Applications of Algebraic Topology to the Detection of Ventricular Tachycardia

Justin Zhang
Under the direction of Dr.Krassimir Penev
Bergen County Academies

Abstract

Algebraic topological techniques can reveal meaningful signatures in complex physiological time series data. This study applied persistent homology to electrocardiogram (ECG) data for detecting ventricular tachycardia, a form of heart arrhythmia. The ECG time series from 35 patients with ventricular tachycardia and 20 healthy controls were embedded into 3D point clouds using time delays chosen by mutual information. Vietoris-Rips filtrations generated simplicial complex sequences representing topological features at different scales. Tracking the evolution of connected components in these sequences via persistent homology allowed the identification of significant differences between the ventricular tachycardia patients and controls. Notably, the longest-lived components had substantially higher death times for the 35 ventricular tachycardia cases compared to the 20 healthy controls. These topological distinctions demonstrate the potential of harnessing algebraic topology for extracting diagnostic markers from ECG data. Further work could develop the technique into a real-time machine learning algorithm for ventricular tachycardia detection and compare performance against existing methods.

1 Introduction

Big data has become an increasingly common term. However, often, this data is large, complex, and full of noise. In order to make a valid conclusion from big data, we must have a way to extract certain features of the data without losing others. Topological Data Analysis, or TDA for short, is a novel new method to address this challenge. TDA is based on the idea that data has shape; this shape can be exploited via topological and geometrical tools. These can provide a powerful way to obtain useful qualitative information about the data. In recent years, TDA has provided mathematical and statistical methods for analyzing complex data, and many software libraries have become available for this type of analysis, including the GUDHI library in Python, R studio, and Giotto.

In this context, data refers to a collection of points along a time series. The horizontal axis of the variable will be the time coordinate, while the vertical coordinate of the point is the metric we are studying. When we have a collection of these data points with a notion of distance that satisfies the criteria for a metric space, we refer to this collection as a point cloud.

In this paper, we will use the technique of persistent homology to count many topological features of the time series embedding, including connected components, as well as other higher dimensional topological properties, such as holes, tunnels, and voids. In this paper, we will study the 0 dimensional topological structures of the time series embeddings of patients suffering from Ventricular Tachycardia and those of healthy controls. It is good practice to analyze the 0 dimensional homology of data rather than 1 dimensional homology and higher dimensional homologies because while higher dimensional homologies can occasionally yield interesting results that cannot be seen from lower dimensional homologies, it is much more difficult to do and the results are much harder to interpret. It is still not agreed upon how to interpret the Betti sequences of higher dimensional homologies, and the topological features identified from a higher dimensional analysis can vary greatly from a change in the time series embedding dimension or the time delay. As a result, the conclusions drawn from higher dimensional analysis are often questionable.

Throughout this paper, 0 dimensional topological features are referring to the connected parts between the sublevel subsets within the time series embedding. Subsequent to a piecewise linear interpolation of this time series embedding, we can begin to analyze the topological features of this 3 dimensional point cloud. As the time series embedding is thickened (aka. the sublevel subsets of the embedding increase) and previously disjoint connected components merge together and new connected parts are formed, the number of connected components either increases or decreases. When a new connected component is added, we refer to this as the *birth* of that topological feature. Similarly, when two connect components merge together and one of the components is lost, we denote this as an instance of the *death* of a connected component. To depict all of these instances of birth and death as the sublevel subsets increase, we use a depiction known as a *persistence diagram*. An alternative representation is a *persistence barcode*, which encodes the same information. In this paper, we will mainly use persistence diagrams, as this is the most standard way to depict birth and death times in the current topological literature.

The topic of analyzing the heart rate variability (HRV) from ECG/EKG data has been an important and well-studied area, and in this paper, we continue to build on this research as we introduce and examine the efficacy of a novel method of detecting Ventricular Tachycardia using persistent homology on the time series generated by ECG data. General uses of methods from topological data analysis have shown promise in numerous other recent studies. The analysis of the topology of data has been applied to crack propagation in the field of fracture mechanics, the spread of the COVID-19 epidemic in the United States, the spread of the Zika Virus in Brazil, as well as the detection of general Arrhythmia from time series data. However, previous work has been unable to detect specific properties of Arrhythmia. Arrhythmia is a broad disease, with numerous different types, and there are two properties of Arrhythmia that are crucial for a proper diagnosis: the location of the Arrhythmia as well as the relative rate of the ECG signal (whether it is going faster or slower than normal). In this paper, we study Ventricular Tachycardia, the specific form of Arrhythmia that originates in the ventricles (the lower chamber of the heart) and whose rate is faster than normal. This is important as the origin of the Arrhythmia and the classification of its abnormalities are key to properly treating the patient and protecting them from the possible consequences of leaving Ventricular Tachycardia untreated, which include cardiac arrest, a heart attack, or in severe cases, sudden death. To ensure that we properly extract topological features of the time series embedding, we will use short segments of the recordings of ECG data from [4], [5], [6].

stating the formal definitions of metric spaces, simplicial complexes, Betti numbers, and persistence/persistent homology, with a specific focus on the Cech and Vietoris Rips complexes. The Vietoris Rips complex is the key tool we will use for topological data analysis throughout this paper. Next, we provide details about the data set that we will be doing our analysis on, followed by a discussion of all of the operations that we will apply to the raw data with images of the resulting point clouds and diagrams displaying the relevant topological information We then include a copy of all of the code used in the topological data analysis for completeness and describe our results and conclusions.

2 Methods

2.1 Metric Spaces

First, we should define what a metric space is.

Definition 2.1 (Metric Space). A metric space consists of a set of points M and a distance function $d : M \times M \to \mathbb{R}_{\geq 0}$. This function d must satisfy three criteria.

    (i) d is symmetric ($\forall a, b \in M, d(a, b) = d(b, a)$)

    (ii) $\forall a, b \in M, d(a, b) \geq 0$, with equality if and only if $a = b$. (d is positive definite)

    (iii) d satisfies the triangle inequality ($\forall a, b, c \in M, d(a, b) \leq d(a, c) + d(c, b)$)

Metric spaces are important because they are a type of topological space, and thus we can use

our topological tools on them. In this paper, the metric space will be a set of points in 2D space (a point cloud), with the distance metric being 2D Euclidean distance.

