thus its persistence is approximately 3.536-2.5=1.036. Since 1.036-0.207=0.829, the blue triangle representing the equivalence class of non-contractible loops formed from the subset of the data comprising the larger square lies a distance 0.829 further above the y=x line than the blue triangle representing the equivalence class of non-contractible loops formed from the subset of data comprising the smaller square. This illustrates how a highly persistent 1-dimensional homology feature has a larger "loop-like" structure than a 1-dimensional homology feature of lower persistence.

Also note how there are two choices of subsets of the dataset representative of the equivalence class of non-contractible loops with a smaller persistence, namely the four points that comprise the smaller square and the set of five points where four points comprise the smaller square and the other point is nearby. These choices of data points are referred to as *representative 1-cycles* and can be chosen such that the representative 1-cycle is minimal with respect to either i) the number of data points it consists of or ii) the area it spans. In the first case, we have a *cardinality-minimal* representative 1-cycle for the equivalence class; in the second case, we have a *area-minimal* representative 1-cycle for that equivalence class. In this example, both the cardinality-minimal and area-minimal representative 1-cycle of the equivalence class of non-contractible loops with a lower persistence consists of the four data points comprising the smaller square.

Background on topological data analysis

We now set out to formalize the notion of "equivalence classes of non-contractible loops that persist for a given range of radius values." Given a set of data X represented as a finite set of points in $\mathbb{R}^2$, we construct a simplicial complex as a topological space that approximates the structure of the data.

Definition 0.1 *A* simplicial complex *is a collection K of subsets of a finite set V such that:*

- $\{v\} \in K$ *for all* $v \in V$*, and*
- *if* $\tau \subset \sigma$ *for* $\sigma \in K$*, then* $\tau \in K$*.*

An element of V is referred to as a vertex*, and an element of K with cardinality $n+1$ is referred to as an n-*simplex*.*

There are several ways to construct a simplicial complex given a finite set of points in $\mathbb{R}^2$, and to be consistent with the geometry of the simple dataset example, we consider the radius r Vietoris-Rips complex, a simplicial complex constructed by considering a circle of radius $\frac{r}{2}$ around each point in our dataset and then including

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.478609; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$S \subset X$ as a simplex if the intersection of the balls of radius $\frac{r}{2}$ for each point in S is nonempty. An example of constructing the Vietoris-Rips complex for several values of r is shown in Fig 2.

Fig 2. Relation between the geometric Čech complex of radius r and the radius r Vietoris-Rips Complex A-C: Geometric Čech complex of radius 0.4, 0.6, 0.8 shown in blue. D-E: Geometric realization of radius 0.4, 0.6, 0.8 Vietoris-Rips complex; a point indicates a 0-simplex, a line indicates a 1-simplex whose members are the endpoints of the line, and a filled-in region indicates a n-simplex whose n+1 members are the data on the boundary of the region; the red regions in each plot indicates the simplicies which are born at r=0, r=0.5, and $\frac{\sqrt{2}}{2}$, from left to right.

Definition 0.2 *Given a dataset X represented as a finite subset of $\mathbb{R}^2$, and given a positive real number r, the* radius r Vietoris-Rips complex of X, *denoted $VR_r(X)$, is the simplicial complex given by the collection of all subsets U of X with the property that if $x_1, x_2 \in U$, then $|x_1 - x_2| < r$.*

Note that if $S \subset U$ for $U \in VR_r(X)$, then $|x_1 - x_2| < r$ for all $x_1, x_2 \in U$ implies $|x_1 - x_2| < r$ for all $x_1, x_2 \in S$. Thus the radius r Vietoris Rips complex of a finite subset of $\mathbb{R}^2$ defines a simplicial complex.

We are now in a position to be more concrete about the notion of an "equivalence class of non-contractible loops" within the geometric Čech complex, as discussed in the "Example: a 1-dimensional topological feature in a simple dataset" section. By an "equivalence class of non-contractible loops," we are referring to an element of the 1-dimensional homology group of some radius r Vietoris-Rips complex, which we now set out to define.

Let X be a finite subset of $\mathbb{R}^2$, let r be a positive real number, and let C_n be the vector space over $\mathbb{F}_2$ with basis consisting of the elements of $VR_r(X)$ of cardinality $n+1$ for $n = 0, 1, 2$. Furthermore, suppose there is an ordering on $VR_r(X)$. Consider $0 \xleftarrow{\delta_{-1}} C_0 \xleftarrow{\delta_0} C_1 \xleftarrow{\delta_1} C_2 \xleftarrow{\delta_2} 0$ where $\delta_n([x_0, ..., x_n]) = \sum_{i=0}^{i} (-1)^n [x_0, ..., \hat{x_i}, ..., x_n]$ and $\hat{x_i}$ indicates that x_i is omitted from the ordered simplex. The elements of C_1 are referred to as 1-chains, the elements of $\ker(\delta_0)$ are referred to as 1-cycles, and elements of $\mathrm{im}(\delta_1)$ are referred to as 1-boundaries. Since $\delta_0(\delta_1(v)) = 0$ for all $v \in C_2$, every 1-boundary is an 1-cycle. However, it is not necessarily true that every 1-cycle is an 1-boundary. Intuitively, if we think of X as a cloud of points in the plane, the 1-dimensional homology group of $VR_r(X)$ is defined such that its dimension over $\mathbb{F}_2$ counts the number of "holes" in that cloud.

Definition 0.3 *Given $r > 0$ and $VR_r(X)$ where X is a finite subset of $\mathbb{R}^2$, we follow the construction of $\mathbb{F}_2$-vector spaces C_0, C_1, C_2 and linear transformations δ_{-1}, δ_0, δ_1, δ_2 as outlined above and define the first homology group of $VR_r(X)$ as the quotient vector space $H_1(VR_r(X)) = ker(\delta_0)/im(\delta_1)$. The $\mathbb{F}_2$-vector space dimension $\beta_1 = dim(H_1(VR_r(X))) = dim(ker(\delta_0)) - dim(im(\delta_1))$ of $H_1(VR_r(X))$ is called the* first Betti number.

By increasing r, we create a sequence of Vietoris-Rips Complexes where $VR_r(X) \subset VR_{r'}(X)$ for $r < r'$. We then construct

$$VR_{r_0}(X) \xrightarrow{i_0} VR_{r_1}(X) \xrightarrow{i_1} ... \xrightarrow{i_{m-1}} VR_{r_m}(X)$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.478609; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

where $VR_{r_i}(X)$ is a proper subset of $VR_{r_j}(X)$ for $i < j$ and i_0, i_1,...,i_{m-1} are inclusion homomorphisms. This induces a sequence of $\mathbb{F}_2$-linear functions i_0^*, i_1^*,..., i_{m-1}^* such that

$$H_1(VR_{r_0=0}(X)) \xrightarrow{i_0^*} H_1(VR_{r_1}(X)) \xrightarrow{i_1^*} ... \xrightarrow{i_{m-1}^*} H_1(VR_{r_m}(X))$$

and $i_n^*([c]_V) = [i_n(c)]_W$ for $V = VR_{r_n}(X)$, $W = VR_{r_{n+1}}(X)$, and all $n = 0, 1, ..., m-1$. We now give a name to the "smallest" and "largest" $r > 0$ such that a given 1-cycle belongs to $H_1(VR_r(X))$.

Definition 0.4 *Let $[c] \in H_1(VR_r(X))$ for some $r > 0$. The birth filtration of $[c]$ is defined as the greatest lower bound of the set of all $\epsilon > 0$ such that $[c]$ is in the range of the $\mathbb{F}_2$-linear function $H_1(VR_\epsilon(X)) \to H_1(VR_r(X))$. Similarly, the death filtration of $[c]$ is defined as the least upper bound of the set of all $\epsilon > 0$ such that $[c]$ maps to zero under the $\mathbb{F}_2$-linear function $H_1(VR_r(X)) \to H_1(VR_\epsilon(X))$. The persistence of $[c]$ is defined as the difference between the death filtration and the birth filtration.*

Up to a scaling factor in the variable r, the geometric Čech complex of radius r is homotopy equivalent to the radius r Vietoris-Rips complex due to the Nerve Lemma [29]. Consequently, the definitions of the birth and death radii of an equivalence class of non-contractible loops presented in the "Example: a 1-dimensional topological feature in a simple dataset" section are equivalent to the definitions of the birth and death filtration of a class $[c] \in H_1(VR_r(X))$ given in Definition 0.4. For a more thorough treatment of persistent homology, see [30].

Methods

In this section, we first describe how ECG data are processed to be in a form such that our topological approach can be applied consistently. We then describe our topological method of identifying P,Q,S, and T-waves and measuring the PR-interval, QT-interval, ST-segment, QRS-duration, P-wave duration, and T-wave duration. Lastly, we present the methods used to simulate ECG signals with the PR-interval, QT-interval, ST-segment, QRS-duration, P-wave duration, and T-wave duration determined by parameters of the simulation.

Data processing

Before an ECG signal can be analyzed using our topological approach, the signal is processed such that the topological approach can be consistently applied. An outline of this processing pipeline is depicted in Fig 3. This processing pipeline is the same regardless of whether the ECG signal is simulated or real.

Fig 3. ECG Signal Processing Pipeline
The above flowchart depicts a processing pipeline which is applied to an ECG signal prior to computing topological features of the signal.