2.2 Simplicial Complexes

The main tool for showing the topology behind the complex structures of an object is to use a fundamental object known as a *simplicial complex*. In this section, we define what a simplicial complex is, and in a later section we will introduce the specific complexes that we use in this paper, namely the Cech complex and the Vietoris-Rips complex, which is just a specific case of the Cech complex.

The definition of a *simplex* is a generalized "triangle" of arbitrary dimension. If we denote a subset of the vertices of a point cloud as $\{p_i\}_{i=0}^{q}$, where $\{p_i\}_{i=0}^{q} \in \mathbb{R}^n$, for a certain number of dimensions n, with q and n being positive integers with the number of points in point cloud not exceeding the number of dimensions of the topological space, then we define the notion of a q-simplex as the smallest region that contains the entire set $\{p_i\}_{i=0}^{q}$ aka. the convex hull of all the points. For the sake of this paper, assume that $\{p_i\}_{i=0}^{q}$ is a set of affinely independent points. Thus, using this definition, we can that vertices within the topological space are 0 simplices (0 dimensional homological structures), edges between vertices are 1 simplices (1 dimensional homological structures), "triangles" (three edges with the space in between them) are 2 simplices (2 dimensional homological structures), and "tetrahedra" (a figure with four triangles as faces and the space in between them) are 3 simplices (3 dimensional homological structures). We can continue into higher dimensions, where each n simplex can be defined based on lower dimensional simplices.

Let $\gamma = \text{Simp}(\{p_i\}_{i=0}^{q})$ denote the q simplex of the set $\{p_i\}_{i=0}^{q}$. Vice versa, we can define $V(\gamma) = \{p_i\}_{i=0}^{q}$. As stated previously, any n simplex can be constructed from k dimensional simplices, where $k < n$. We now study the relationship between simplices which have different dimensions. Note that $\forall x \subseteq V(\gamma)$, we have that x is affinely independent (by how we defined $\{p_i\}_{i=0}^{q}$), and the thus the convex hull of all the points in the set x form a $|x|$ simplex. x is known as a *face* of the origianl q simplex γ, more specifically, a $|x|$ face.

A simplicial complex of dimension of $\mathbb{R}^n$ is then just the group of these (finite) simplices γ in $\mathbb{R}^n$ such that $\forall \gamma$ which are in the simplicial complex, any face of γ is also a simplex in the simplicial complex, and for any pair of simplices in the simplicial complex, the intersection of the simplices is a face of both of the original simplices, and so their intersection is also in the original simplicial complex. To put this in a notational form:

Definition 2.2 (Simplicial Complex). A simplicial complex $\mathbb{S}$ in $\mathbb{R}^n$ (of dimension n) is a group of (finite) n dimensional simplices $\{\gamma\}$ such that:

(i) $\forall \gamma_i \in \mathbb{S}$, $\forall f$ such that f is a face of γ_i, $f \in \mathbb{S}$.

(ii) $\forall \gamma_i, \gamma_j \in \mathbb{S}$, the intersection of the two simplices γ_i and γ_j is a face of both γ_i and γ_j. As a consequence of (i), it also follows that the intersection of the two simplices is also a simplex of $\mathbb{S}$.

2.3 Betti Numbers

Betti numbers are used to encode data about all the topological features of a point cloud, including connected components as well as higher dimensional structures such as holes, tunnels, voids, etc. Betti numbers form the fundamental building blocks of persistent homology and are a cornerstone in topological data analysis, so an understanding of them is crucial.

As stated before, when studying simplicial complexes, we are really analyzing the relationship between simplexes which have different dimensions. *Betti numbers* are the conventional way of encoding this information. In general, the dth Betti number represents the number of d dimensional holes that an object has. Thus, the 0th dimensional Betti number, denoted β_0, counts the number of connected components, the 1th dimensional Betti number, denoted β_1 counts the number of tunnels, the 2nd dimensional Betti number, denoted β_2, counts the number of voids, and so on. The Betti number is a significant invariant in topology and can be used to differentiate between different topological spaces. Hence, we see that Betti numbers are a powerful tool to study the holes in topological objects which have different dimensions. Throughout this paper, the topological objects we are studying will typically be simplicial complexes.

Suppose we have a set of q simplices $\gamma_1, \gamma_2, \cdots, \gamma_n$, and collectively, they form a simplicial complex $\mathbb{S}$.

Definition 2.3 (q-chain)**.** Consider all operations over the integers modulo 2 (aka. $\mathbb{Z}_2$). Then given a sequence x_i such that $x_i \in \mathbb{Z}_2$, the sum $\sum_{i=1}^{n} \gamma_i x_i$ is known as a q chain.

We often consider the collection of q chains as a whole, and this set is commonly denoted in literature as $C_q(\mathbb{S})$.

In order to relate the simplicial complexes of different dimensions, we can analyze the relation between their respective chains. This relation is known as a *boundary map*.

Definition 2.4 (Boundary map)**.** The qth *boundary map*, which we will denote as ∂_q (a common notation in the literature), maps $\partial_q : C_q(\mathbb{S}) \to C_{q-1}(\mathbb{S})$, with all operations done in the integers modulo 2 ($\mathbb{Z}_2$),

$$\partial_q(\{p_0, p_1, \cdots, p_q\}) := \sum_{i=0}^{q} (\{p_0, \cdots, \hat{p_i}, \cdots, p_q\})$$

In the above expression, $\{p_0, p_1, \cdots, p_q\}$ is a simplex $\in \mathbb{S}$ and $\hat{p_i}$ means that this set does not contain the ith element, aka. the "drop out" operation. The *boundary map* is also called the *boundary homomorphism*, because $\partial_q : C_q(\mathbb{S}) \to C_{q-1}(\mathbb{S})$ is a homomorphism between the chain groups.

Note: Notice that in the above boundary map expression, the coefficients are very simple and we do not have to account for sign changes because we in $\mathbb{Z}_2$. If we were not doing all of our operations in $\mathbb{Z}_2$, then we would no longer be able to ignore the signs and the coefficients for the boundary map/homomorphism would likely be more complex, with alternating signs throughout the expression.

Now that we have defined the boundary map, let us study the toy example where we have that $\mathbb{S} = \{\{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\}$ aka. a triangle. Notice that $\partial_2\{a,b,c\} = \{a,b\} + \{b,c\} + \{a,c\}$, and is thus nontrivial. However, $\partial_1(\partial_2(\{a,b,c\})) = \partial_1(\{a,b\} + \{b,c\} + \{a,c\}) = \{a\} + \{b\} + \{c\} + \{a\} + \{b\} + \{c\} = 0$, and thus, is trivial. In fact, this phenomenon holds in general as well.