Given an ECG signal (D, f) where D denotes the set of time indices and $f : D \to \mathbb{R}$ is such that $f(t)$ represents the amplitude of the signal at time t, we first nearly double the number of points comprising the signal by including the point $(t + \frac{h}{2}, \frac{f(t)+f(t+h)}{2})$ for every pair of adjacent points $(t, f(t))$ and $(t + h, f(t + h))$ of the signal. The signal is then normalized by mapping every ordered pair $(t, f(t))$ to $(t, \frac{f(t)}{M})$ where $M = \underset{t \in D}{\text{maximum}}\{f(t)\}$. The coordinates of the R-wave peaks are used in identifying P,Q,S, and T-waves, and the coordinate pair of the i-th R-wave peak is computed as the

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.478609; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ordered pair $(t_{R,i}, f(t_{R,i}))$ where $0.6 < f(t_{R,i})$, $f(t_{R,i})$ is a local maximum of the signal, and there are exactly $i-1$ local maxima $f(t_{R,j})$ such that $t_{R,j} < t_{R,i}$ and $0.6 < f(t_{R,j})$, $j = 1, 2, ..., i-1$ [31]. Next, an isoelectric line is incorporated into the signal by mapping $(t, \frac{f(t)}{M})$ to $(t, baseline)$ where $baseline$ is the median value of the set $\{f^*(t) \mid t \in D$ and $f^*(t)$ is $f(t)$ rounded to the nearest hundredth$\}$ for all $t \in D$ such that $t = h(2k)$ for some integer k. Lastly, we consider the restriction of the processed ECG signal such that the signal begins and ends with an R-wave peak as the input of our topological method.

Extracting features of processed ECG signals using TDA

Given a processed ECG signal (D, f), we are interested in i) identifying subsets of the signal as P,Q,S, and T-waves using cycle reconstructions of 1-dimensional homology classes and ii) using the identified waves to measure the PR-interval, QT-interval, ST-segment, QRS-duration, P-wave duration, and T-wave duration. To describe how P,Q,S, and T-waves are identified as 1-cycles with certain properties, if 1-cycles with these properties exist, we first give a general description of five properties of ECG signals that are used in this process:

- **Persistence:** Due to the the addition of the isoelectric line, specifically under the P,T-waves and over the Q,S-waves, there is a loop-like structure of data comprising the P,Q,S, and T-waves as shown in Fig 4A. Thus for a 1-cycle to be identified as a P,Q,S, or T-wave, we impose that it must have a persistence within some bounded interval depending on the specific wave.

- **Birth filtration upper bound:** It is a property of ECG signals that adjacent points $(t, f(t))$, $(t+h, f(t+h))$ within each P,Q,S, and T-wave are within some distance δ from one another where δ is small relative to the distance between $(t, f(t))$ and a point $(t', f(t'))$ on a different wave. Thus we impose an upper bound on the birth filtration of each potential 1-cycle to exclude 1-cycles constructed from points corresponding to multiple waves, e.g. a 1-cycle constructed from both a P-wave and an R-wave as shown in Fig 4B.

- **Time coordinate of centroid of 1-cycle:** It is a property of ECG signals that a T-wave follows a QRS-complex. We also expect a detected P-wave to necessarily precede a QRS-complex due to our simulations not featuring atrial activity with missed ventricular beats. Thus we expect that the centroids of 1-cycles comprising a T-wave and P-wave lie between a pair of adjacent R-waves such that the T-wave is some distance after the preceding R-wave, the P-wave is some distance prior to the following R-wave, and the P-wave follows the T-wave. Similarly, we impose that a 1-cycle comprising a detected Q-wave and S-wave slightly precedes an R-wave and shortly follows an R-wave, respectively.

- **Amplitude coordinate of centroid of 1-cycle:** It is expected that over a given heartbeat, the average amplitude of the signal in the P,T-wave be larger than the signal amplitude at baseline. Additionally, we expect there to be a positive upper bound on the average amplitude of the signal in the P,T-wave. Conversely, it is expected that the average amplitude of the signal in the Q,S-wave be less than the signal amplitude at baseline.

- **Optimality:** Of the 1-cycles satisfying the persistence, birth filtration, and centroid coordinate conditions above, the 1-cycle representing the given P,Q,S, or T-wave is expected to be optimal in the sense of i) having a minimal number of data points comprising the 1-cycle and/or ii) having a minimal-area convex hull. These optimal cycles were computed using HomCloud [32].

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.478609; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fig 4. Illustration of properties of an ECG used to detect P,Q,S, and T-waves
A: The red loops illustrate elements from equivalence classes of non-contractible loops corresponding to the P,Q,S, and T-wave. Notice that without the addition of the isoelectric baseline, we would be unable to draw non-contractible loops corresponding to the P,Q,S, and T-waves within the geometric Čech complex. B: Example of output without imposing an upper bound on the birth filtration of a 1-cycle identified as a P-wave.

We now present the specific criteria that identifies 1-cycles with certain properties related to those described above as being a P-wave or T-wave. We note that differences in the classification criteria for P-waves and T-waves involves changes in the bounds of the persistence range and changes to the upper bound of the birth filtration. Additionally, the centroid of a 1-cycle identified as a P-wave is expected to be closer to the subsequent R-wave than the preceding R-wave, and vice versa for the centroid of a 1-cycle identified as a T-wave.

P-wave Identification: An optimal 1-cycle satisfying the following conditions is identified as a P-wave (U_P, f) with collection of ordered pairs $\{(t, f(t)) \mid t \in U_P\}$ and time index set U_P:

- the persistence of the 1-cycle is in [0.001,0.2]
- the birth filtration of the 1-cycle is less than 0.03
- $0 < \underset{t \in U_P}{\text{avg}} \{f(t)\} < 0.15$
- $\text{R_peak_time_coordinate[i]} - 0.35 * \text{RR_interval} < \underset{t \in U_P}{avg} \{t\} <$ $\text{R_peak_time_coordinate[i]} - 0.06 * \text{RR_interval}$ for some i.

T-wave Identification: An optimal 1-cycle satisfying the following conditions is identified as a T-wave (U_T, f) with collection of ordered pairs $\{(t, f(t)) \mid t \in U_T\}$ and time index set U_T:

- the persistence of the 1-cycle is in [0.01,0.6]
- the birth filtration of the 1-cycle is less than 0.04
- $0 < \underset{t \in U_T}{\text{avg}} \{f(t)\} < 0.4$
- $\text{R_peak_time_coordinate[i]} + 0.15 * \text{RR_interval} < \underset{t \in U_T}{avg} \{t\} <$ $\text{R_peak_time_coordinate[i]} + 0.5 * \text{RR_interval}$ for some i.

We continue with describing the criteria for a 1-cycle to be identified as a Q-wave or S-wave. The interval which the persistence of a 1-cycle must belong to for the 1-cycle to be classified as a Q-wave or S-wave includes a smaller lower bound and smaller upper bound compared to the interval of valid persistence values for a 1-cycle to be classified as a P,T-wave. This is due to the data comprising Q-waves and S-waves having less of a loop-like structure than data comprising the P-waves and T-waves. We also impose that the centroid of a 1-cycle identified as a Q-wave or S-wave more closely precede or follow the nearby R-wave peak relative to a P-wave or T-wave, respectively.

Q-wave Identification: An optimal 1-cycle satisfying the following conditions is identified as a Q-wave (U_Q, f) with collection of ordered pairs $\{(t, f(t)) \mid t \in U_Q\}$ and time index set U_Q:

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.478609; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

- the persistence of the 1-cycle is in [0.007,0.1]
- the birth filtration of the 1-cycle is less than 0.06
- $\underset{t \in U_Q}{\text{avg}} \{f(t)\} < 0$
- $\text{R_peak_time_coordinate[i]} - 0.12 * \text{RR_interval} < \underset{t \in U_Q}{avg} \{t\} <$ $\text{R_peak_time_coordinate[i]}$ for some i.

S-wave Identification: An optimal 1-cycle satisfying the following conditions is identified as a S-wave (U_S, f) with collection of ordered pairs $\{(t, f(t)) \mid t \in U_S\}$ and time index set U_S:

- the persistence of the 1-cycle is in [0.007,0.1]
- the birth filtration of the 1-cycle is less than 0.06
- $\underset{t \in U_S}{\text{avg}} \{f(t)\} < 0$
- $\text{R_peak_time_coordinate[i]} < \underset{t \in U_S}{avg} \{t\} <$ $\text{R_peak_time_coordinate[i]} + 0.12 * \text{RR_interval}$ for some i.

In the above criteria, "optimal 1-cycle" refers to a 1-cycle computed as being minimal with respect to i) the cardinality of the set of data points comprising the 1-cycle or ii) the area spanned by the convex hull of the 1-cycle. Both types of optimal 1-cycles were computed in our analysis and are compared. An example of area-minimal 1-cycles identified as P,Q,S, and T-waves using the above criteria depicted on an ECG signal is shown in Fig 5. Cardinality-minimal and area-minimal 1-cycles identified as P,Q,S, and T-waves for the 400 simulated ECG signals is provided as Supplementary File 1 and Supplementary File 2. Additionally, cardinality-minimal and area-minimal 1-cycles identified as P,Q,S, and T-waves for 200 Lead II ECG signals randomly sampled from a pool of 10646 denoised signals with 11 common rhythms as described in [33] are shown in Supplementary File 3 and Supplementary File 4, respectively.

Fig 5. Example of simulated ECG signal and persistence diagram with area-optimal 1-cycles identified as P,Q,S, and T-waves
The points comprising P,Q,S, and T-waves were identified as area-optimal 1-cycles with certain properties depending on their birth filtration, persistence, and centroid. The Q-wave and S-wave 1-cycles drawn have the left-most and right-most centroid of all Q-wave and S-waves detected within a given heartbeat, respectively. This is consistent with the choice of Q-waves and S-waves used to measure intervals of interest. The color of 1-cycles identified as P,Q,S, and T-waves matches the color of the corresponding points in the persistence diagram.

We now describe how the PR-interval, QT-interval, ST-segment, QRS-duration, P-wave duration, and T-wave duration are measured using 1-cycles identified as P,Q,S, and T-waves. To measure these intervals solely using topological features of the data, the PR-interval and QR-interval of a given heartbeat are only measured if a Q-wave is detected. Similarly, the ST-segment of a given heartbeat is only measured if an S-wave is detected. The QRS-duration is only measured if both a Q-wave and an S-wave are detected. The intervals of interest are computed as the difference between their offset and onset. Given a heartbeat of a processed ECG signal with at least one detected Q,S-wave and one detected P,T-wave, we measure the onset and offset of the intervals of interest as follows:

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.478609; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

- The onset of the PR-interval and onset of the P-wave duration is minimum$\{U_P\}$.
- The offset of the PR-interval, onset of the QT-interval, and onset of the QRS-duration is given by minimum$\{U_Q^*\}$ where U_P^* is the set of time indices of the 1-cycle satisfying the Q-wave criteria such that the time coordinate of its centroid is less than the time coordinate of the centroid of every other 1-cycle identified as a Q-wave within the heartbeat.
- The onset of the ST-segment and offset of the QRS-duration is given by minimum$\{U_S^*\}$ where U_S^* is the set of time indices of the 1-cycle satisfying the S-wave criteria such that the time coordinate of its centroid is greater than the time coordinate of the centroid of every other 1-cycle identified as an S-wave within the heartbeat.
- The offset of the QT-interval and offset of the T-wave duration is minimum$\{U_T\}$.
- The offset of the ST-segment and onset of the T-wave duration is minimum$\{U_T\}$.
- The offset of the P-wave duration is maximum$\{U_P\}$.

The average PR-interval, QT-interval, ST-segment, QRS-duration, P-wave duration, and T-wave duration were calculated using i) the model's parameters, ii) cardinality-minimal 1-cycle reconstructions, and iii) area-minimal 1-cycle reconstructions for all 1000 simulated ECG signals. The percent difference between the model's measurement of each interval and the optimal 1-cycle's measurement of each interval for both cardinality-minimal and area-minimal 1-cycles were calculated for each simulation.

ECG simulation

ECG signals were generated with a sampling frequency of 500Hz using an ECG Simulator in MATLAB [34]. The set of parameters that completely characterizes what we refer to as a simulated ECG signal includes the heart rate, the amplitude and duration of the R-wave, and the amplitude, duration, and location with respect to i) the subsequent R-wave for the P,Q-wave and ii) the preceding R-wave for the S,T-wave. We now go into detail about how this parameter set is used to construct a simulated ECG signal with a given PR-interval, QT-interval, ST-segment, QRS-duration, P-wave duration, and T-wave duration.

Let $h = \frac{4-0}{4*500} = 0.002$ be the time-step and $D = \{x \in [0,4] \mid x = hk \text{ for } k \in \mathbb{Z}\}$ be the set of time indices of a simulated ECG signal. A simulated ECG signal (D, f) is given by a set of ordered pairs $(t, f(t))$ where $t \in D$ and $f : [0,4] \to \mathbb{R}$ is such that $f(t)$ represents the amplitude of the signal at time t. Furthermore, the simulated ECG signal is periodic, i.e. $f(t) = f(t + rr_interval)$ where $rr_interval = \frac{60}{heart_rate}$. Thus to completely characterize a simulated ECG signal, we need only know how f is defined on a contiguous subset E of D which ranges over one period of f. We also impose the property of a simulated ECG signal such that any choice of E contains five subsets of the simulated ECG signal which we refer to as the P,Q,R,S, and T-waves. Depending on the choice of E and the parameters of the model, the P,Q,R,S, and T-waves are either connected by an isoelectric segment on which f is constant or intersect one another.

We now briefly describe the P,Q,R,S, and T-waves. The QRS complex consists of six line segments, and the shape of the P and T-wave is dictated by a Fourier series [34]. The amplitude and duration of the X-wave is given by the parameters amplitude_X_wave and duration_X_wave, respectively, for X=P,Q,R,S,T. The distance along the time-axis from the centroid of the ordered pairs comprising the

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.478609; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

P,Q-wave to the coordinate pair of the following R-wave peak is given by the parameter location_P,Q_wave. Similarly, the distance along the time-axis from the coordinate pair of an R-wave peak to the centroid of the ordered pairs $(t, f(t))$ comprising the subsequent S,T-wave is given by the parameter location_S,T_wave.

We now describe how the intervals of interest, i.e. the PR-interval, QT-interval, ST-segment, QRS-duration, P-wave duration, and T-wave duration, are constructed from the model parameters. We consider the location on the time axis corresponding to the R-wave peak as the origin and measure distance as being positive in both directions along the time axis. Thus, both the onset and offset of the R-wave is given by $\frac{duration_R_wave}{2}$. The onset of the P,Q-wave and the offset of the S,T-wave is taken to be $location_X_wave + \frac{duration_X_wave}{2}$ for X=P,Q,S,T. Similarly, the offset of the P,Q-wave and onset of the S,T-wave is given as $location_X_wave - \frac{duration_X_wave}{2}$ for X=P,Q,S,T. We now define the simulation's measure of these intervals as

- PR-interval = onset of P-wave - onset of Q-wave
- QT-interval = onset of Q-wave + offset of T-wave
- ST-segment = onset of T-wave - offset of S-wave
- QRS-duration = onset of Q-wave + offset of S-wave
- P-wave duration = onset of P-wave - offset of P-wave
- T-wave duration = offset of T-wave - onset of T-wave

where if the amplitude of the Q-wave or S-wave is below 0.02, then the onset of the Q-wave or offset of the S-wave is given by the onset or offset of the R-wave, respectively.

By uniformly sampling the heart rate from [60,100] and the parameter values of the P,Q,R,S, and T-waves from the ranges in Table 1, 1000 simulated ECGs were generated.

Table 1. Parameter ranges of simulated ECG signals.

	P-wave	**Q-wave**	**R-wave**	**S-wave**	**T-wave**
Amplitude	[0, 0.4]	[0, 0.2]	N/A	[0, 0.15]	[0.15, 0.6]
Duration	[0.08, 0.15]	[0.04, 0.07]	[0.03, 0.04]	[0.04, 0.07]	[0.08, 0.2]
Location	[0.17*rr_int, 0.3*rr_int]	[0.04*rr_int, 0.07*rr_int]	N/A	[0.04*rr_int, 0.07*rr_int]	[0.26*rr_int, 0.38*

Note that R-waves are evenly spaced a distance of $\frac{60}{heart\ rate}$ from one another and that the location of the P,Q,S, and T-waves are defined with respect to the subsequent R-wave in the case of P,Q-waves and to the preceding R-wave in the case of S,T-waves.

Results/discussion

The PR-interval, QT-interval, ST-segment, QRS-duration, P-wave duration, and T-wave duration of 1000 simulated ECG signals were measured using i) cardinality-minimal 1-cycle reconstructions and ii) area-minimal 1-cycle reconstructions and compared with the interval measurements made by model parameters. For each simulation, the percent difference between the model's measurement of each interval and the measurement of each interval performed using optimal 1-cycle reconstructions was calculated. The distribution of these measurements using cardinality-minimal and area-minimal 1-cycles are shown in Fig 6 and Fig 7, respectively. The average percent difference and the

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.478609; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

standard deviation of the percent difference measurements for each interval of interest across the 1000 simulated ECG signals is shown in Table 2.

Fig 6. Distribution of percent difference measurements using cardinality-optimal 1-cycles
Each histogram depicts the distribution of percent difference measurements for a single interval of interest. Left to right across the top row: PR-interval, QT-interval, ST-segment. Left to right across the bottom row: QRS-duration, P-wave duration, T-wave duration.

Fig 7. Distribution of percent difference measurements using area-optimal 1-cycles
Each histogram depicts the distribution of percent difference measurements for a single interval of interest. Left to right across the top row: PR-interval, QT-interval, ST-segment. Left to right across the bottom row: QRS-duration, P-wave duration, T-wave duration.

Table 2. Average percent difference of interval measurements.

	Area-minimal 1-cycles	**Cardinality-minimal 1-cycles**
	mean ± standard deviation	mean ± standard deviation
$PR-interval$	9.712 ± 14.581	9.712 ± 14.581
$QT-interval$	4.668 ± 3.677	4.668 ± 3.675
$ST-segment$	25.122 ± 38.417	25.122 ± 38.417
$QRS-duration$	25.284 ± 23.791	25.285 ± 23.790
$P-wave\ duration$	7.473 ± 17.146	7.473 ± 17.146
$T-wave\ duration$	2.972 ± 9.136	2.972 ± 9.136

The intervals of interest of 1000 simulated ECG signals were measured using parameters of the model, cardinality-minimal 1-cycle reconstructions, and area-minimal 1-cycle reconstructions as outlined in the Methods section. The reported percent difference values refer to the percent difference in interval measurements between the model's measurements and measurements made using optimal 1-cycles.

From comparing cardinality-minimal 1-cycles identified as P,Q,S, and T-waves to area-minimal 1-cycles identified as P,Q,S, and T-waves in Supplementary File 1 and Supplementary File 2, we see that these cycles are often "close" to one another, likely resulting in the similar percent error distributions seen in Fig 6, Fig 7, and Table 2. It is also evident from Supplementary File 1 and Supplementary File 2 that the algorithm could be improved by using a more stable method to include an isoelectric baseline in the signal. Additionally, to ensure that P,Q,S, and T-wave identifications and interval measurements are not compromised due to misidentified or undetected R-wave peaks, a more robust R-wave peak detection algorithm could be used such as those in [35], [36], [37].

The parameters in the criteria used to identify 1-cycles as P,Q,S, and T-waves are currently determined from several iterations of identifying 1-cycles as P,Q,S, and T-waves and adjusting the parameter values to obtain more accurate identification based on visual inspection. A more robust method of optimizing this parameter set could result in more accurate identification of 1-cycles which better represent the shape and/or endpoints of the P,Q,S, and T-waves. A difficulty in this is that it may be difficult to know what a "correct" subset of a signal is in order for it to be considered a P,Q,S, or T-wave. Additionally, information about optimal 1-cycles identified as P,Q,S, and T-waves could be combined with other existing approaches such as analyzing other

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.478609; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

persistent homology statistics, wavelet decompositions, and machine learning for automated arrhythmia detection in future work [12] [13] [17] [19] [21].