Theorem 1 (Composition of two consecutive boundary maps)**.** The composition of any two consecutive boundary homomorphisms is a trivial homomorphism aka. $\forall q, \partial q - 1 \circ \partial q = 0$

Proof. Refer to [7]. $\qquad\square$

Now that we have the composition of consecutive boundary maps is trivial, we can define a nested mapping between chains.

Definition 2.5 (Chain Complex)**.** The nested mapping

$$0 \xrightarrow{\partial_{q+1}} C_q(\mathbb{S}) \xrightarrow{\partial_q} C_{q-1}(\mathbb{S}) \xrightarrow{\partial_{q-1}} \cdots \xrightarrow{\partial_2} C_1(\mathbb{S}) \xrightarrow{\partial_1} C_0(\mathbb{S}) \xrightarrow{\partial_0} 0$$

is known as the *chain complex.*

Let us denote $G_q(\mathbb{S}) := \ker(\partial_q)$ and $B_q(\mathbb{S}) := \operatorname{im}(\partial_{q+1})$, where ker is the kernel of the boundary homomorphism and im is image of the homomorphism. Now we can define the qth homology.

Definition 2.6 (qth Homology)**.** Note that all the $B_q(\mathbb{S}) := \operatorname{im}(\partial_{q+1})$, are subspaces of their respective $G_q(\mathbb{S}) := \ker(\partial_q)$, and so we can define, using the standard notation in literature,

$$H_q(\mathbb{S}) := \frac{\ker(\partial_q)}{\operatorname{im}(\partial_{q+1}}) = \frac{G_q(\mathbb{S})}{B_q(\mathbb{S})}.$$

The $H_q(\mathbb{S})$ quotient space is known as the qth homology.

Using all of these, we can define the Betti number as the dimension of this qth homology.

Definition 2.7 (Betti Number)**.** Let $\beta_q(\mathbb{S})$ denote the qth Betti number. Then

$$\beta_q(\mathbb{S}) = \dim(H_q(\mathbb{S})).$$

Note that $H_q(\mathbb{S})$ is sometimes also known as the qth homology group, and qth Betti number β_q is then just $\operatorname{rank} H_q$.

Using Betti numbers, we can compute the number of q dimensional holes of a topological object, and this will be the core building block of persistent homology, the main technique we will use in our topological data analysis.

2.4 Persistence Homology and Persistence Diagrams

Persistence Homology/Persistent Homology is in essence an extension of normal homology. The advantage of using persistent homology as compared to homology is that persistent homology helps to get rid of the "noise" data in a point cloud. In order to get rid of this noise in persistent homology, we apply a technique known as a *filtration*. This can be visualized as a "thickening" of the point cloud, where a sequence of simplicial complexes are generated by this filtration, each one with a larger "thickening" than the last. Through the sequence of simplicial complexes, we can observe the changes in the topological features of the point cloud and the birth and death of various holes of specified dimensions through the filtration. Throughout this section, we will use notation consistent with that in [8].

Definition 2.8. [8], Filtration

For an index set I, a filtration is a sequence of simplicial complexes, $\{\mathbb{K}_t\}_{t \in I}$, satisfying

$$\mathbb{K}_{t_1} \subseteq \mathbb{K}_{t_2}$$

whenever $t_1 \leq t_2$.

Intuitively, the above definition should make sense. After a set of connected components are merged, they should not separate again later on in the sequence of the filtration.

In the previous section, we showed how given a simplicial complex, we could find the homology group and Betti number of it. In the sequence of simplicial complexes defined from a filtration, one can observe from the previous definition that every previous simplicial complex is contained in the current one. As a result, we can think of these simplicial complexes as being "nested" within the subsequent simplicial complexes, and thus, we can observe the alterations that occur in the Betti numbers and the homology groups of the simplicial complexes in the sequence from the filtration. To see this, $\forall i \leq j$, we can construct a homomorphism $f_p^{i,j} : H_p(\mathbb{K}_i) \to \mathbb{H}_p(\mathbb{K}_j)$ for all dimensions p. The filtration can then be thought of as chain of homology groups of each of the respective simplicial complexes as follows:

$$0 = H_p(\mathbb{K}_0) \to H_p(\mathbb{K}_1) \to \cdots \to H_p(K_n).$$

As we move through the chain, some holes are lost while at other points, some holes are gained.

We now formally define the notion of a persistent homology.

Definition 2.9. [8], Persistent Homology/Persistent Homology Groups

Let $\{\mathbb{K}_i\}_{i=0}^n$ be a filtration of simplicial complexes. For $q \in \mathbb{Z}_{\geq 0}$ and $i, j \in \mathbb{Z}_{\geq 0}$ with $i \leq j$, we define the persistent homology (group) as $H_q^{i,j} := \frac{Z_q(\mathbb{K}_i)}{B_q(\mathbb{K}_j) \cap Z_q(\mathbb{K}_i)}$. The persistent homology group consists of all the homology classes (aka. holes) of K_i that are still alive at K_j.

Because the simplicial complexes in sequence of filtrations are nested, the q chains of these simplicial complexes are also nested, and so the denominator of the persistent homology is a valid subspace, and it is well-defined with respect to $Z_q(\mathbb{K}_i)$.

Definition 2.10 ([8]. ,[9], Birth and Death] Let $\{K_i\}_{i=0}^n$ be a filtration of simplicial complexes and let γ be a pth hole in $H_p(\mathbb{K}_i)$. We say that γ is *born* at K_i if $\gamma \notin H_p^{i-1,i}$. Furthermore, if γ is born at $\mathbb{K}_i$ then it *dies* entering $\mathbb{K}_j$ if it merges with another hole as we go from $\mathbb{K}_{j-1}$ to $\mathbb{K}_j$, that is $f_p^{i,j-1}(\gamma) \notin H_p^{i-1,j-1}$ but $f_p^{i,j}(\gamma) \in H_p^{i-1,j}$.