Supporting information

The code used to produce simulated ECG signals is available at `https://www.mathworks.com/matlabcentral/fileexchange/10858-ecg-simulation-using-matlab/` [34]. The real ECG data used in this study is publicly available and can be freely downloaded at `https://www.nature.com/articles/s41597-020-0386-x/` [33]. The code used to identify P,Q,S, and T-waves and measure intervals of interest based off of topological features of the ECG signal can be found in the GitHub repository at `https://github.com/hdlugas/ekg_tda`.

Supplementary File 1 Cardinality-minimal 1-cycle reconstructions identified as P,Q,S,T-waves for 400 simulated ECG signals

Supplementary File 2 Area-minimal 1-cycle reconstructions identified as P,Q,S,T-waves for 400 simulated ECG signals

Supplementary File 3 Cardinality-minimal 1-cycle reconstructions identified as P,Q,S,T-waves for 200 ECG signals randomly sampled from the pooled described in 'Methods-Extracting Features of Processed ECG Signals using TDA'

Supplementary File 4 Area-minimal 1-cycle reconstructions identified as P,Q,S,T-waves for 200 ECG signals randomly sampled from the pooled described in 'Methods-Extracting Features of Processed ECG Signals using TDA'

Supplementary File 5 File names of the randomly sampled ECG signals analyzed using cardinality-minimal 1-cycles.

Supplementary File 6 File names of the randomly sampled ECG signals analyzed using cardinality-minimal 1-cycles.

Acknowledgments

The author would like to thank Andrew Salch for discussions on interpreting topological features of data and for guidance towards submitting this paper. The author would also like to thank Hassan Abdallah for discussions on i) persistent homology software and ii) a statistical analysis comparing interval measurements performed using cardinality-minimal and area-minimal 1-cycle reconstructions.

References

1. World Health Organization: The top 10 causes of death .

2. Centers for Disease Control and Prevention: Leading causes of death .

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.478609; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3. GA Oksak GI. Contribution of mortality from cardiovascular disease to overall mortality. Wiad Lek. 2017 .

4. Pyakillya B, Kazachenko N, Mikhailovsky N. Deep Learning for ECG Classification. Journal of Physics: Conference Series. 2017;913:012004. doi:10.1088/1742-6596/913/1/012004.

5. Rahhal MMA, Bazi Y, AlHichri H, Alajlan N, Melgani F, Yager RR. Deep learning approach for active classification of electrocardiogram signals. Information Sciences. 2016;345:340–354. doi:https://doi.org/10.1016/j.ins.2016.01.082.

6. Feeny AK MAea Chung MK. Artificial Intelligence and Machine Learning in Arrhythmias and Cardiac Electrophysiology. Circ Arrhythm Electrophysiol. 2020;doi:10.1161/CIRCEP.119.007952.

7. Wang G, Zhang C, Liu Y, Yang H, Fu D, Wang H, et al. A global and updatable ECG beat classification system based on recurrent neural networks and active learning. Information Sciences. 2019;501:523–542.

8. Guo L, Sim G, Matuszewski B. Inter-patient ECG classification with convolutional and recurrent neural networks. Biocybernetics and Biomedical Engineering. 2019;39(3):868–879.

9. Chen S, Hua W, Li Z, Li J, Gao X. Heartbeat classification using projected and dynamic features of ECG signal. Biomedical Signal Processing and Control. 2017;31:165–173.

10. Asgharzadeh-Bonab A, Amirani MC, Mehri A. Spectral entropy and deep convolutional neural network for ECG beat classification. Biocybernetics and Biomedical Engineering. 2020;40(2):691–700.

11. Sannino G, De Pietro G. A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. Future Generation Computer Systems. 2018;86:446–455.

12. Ye C CM Kumar BV. Heartbeat classification using morphological and dynamic features of ECG signals. IEEE Trans Biomed Eng. 2012;doi:10.1109/TBME.2012.2213253.

13. Kutlu Y KD. Feature extraction for ECG heartbeats using higher order statistics of WPD coefficients. Comput Methods Programs Biomed. 2012;doi:10.1016/j.cmpb.2011.10.002.

14. Elhaj FA, Salim N, Harris AR, Swee TT, Ahmed T. Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals. Computer Methods and Programs in Biomedicine. 2016;127:52–63.

15. Özal Yıldırım, Pławiak P, Tan RS, Acharya UR. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. Computers in Biology and Medicine. 2018;102:411–420.

16. Özal Yildirim. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Computers in Biology and Medicine. 2018;96:189–202.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.478609; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

17. Dindin M, Umeda Y, Chazal F. Topological Data Analysis for Arrhythmia Detection Through Modular Neural Networks. In: Goutte C, Zhu X, editors. Advances in Artificial Intelligence. Springer International Publishing; 2020.

18. Safarbali B, Hashemi Golpayegani SMR. Nonlinear dynamic approaches to identify atrial fibrillation progression based on topological methods. Biomedical Signal Processing and Control. 2019;53:101563. doi:https://doi.org/10.1016/j.bspc.2019.101563.

19. Ignacio PS, Dunstan C, Escobar E, Trujillo L, Uminsky D. Classification of Single-Lead Electrocardiograms: TDA Informed Machine Learning. In: 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA); 2019. p. 1241–1246.

20. Ni Y, Sun F, Luo Y, Xiang Z, Sun H. A Novel Heart Disease Classification Algorithm based on Fourier Transform and Persistent Homology; 2021.

21. Graff G, Graff B, Pilarczyk P, Jabłoński G, Gąsecki D, Narkiewicz K. Persistent homology as a new method of the assessment of heart rate variability. PLOS ONE. 2021;16(7):1–24. doi:10.1371/journal.pone.0253851.

22. Chung YM, Hu CS, Lo YL, Wu HT. A Persistent Homology Approach to Heart Rate Variability Analysis With an Application to Sleep-Wake Classification. Frontiers in Physiology. 2021;12:202.

23. Altındiş F, Yılmaz B, Borisenok S, İçöz K. Parameter investigation of topological data analysis for EEG signals. Biomedical Signal Processing and Control. 2021;63:102196.

24. Ormrod Morley D WM Salmon PS. Persistent homology in two-dimensional atomic networks. J Chem Phys. 2021;doi:10.1063/5.0040393.

25. Hiraoka Y, Nakamura T, Hirata A, Escolar EG, Matsue K, Nishiura Y. Hierarchical structures of amorphous solids characterized by persistent homology. Proceedings of the National Academy of Sciences. 2016;113(26):7035–7040. doi:10.1073/pnas.1520877113.

26. Bauer U. Ripser: efficient computation of Vietoris-Rips persistence barcodes. Journal of Applied and Computational Topology. 2021;doi:10.1007/s41468-021-00071-5.

27. Maria C, Boissonnat JD, Glisse M, Yvinec M. The Gudhi Library: Simplicial Complexes and Persistent Homology. In: Hong H, Yap C, editors. Mathematical Software – ICMS 2014. Springer Berlin Heidelberg; 2014.

28. Hatcher A. Algebraic Topology. Cambridge University Press; 2002.

29. Meunier F, Montejano L. Different versions of the nerve theorem and colourful simplices. Journal of Combinatorial Theory, Series A. 2020;169:105125. doi:https://doi.org/10.1016/j.jcta.2019.105125.

30. Tamal Dey YW. Computational Topology for Data Analysis. Cambridge University Press; 2021.

31. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods. 2020;17:261–272. doi:10.1038/s41592-019-0686-2.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.478609; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

32. Obayashi I. Volume-Optimal Cycle: Tightest Representative Cycle of a Generator in Persistent Homology. SIAM Journal on Applied Algebra and Geometry. 2018;2(4):508–534.

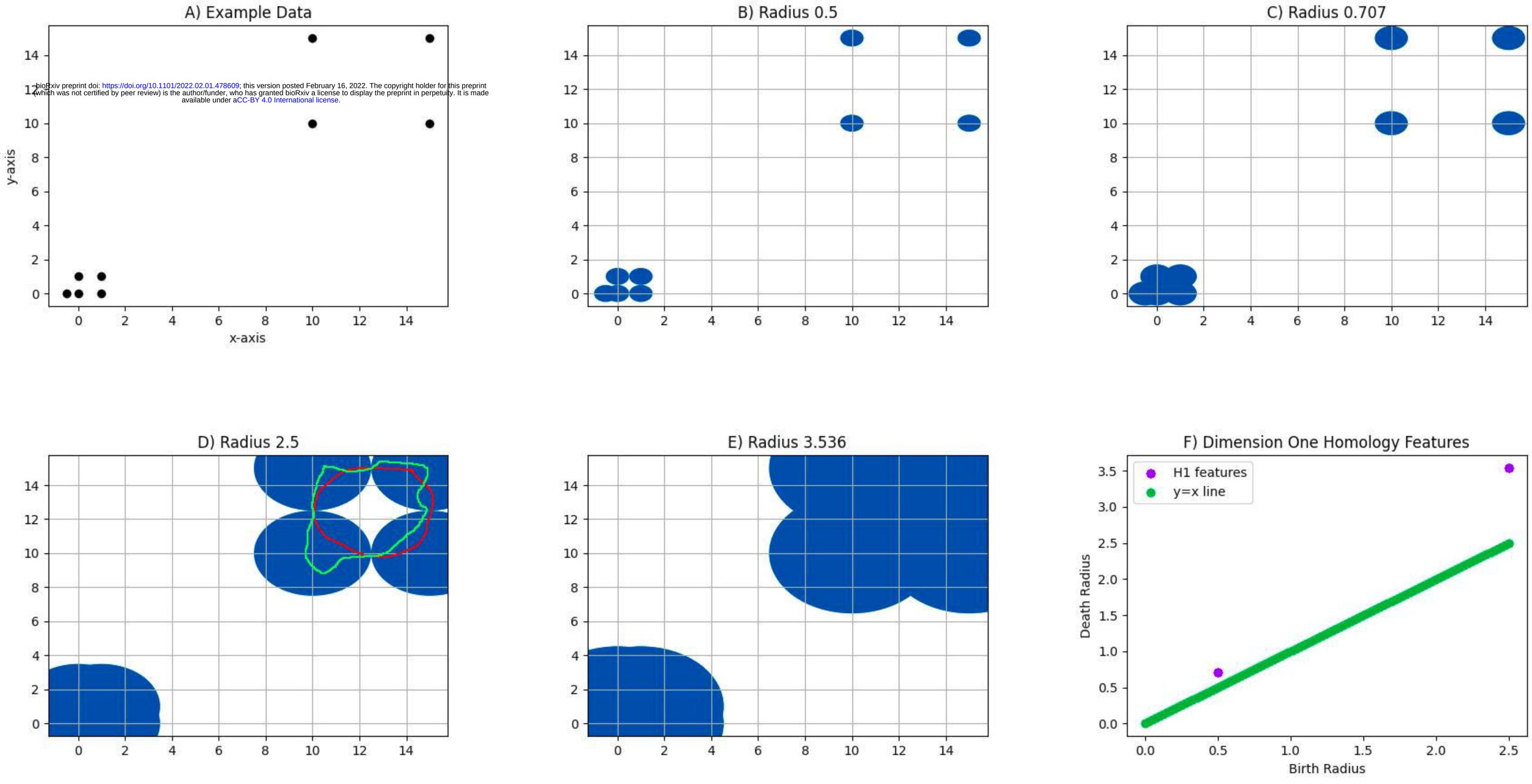
33. Jianwei Zheng SDHYHGCR Jianming Zhang. A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. Sci Data;doi:10.1038/s41597-020-0386-x.