We can depict the persistent Betti numbers (the ranks of the persistence groups) by constructing a point cloud in two dimensions known as a *persistence diagram*. Let $\mu_p^{i,j}$ be the number of p dimensional holes which were born at K_i in the filtration and died at K_j, then we have $\mu_p^{i,j} = (\beta_p^{i,j-1} - \beta_p^{i,j}) - (\beta_p^{i-1,j-1} - \beta_p^{i-1,j}) \, \forall i < j, p$. To see that this is the case, just note that the first term represents the number of holes that were born at or before K_i in the sequence and that died when entering K_j, and the second term is the number that are born at before K_{i-1} in the sequence and that died when entering K_j.

To construct the pth persistence diagram, which depicts all the birth and death times of p-dimensional holes in the structures, we construct all points (a_i, a_j) with a multiplicity of $\mu_p^{i,j}$, and we can denote this persistence diagram as $\mathrm{Dgm}_p(f)$. If we define the *persistence* of a point to be $a_j - a_i$, then the vertical distance to the diagonal of the persistence diagram is equivalent to the value of the persistence, and as is usual convention, we add in all the points on the diagonal in the persistence diagram as well. A persistence diagram is useful because it lets us quickly compute persistent Betti numbers. To compute $\beta_p^{k,l}$, we just compute the number of points in the upper left quadrant with corner point (a_k, a_l).

We state the following theorem without proof.

Theorem 2 ([8].], Fundamental Theorem of Persistent Homology] Let $\phi = \mathbb{K}_0 \subseteq \mathbb{K}_1 \subseteq\subseteq \cdots \subseteq K_n = K$ be a filtration. For every pair of indices $0 \le k \le l \le n$ and every dimension of p, the p-th persistent Betti numbers is $\beta_p^{k,l} = \sum_{i \le k} \sum_{j > l} \mu_p^{i,j}$.

This theorem establishes the persistence diagram as a viable method of studying the topological structures of a point cloud. Specifically, it tell us that the persistence diagram contains all the information about the persistence homology of a filtration, and so by studying the persistence diagram, we are able to study the topology of the entire point cloud as a whole.

2.5   The Cech and Vietoris Rips Complexes

In this paper, the simplicial complex that we will use is the Vietoris Rips Complex and the filtration will be the Rips filtration. In order to apply the Vietoris Rips complex to a time series, we must first use Taken's embedding to embed the time series in a higher dimensional point cloud.

Theorem 3. Taken's Embedding Theorem Denote a time series by $\{x_t, t \in T\}$. Choose a time delay of ρ and an embedding dimension of size d. Then the time series embedding will become a set of d dimensional vectors, specifically, if we represent the time series as a function $f : A \to B$ and denote the embedding by T, then $T(f) = \{[f(t), f(t - \rho), f(t - 2\rho), \cdots, f(t - (d-1)\rho)]^T | t \in A\}$.

There is no universally agreed upon method to choosing the parameters ρ and d. In this paper, we use the concept of mutual information to determine ρ and set $d = 3$ for the ease of computation and visualization.

Theorem 4 ([10]. , Mutual information] To determine the best ρ we first calculate the maximum $x_{\max}$ and $x_{\min}$ values of the time series, and divide the interval $[x_{\min}, x_{\max}]]$ into a large number of bins. We let p_k be the probability that an element of the time series is in the kth bin and $p_{j,k}$ be the probability that x_i is in the jth bin while $x_{i+\rho}$ is in the kth bin. Then the mutual information is defined as $I(\rho) = -\sum_{j=1}^{n_{\text{bins}}} \sum_{k=1}^{n_{\text{bins}}} p_{j,k}(\rho) log \frac{p_{j,k}(\rho)}{p_j p_k}$. The first minimum of $I(\rho)$ gives the optimal time delay since there we get the most information by adding $x_{i+\rho}$.

Once we have embedded the time series into a point cloud, we construct simplices between points that are within a certain distance, where distance in this paper is defined by the Euclidean metric. For a formal definition, we have

Definition 2.11. [11], Vietoris Rip's Complex Let $X = \{x_1, x_2, \cdots, x_N\} \in \mathbb{R}^p$ be a point cloud and take some $\epsilon > 0$. The *Vietoris Rips complex* is a collection of all simplexes σ with vertices in X with $\text{diam}(\sigma) \leq 2\epsilon$, where $\text{diam}(\sigma)$ denotes the diameter of σ. $VR(X, \epsilon) := \cup_{q=0}^{p} \{q - \text{simplex}(\sigma) | \text{diam}(\sigma) \leq 2\epsilon, V(\sigma) \subseteq X\}$

It should be intuitive that for an increasing sequence $\epsilon_1, \epsilon_2, \cdots, \epsilon_{N-1}, \epsilon_N$, the Vietoris Rips complexes construct a filtration, as if two vertices are within ϵ_i of each other, then they are obviously within ϵ_j of each for any $j > i$ because $\epsilon_j > \epsilon_i$.
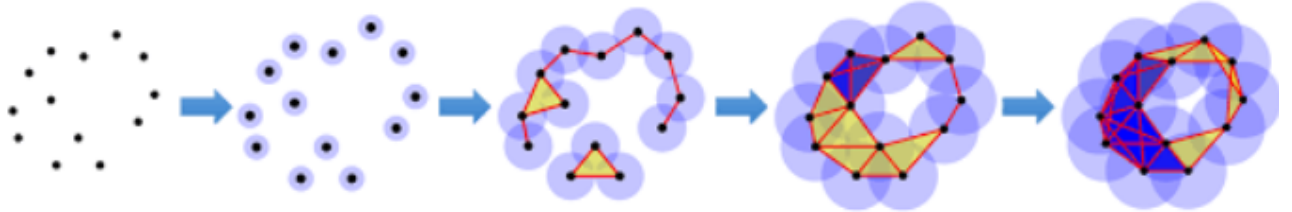

Figure 1: An application of persistent homology to a point cloud.

To demonstrate this method of a Rips filtration, we show here a toy example. As exhibited in Figure 1, it seems that the points lie in the shape of an annulus. To see whether this is true, we apply persistent homology using the Vietoris Rips complex to the given point cloud. Let us encode ϵ as the radius of the thickening of the point cloud. As we increase ϵ, we can see certain topological features, such as holes forming, which we denote as an instance of birth, and also other features disappearing when two components connect together, which we denote as death. Also note of the edges, triangles, and higher dimensional shapes that are forming, which we call simplicies. We connect two points, denote them as a and b, with an edge if and only if $d(a, b) <= \epsilon$ where $d(x, y)$ denotes the distance metric applied to x and y. Building off of this definition, we construct a triangle if and

only if all the sides of the triangle are edges, and we build a tetrahedron if and only if all of the faces of this tetrahedron are triangles, and so on and so forth. These "generalizations" of triangles are the simplicial complexes that we have discussed.

3   Information on Datasets

The data that we will be using come from the datasets "CU Ventricular Tachyarrhythmia Database" and "Combined measurement of ECG, Breathing and Seismocardiograms" available at the PhysioNet database.

The dataset "CU Ventricular Tachyarrhythmia Database" was published on May 2, 2007 and contains 35 eight minute long ECG recordings of patients who were experiencing episodes of sustained Ventricular Tachycardia. The recordings were originally collected by Floyd M. Nolle at the Creighton University Cardiac Center. All signals were passed through an active second-order Bessel low-pass filter with a cutoff of 70 Hz, and were digitized at 250 Hz with 12-bit resolution over a 10 V range (10 mV nominal relative to the unamplified signals). Each record contains 127,232 samples (slightly less than 8.5 minutes). [4]

The dataset "Combined measurement of ECG, Breathing and Seismocardiograms" was published on December 12, 2014. The ECG data was collected from 20 presumed healthy volunteers. Subsequent to attaching the ECG sensors to the subjects, the basal state of the subjects was recorded by measuring their ECG data for 5 minutes. The data was acquired using a Biopac MP36 data acquisition system (Santa Barbara, CA, USA). [6]

4   Application to Ventricular Tachycardia

The original time series data from the sick patients is shown on the following page. Note that all of this data analysis is done on the first lead of the ECG signal.
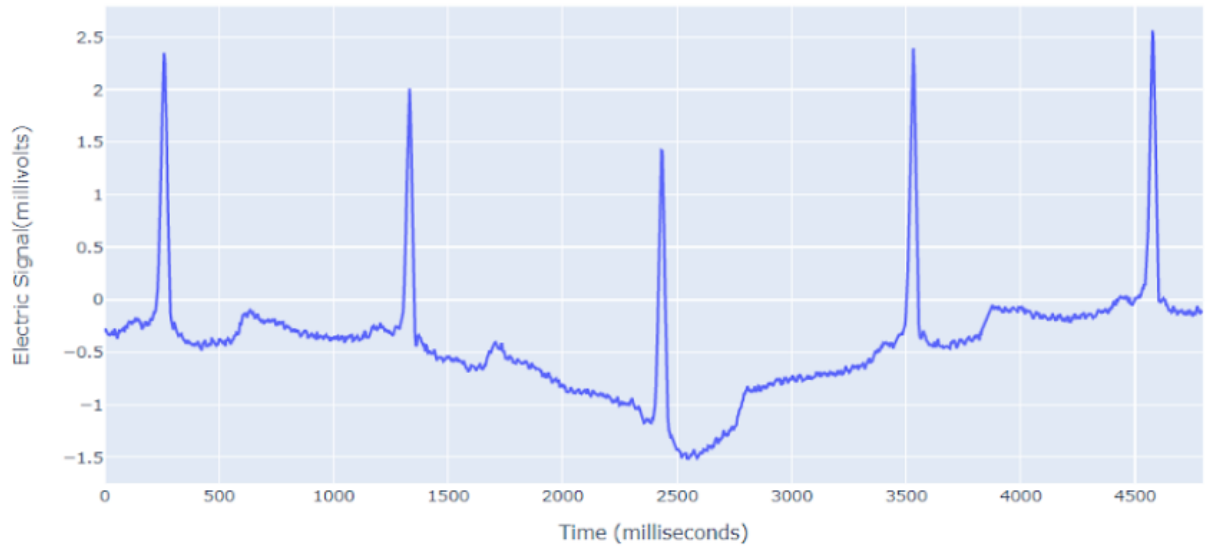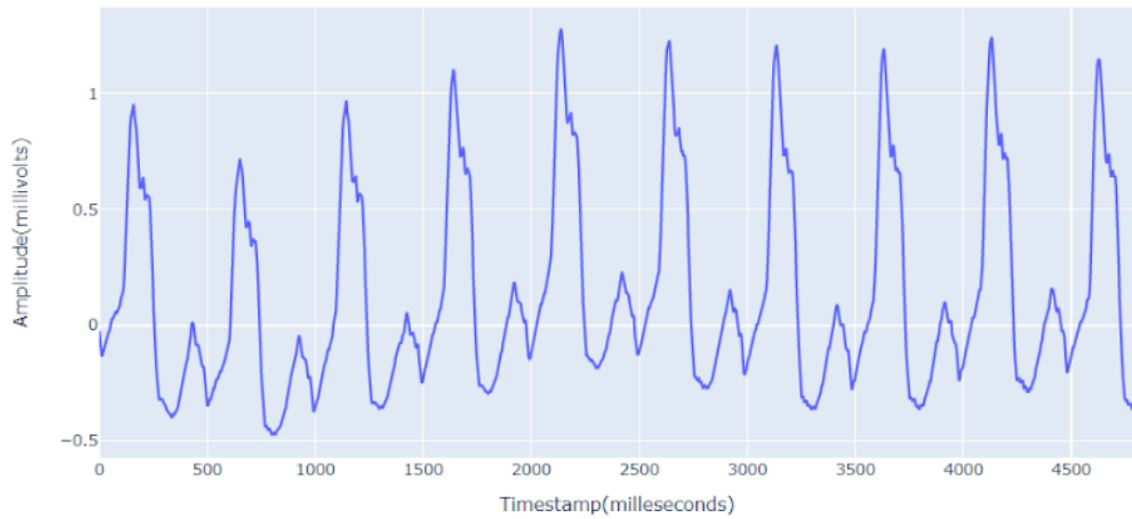
Figure 2: Time Series Data of the 1st patient