34. Raviprakash K. ECG simulation using MATLAB, MATLAB Central File Exchange .

35. R-Peak Detection Algorithm for Ecg using Double Difference And RR Interval Processing. Procedia Technology. 2012;4:873–877. doi:https://doi.org/10.1016/j.protcy.2012.05.143.

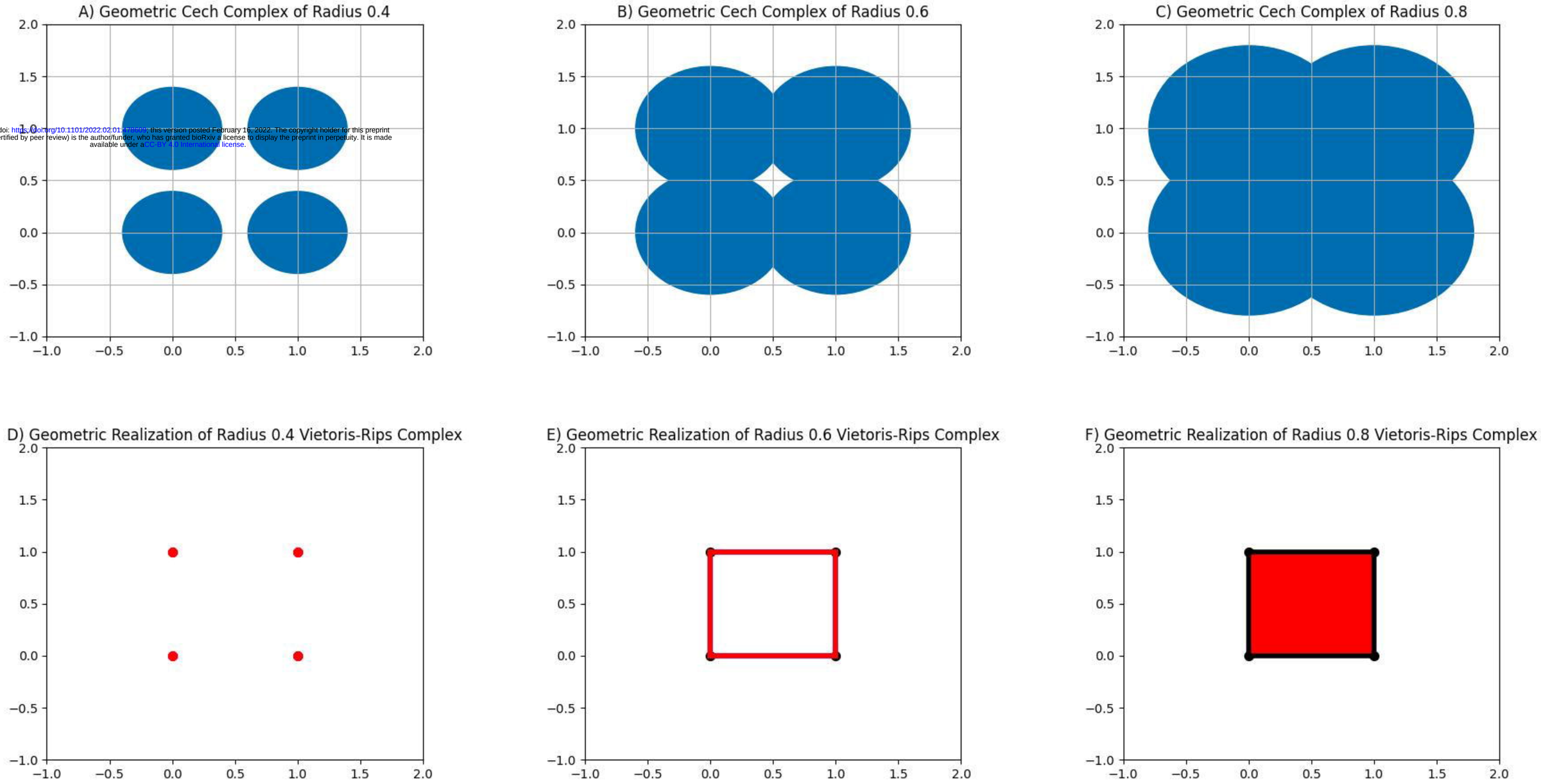
36. Zou Y, Han J, Xuan S, Huang S, Weng X, Fang D, et al. An Energy-Efficient Design for ECG Recording and R-Peak Detection Based on Wavelet Transform. IEEE Transactions on Circuits and Systems II: Express Briefs. 2015;62(2):119–123. doi:10.1109/TCSII.2014.2368619.

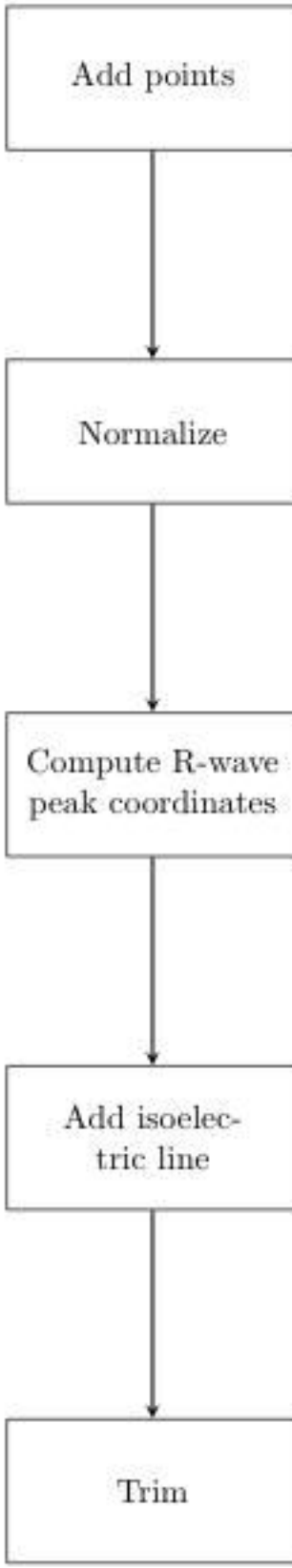
37. Sathyapriya L, Murali L, Manigandan T. Analysis and detection R-peak detection using Modified Pan-Tompkins algorithm. In: 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies; 2014. p. 483–487.



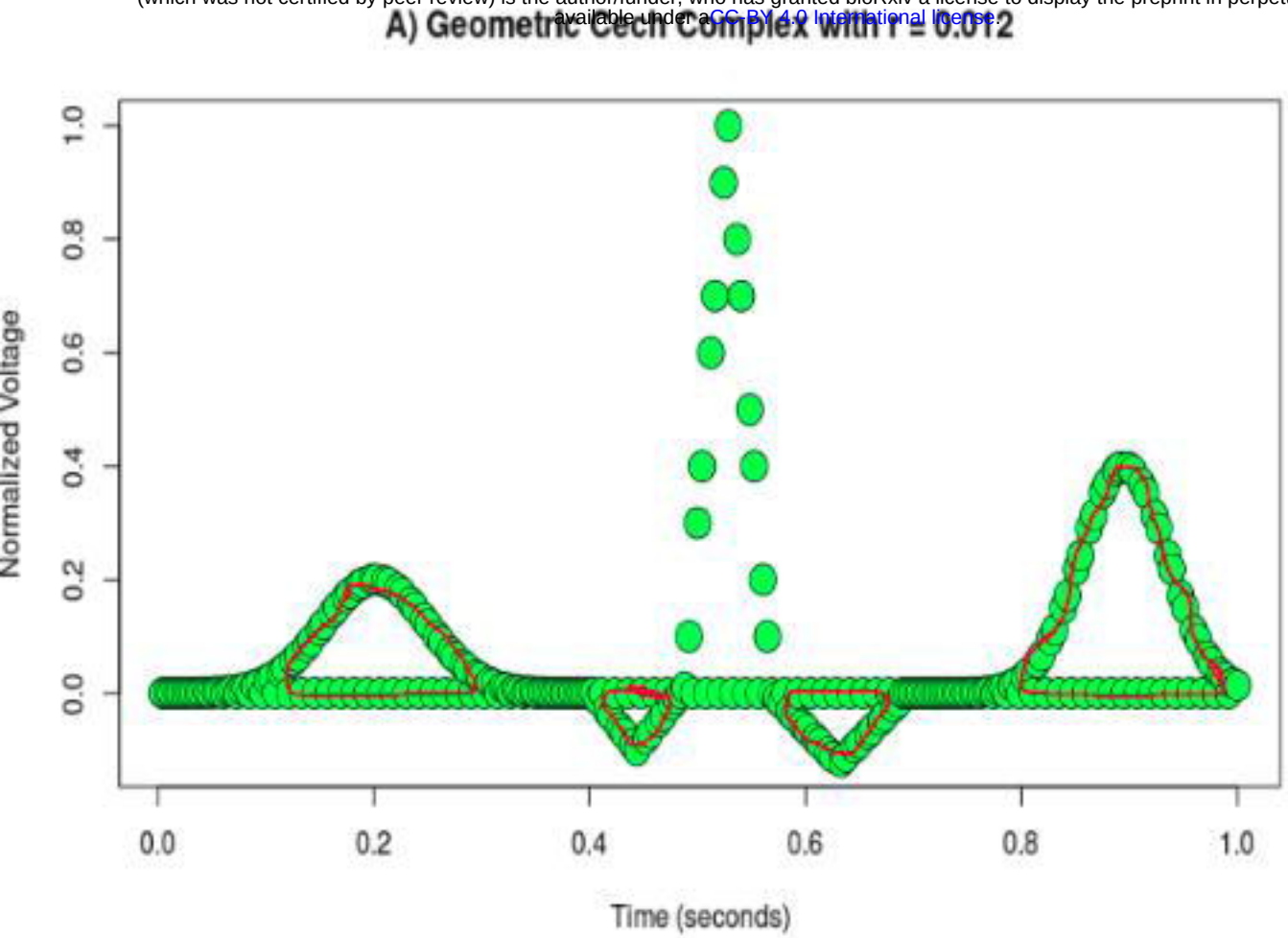
bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.478609; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

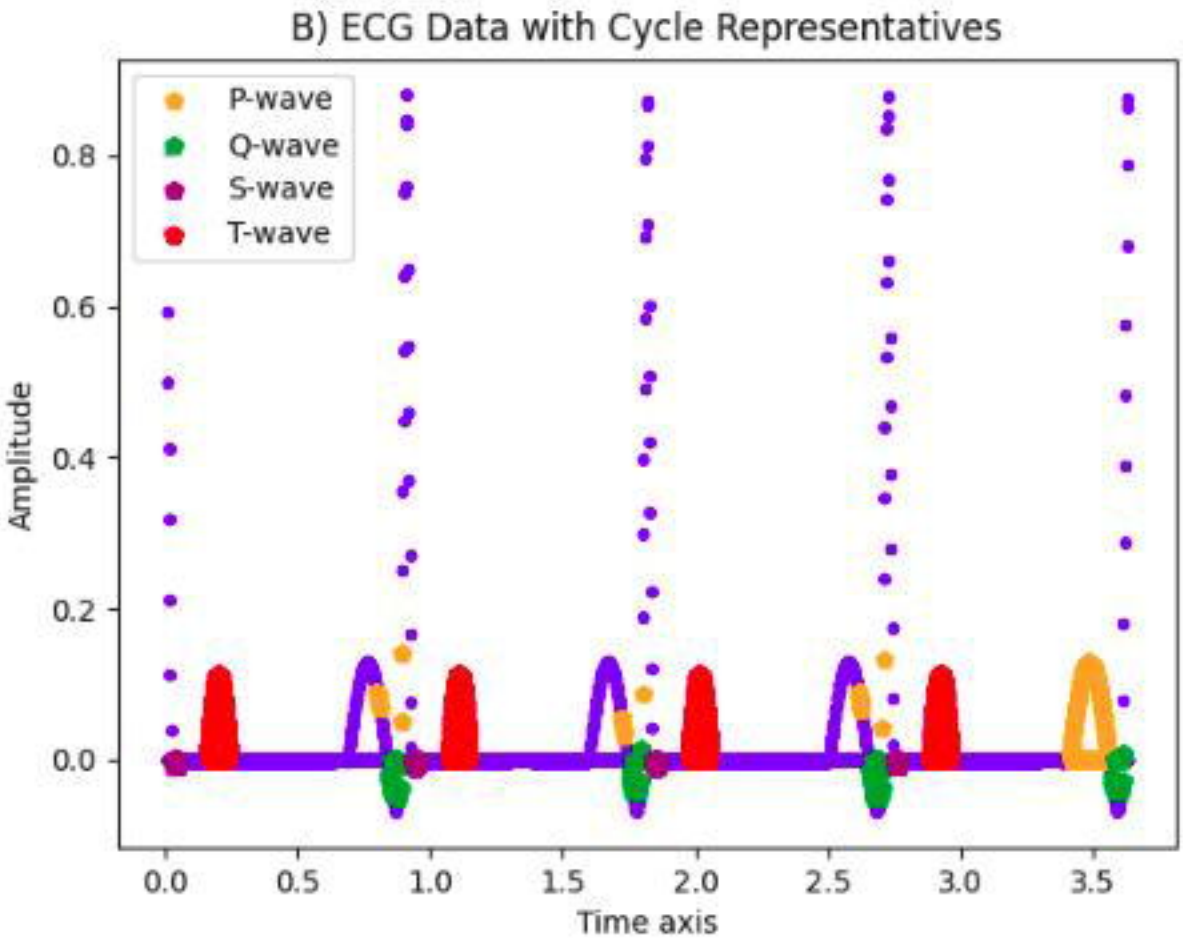
bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.479599; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.



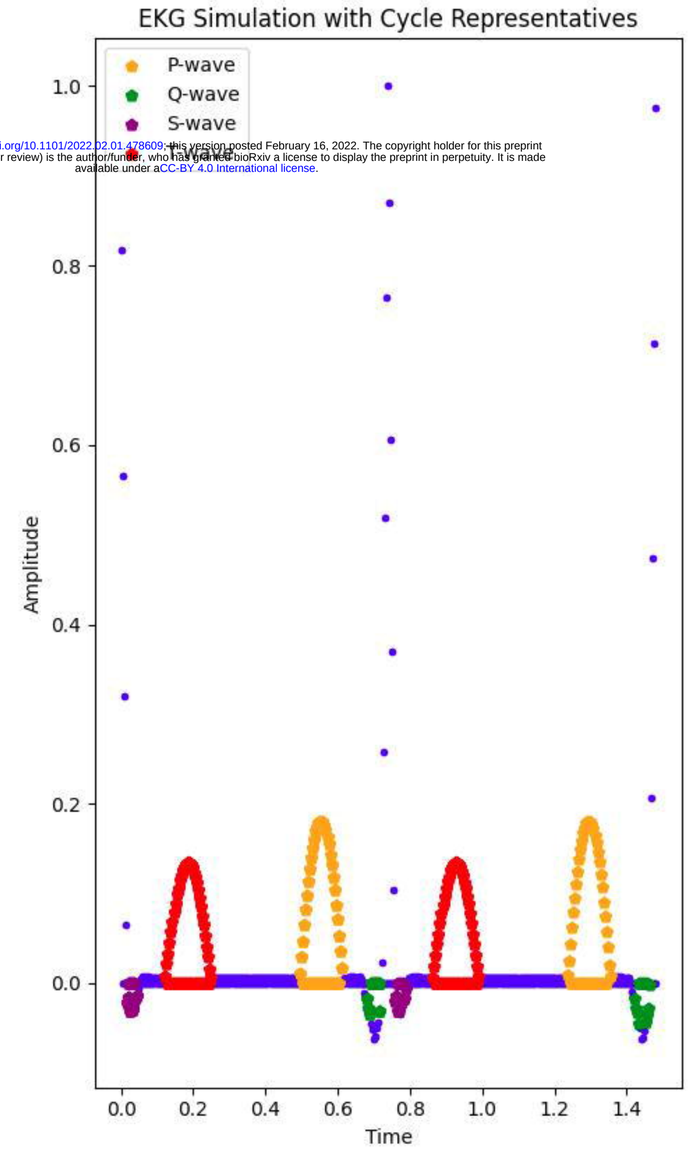

Add points
Normalize
Compute R-wave peak coordinates
Add isoelec-tric line
Trim

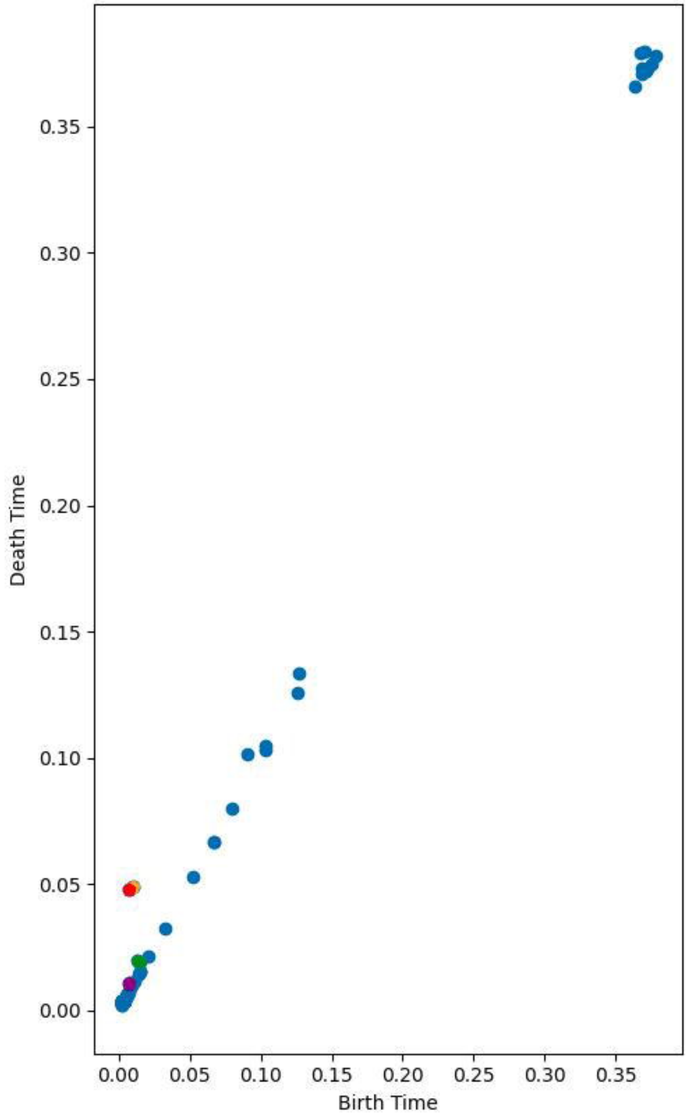
(which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuit
available under a CC-BY 4.0 International license.