Figure 3: Time Series Data of the 2nd patient

We then apply Taken's embedding theorem, using the embedding dimension $d = 3$ and determining ρ using the method of mutual information. Below, we show the embedding for both patients.
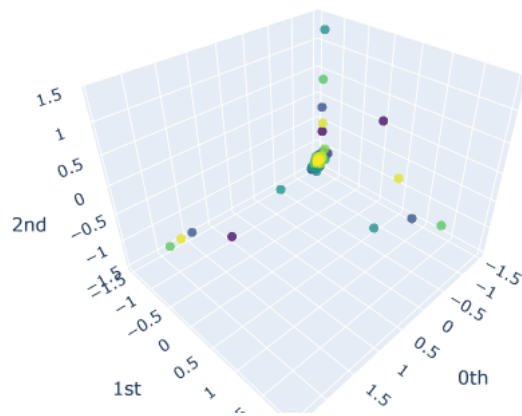
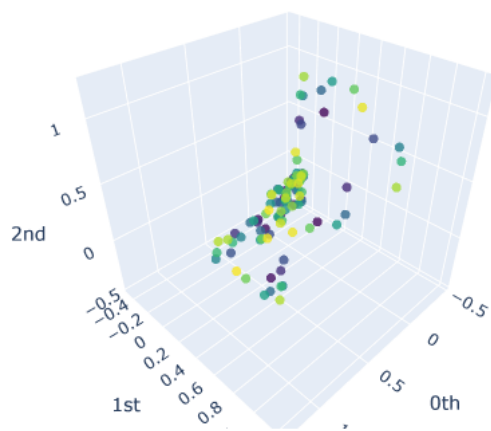Figure 4: Time Series Embedding for the 1st Patient

Figure 5: Time Series Embedding for the 2nd patient

Now using the Vietoris Rips Complex and the Rips filtration, we generate the persistence diagrams for the patients with Ventricular Tachycardia.
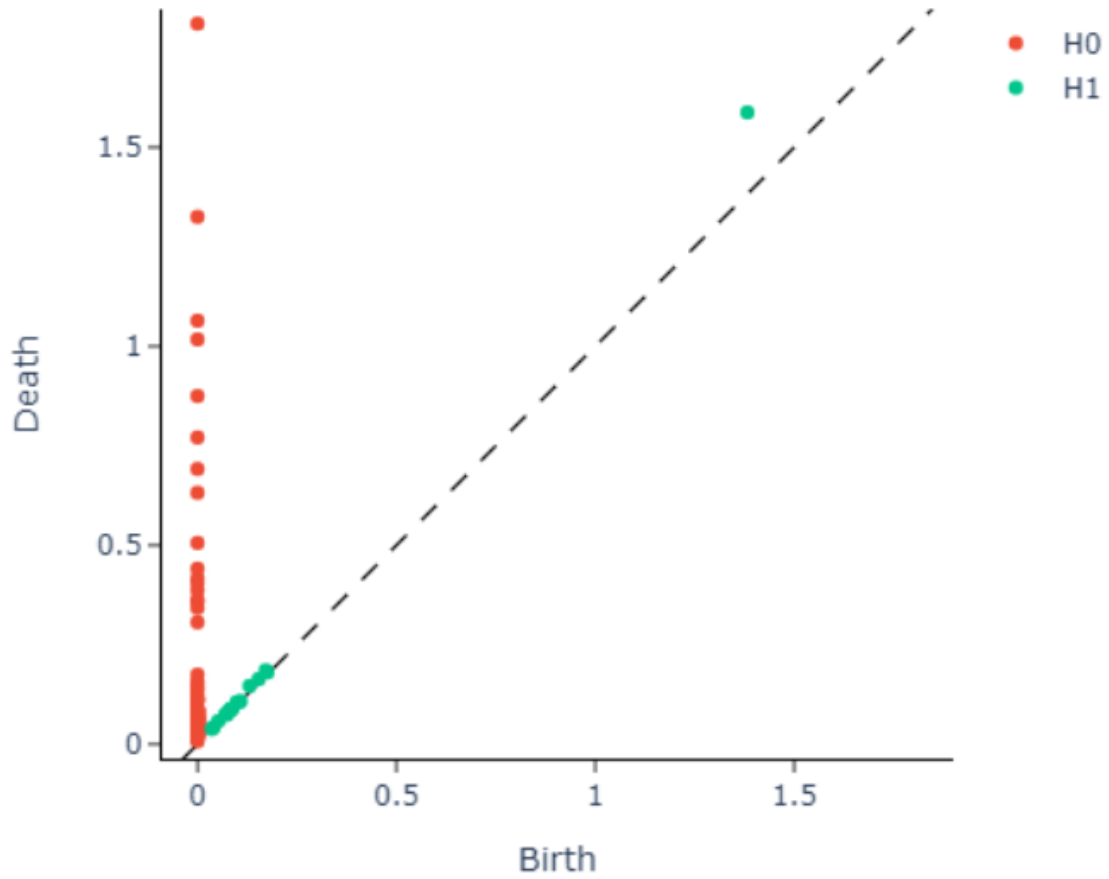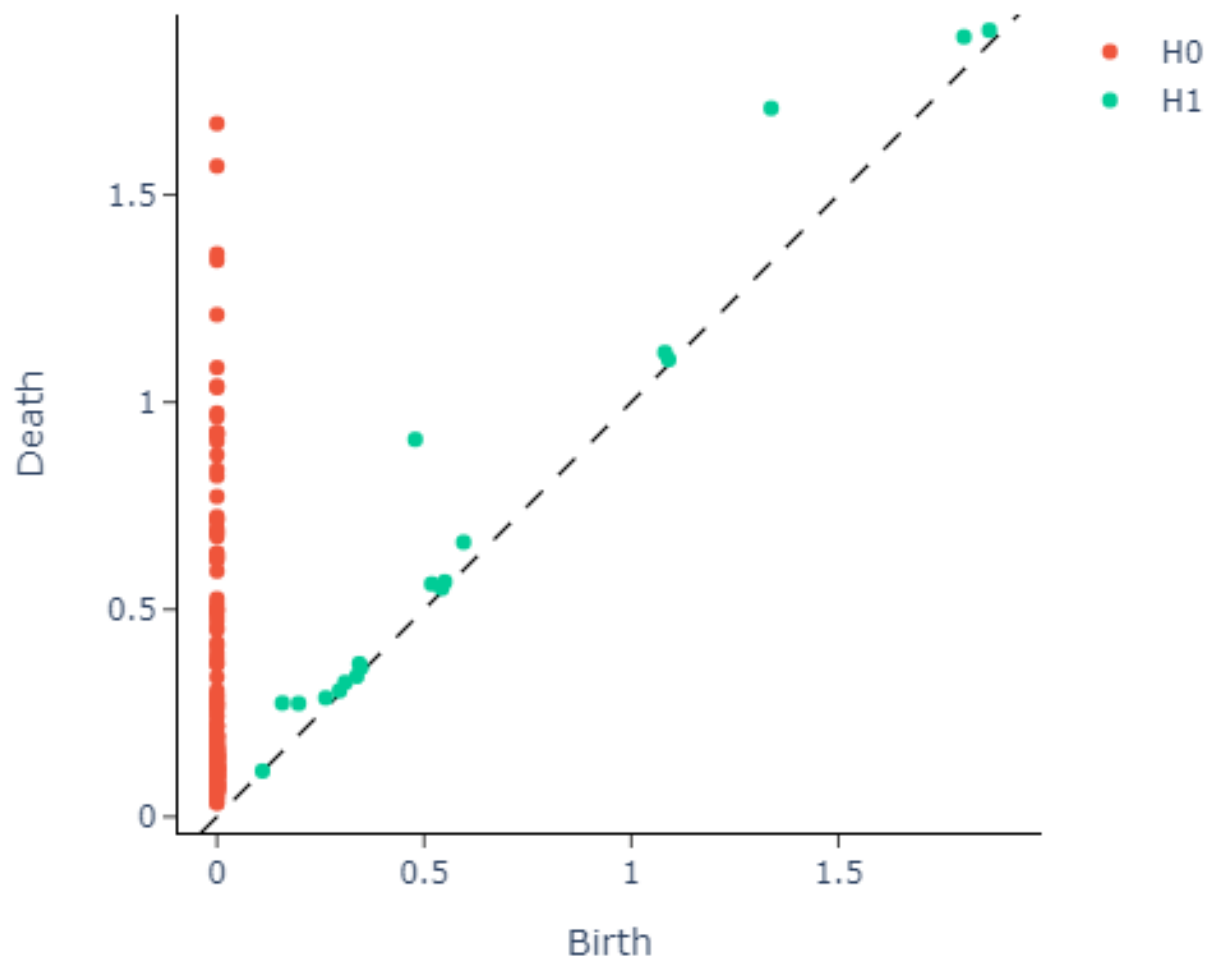