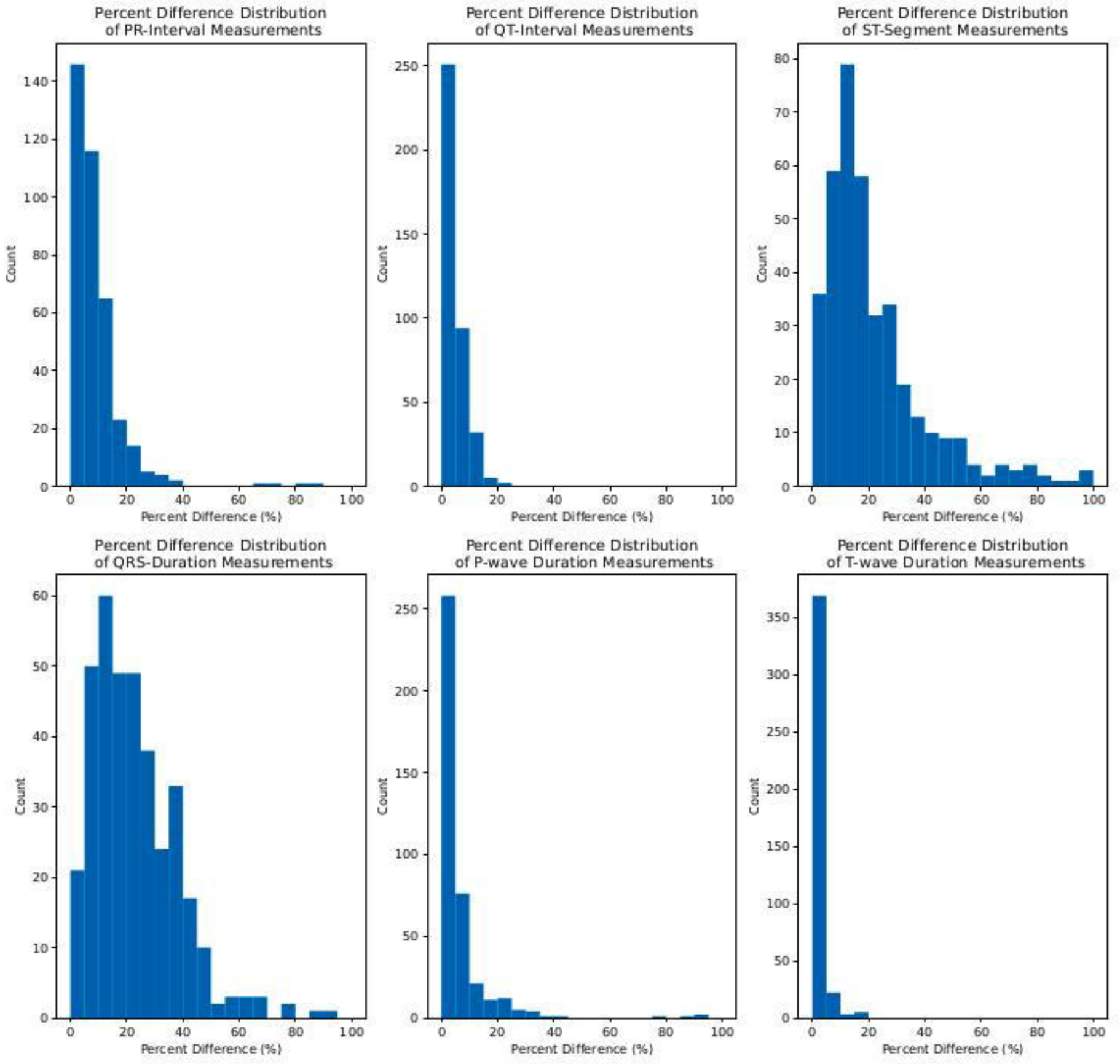
bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.478609; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