Figure 6: Persistence Diagram for the 1st patient

Figure 7: Persistence Diagram for the 2nd patient

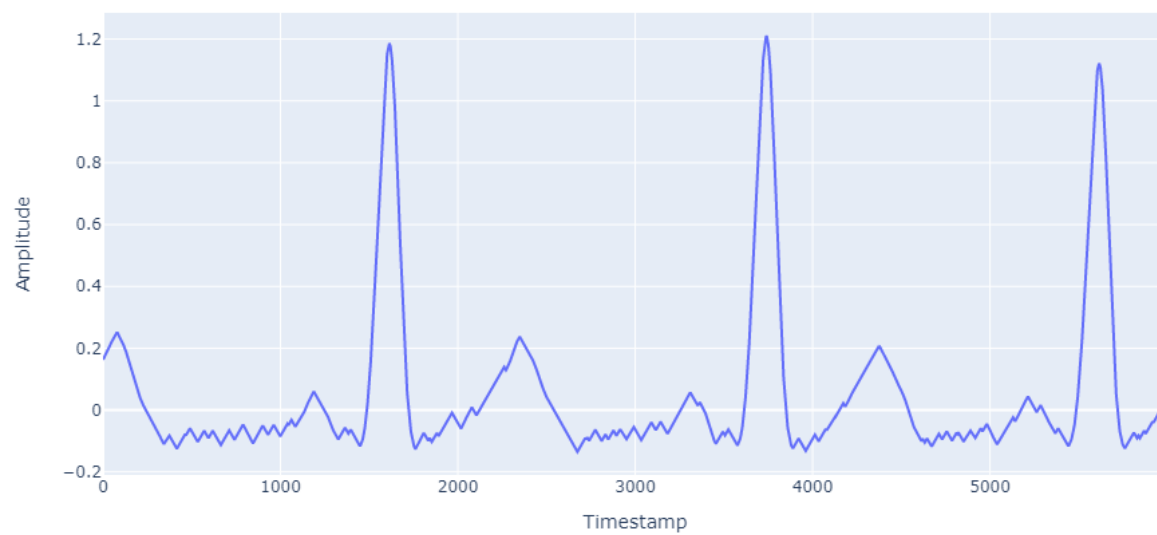We do the same procedure as above for a healthy control.

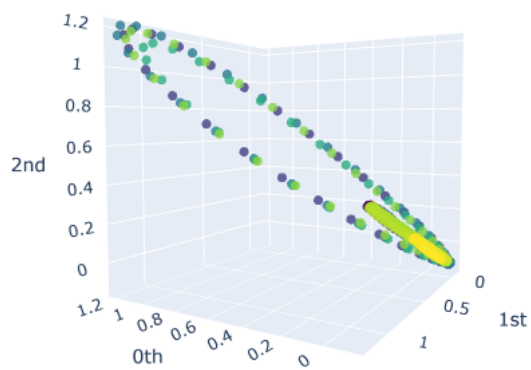Figure 8: Time Series Embedding for the healthy control