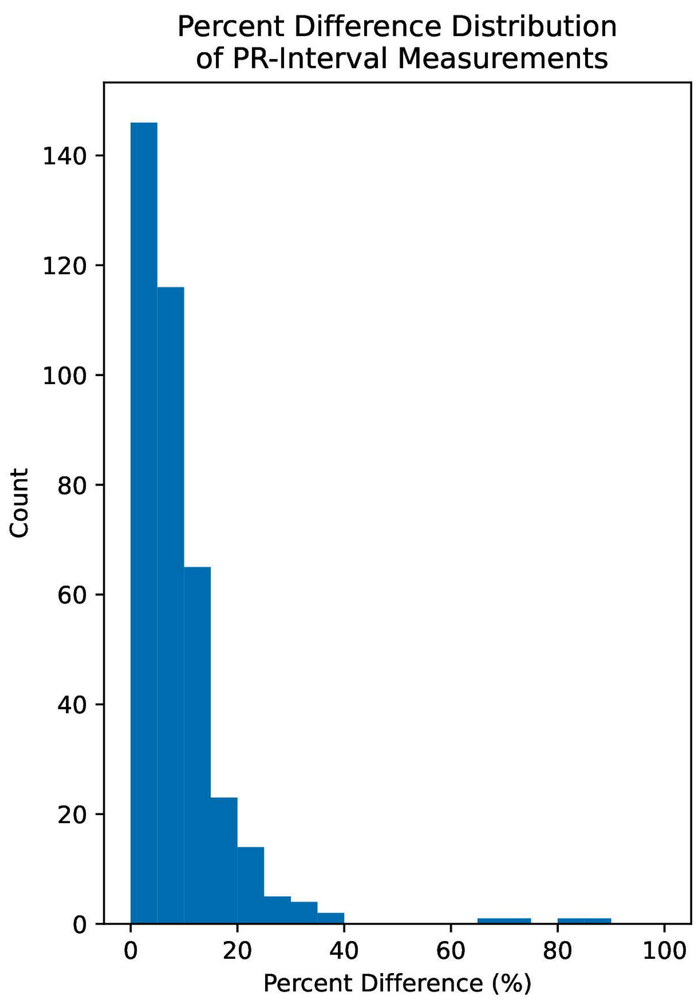


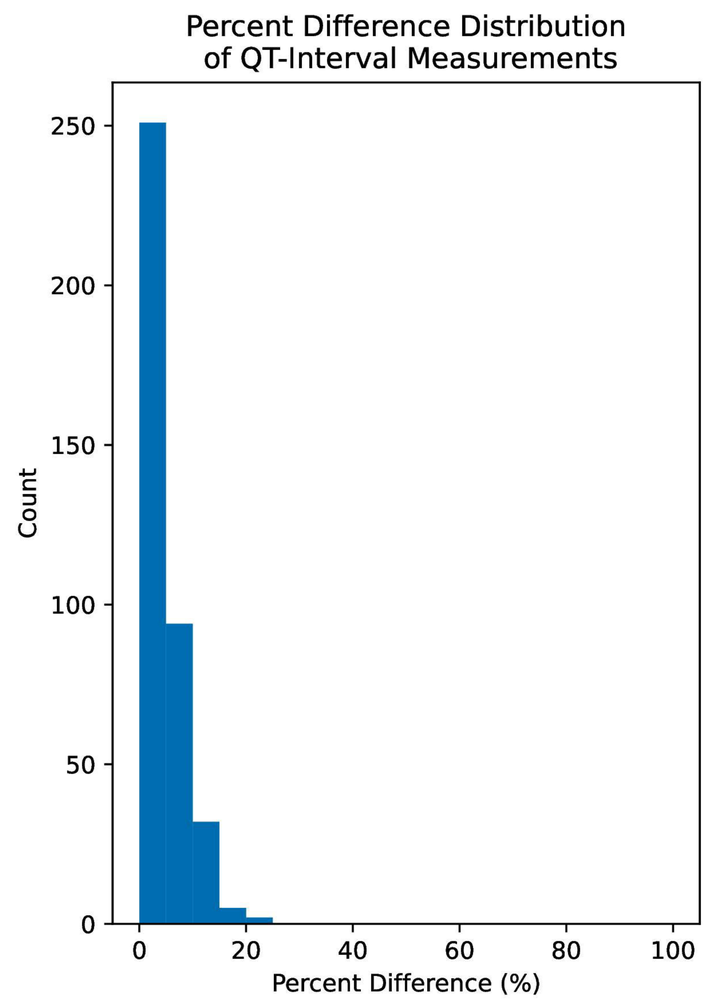


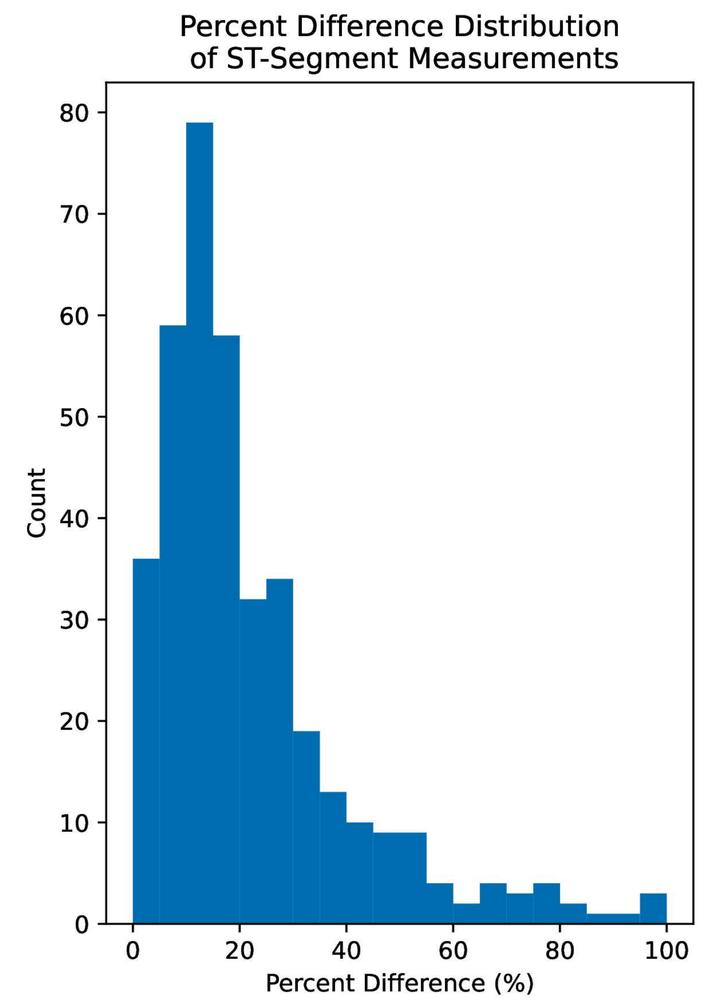
Simulation-Measured Intervals:
PR-interval: 0.173
QT-interval: 0.324
ST-segment: 0.069
QRS-duration: 0.127
P-wave duration: 0.119
T-wave duration: 0.128

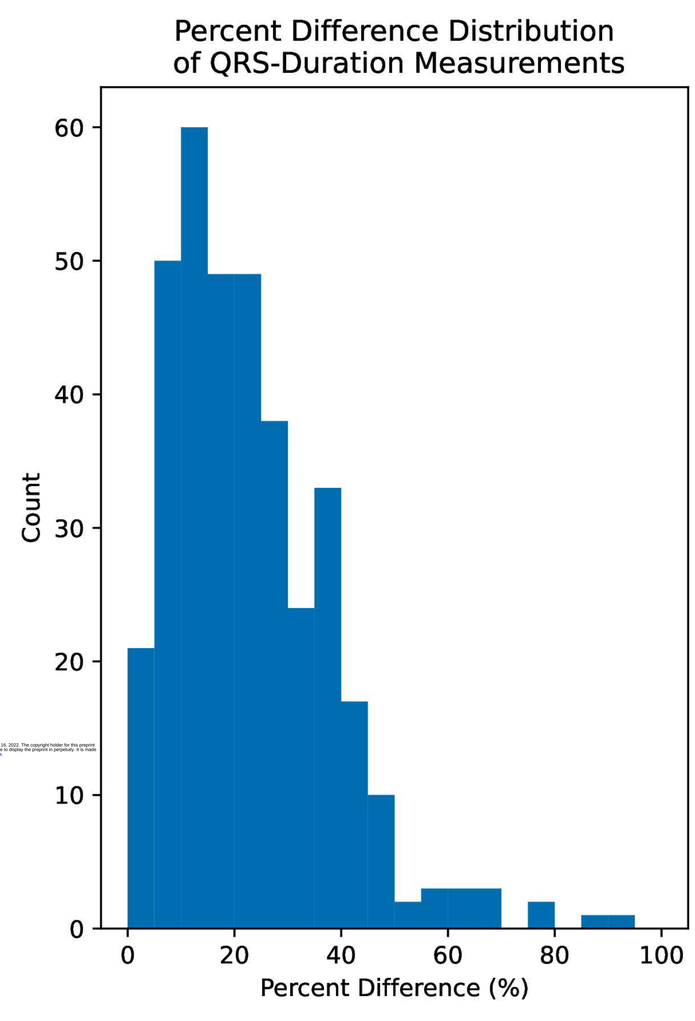
TDA-Measured Intervals:
PR-interval: 0.184
QT-interval: 0.312
ST-segment: 0.078
QRS-duration: 0.108
P-wave duration: 0.118
T-wave duration: 0.128

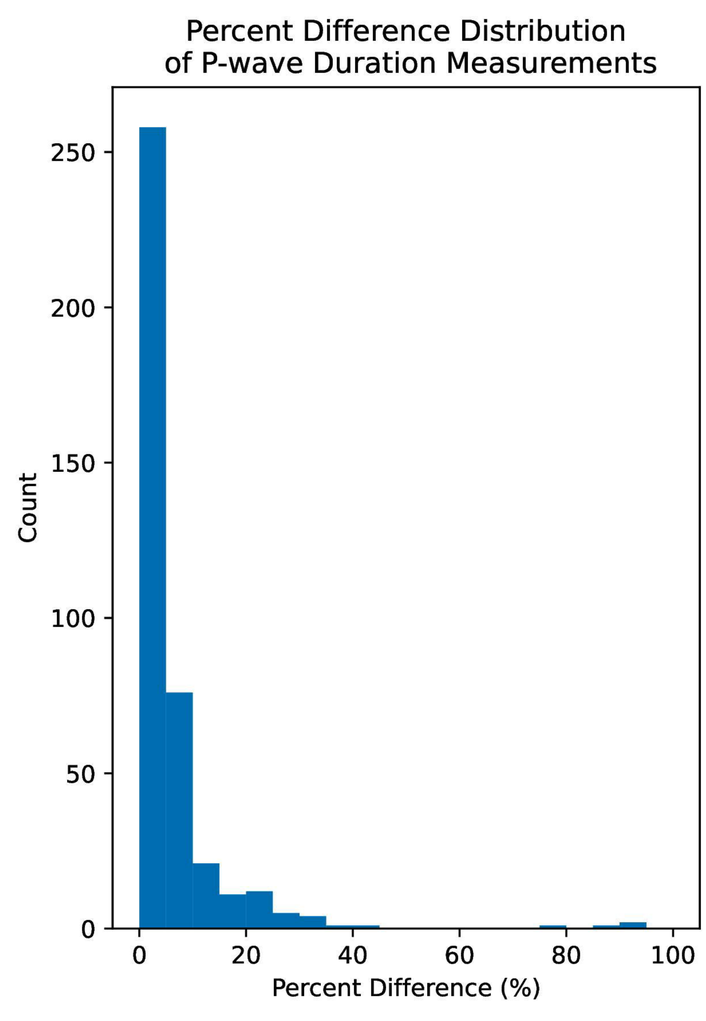

Percent Difference Distribution
of PR-Interval Measurements
Percent Difference Distribution
of QT-Interval Measurements
Percent Difference Distribution
of ST-Segment Measurements
Percent Difference Distribution
of QRS-Duration Measurements
Percent Difference Distribution
of P-wave Duration Measurements
Percent Difference Distribution
of T-wave Duration Measurements
Count
Percent Difference (%)

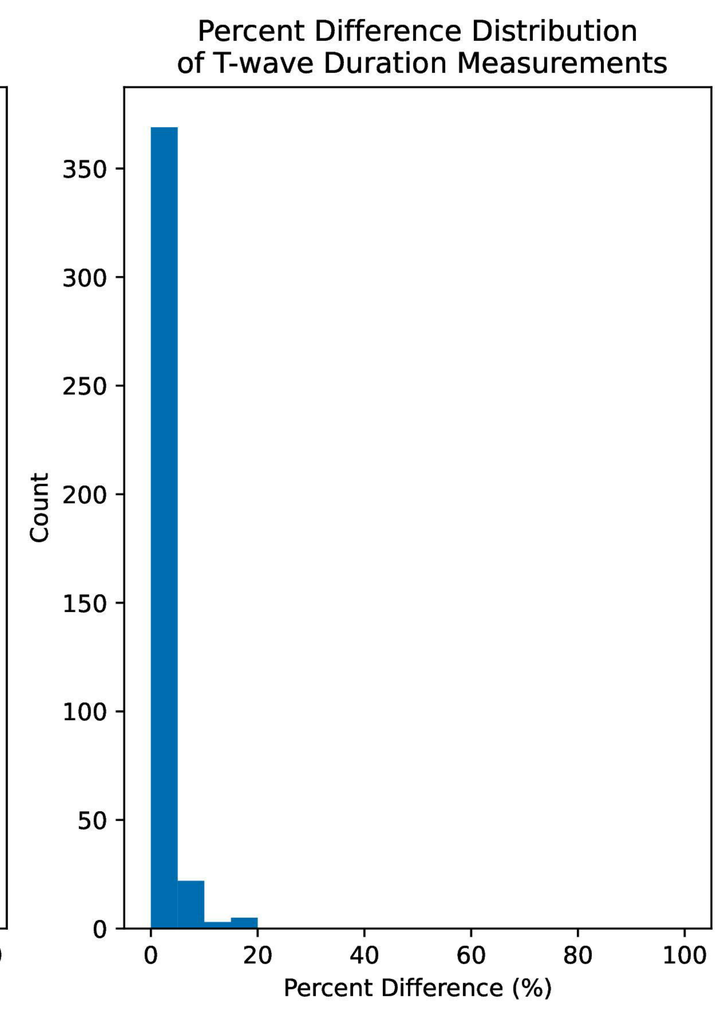












bioRxiv preprint doi: https://doi.org/10.1101/2022.02.01.479600; this version posted February 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.