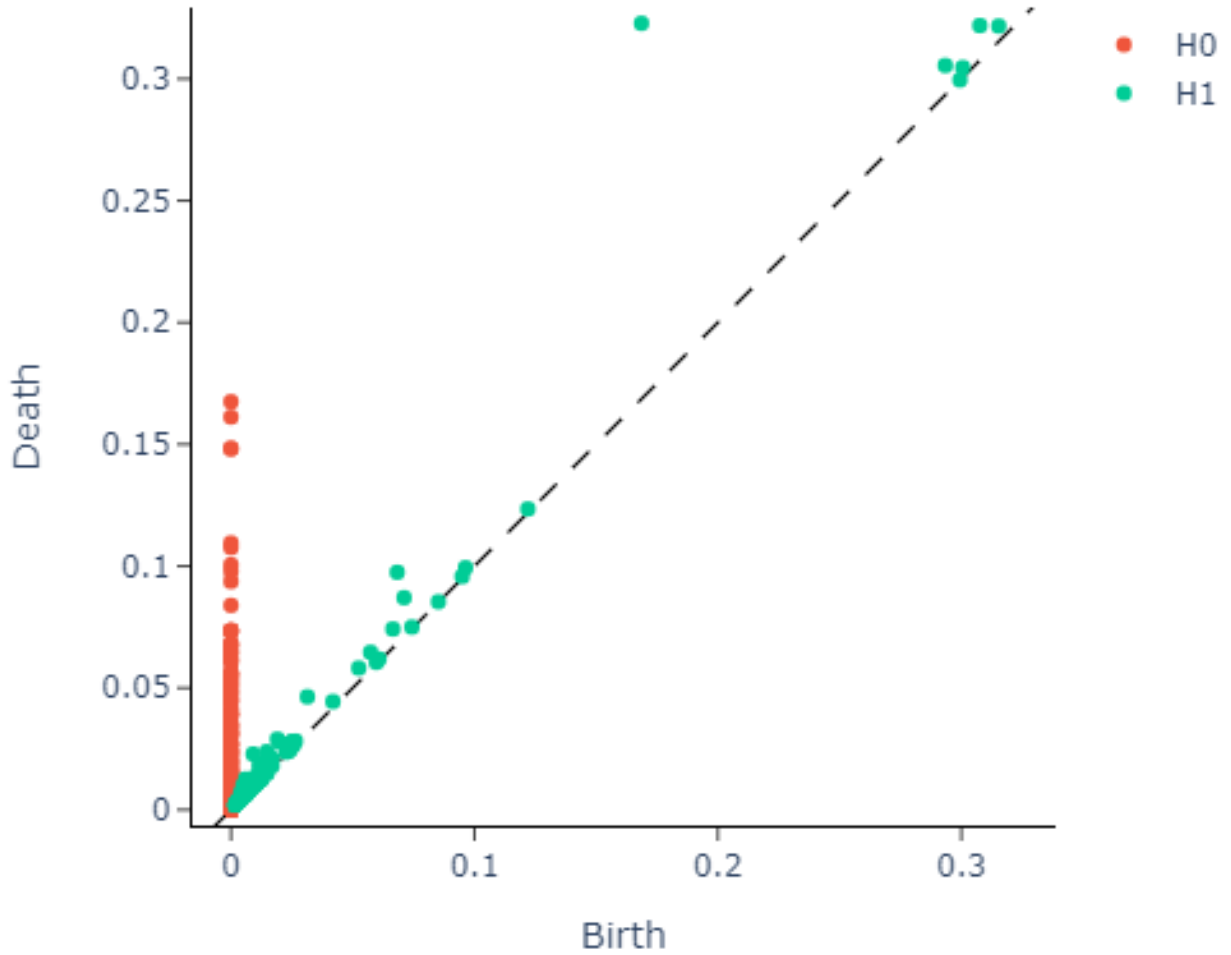Figure 9: Time Series Embedding for the healthy control

Figure 10: Time Series Embedding for the healthy control

5   Code

For completeness, we provide here the code that was used to construct the time series embedding and persistence diagram for the first patient with Ventricular Tachycardia.

Below is the code used to generate the Taken's Embedding into three dimensional topological space for the first patient with Ventricular Tachycardia.

```python
from gtda.time_series import SingleTakensEmbedding

embedding_dimension_periodic = 3
embedding_time_delay_periodic = 8
stride = 10

embedder_periodic = SingleTakensEmbedding(
    parameters_type="fixed",
    n_jobs=2,
    time_delay=embedding_time_delay_periodic,
    dimension=embedding_dimension_periodic,
    stride=stride,
)

y_periodic_embedded = embedder_periodic.fit_transform(y_list1)
print(f"Shape of embedded time series: {y_periodic_embedded.shape}")


from gtda.plotting import plot_point_cloud

plot_point_cloud(y_periodic_embedded)
```

Below is the code used to generate the persistence diagram for the first patient with Ventricular Tachycardia.

```
y_periodic_embedded = y_periodic_embedded[None, :, :]
from gtda.homology import VietorisRipsPersistence

# 0 - connected components, 1 - loops, 2 - voids
homology_dimensions = [0, 1, 2]

periodic_persistence = VietorisRipsPersistence(
    homology_dimensions=homology_dimensions, n_jobs=6
)
print("Persistence diagram for periodic signal")
periodic_persistence.fit_transform_plot(y_periodic_embedded)

nonperiodic_persistence = VietorisRipsPersistence(
    homology_dimensions=homology_dimensions, n_jobs=6
)
print("Persistence diagram for nonperiodic signal")
nonperiodic_persistence.fit_transform_plot(y_nonperiodic_embedded);
```

6   Results and Conclusions

As can be seen from the persistence diagrams of the healthy patient and the sick patients, one key distinction between the two is that the latest death time of a connect component of patients with Ventricular Tachycardia is much greater than the latest death time of a connected component of a healthy patient. The latest death time for the first sick patient was around 1.9 in the persistence diagram for the first patient and around 1.5 in the persistence diagram for the second patient. These are both significantly higher than the latest death of a component for the healthy patient, which around 0.18. Furthermore, the claim that the one dimensional homology of point clouds is not particularly useful for distinguishing them due to their large dependence on the parameters use in the time series embedding is supported here. There does not appear to be any significant correlation between the one dimensional topological features of the sick patients or the healthy patients.

7   Future Work

Throughout this project, the efficacy of persistent homology has been shown through how the topological features of data can be exploited to demonstrate real world phenomena. Future work could revolve around the construction of a machine learning algorithm that would be able to identify Ventricular Tachycardia given a recording of ECG data from a live patient using the methods described in this paper. In addition, future work could explore the higher dimensional holes and homology groups, as these higher dimensional holes often reveal new information that analysis of lower dimensional holes does not provide. In addition, if a machine learning algorithm is constructed, future work could test the model against existing methods of detecting Ventricular Tachycardia and compare the performances.

8   Acknowledgements

References

[1] Chad Giusti. *Applied Topology Lecture Notes*. URL: http://www.chadgiusti.com/uploads/4/9/9/2/49923137/567_notes_week_4_wed.pdf

[2] Karol Borsuk. *On the imbedding of systems of compacta in simplicial complexes*. URL: http://matwbn.icm.edu.pl/ksiazki/fm/fm35/fm35119.pdf

[3] Munson, Brian A.; Volić, Ismar Cubical homotopy theory. New Mathematical Monographs, 25. Cambridge University Press, Cambridge, 2015.

[4] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220.

[5] Nolle FM, Badura FK, Catlett JM, Bowser RW, Sketch MH. CREI-GARD, a new concept in computerized arrhythmia monitoring systems. Computers in Cardiology 13:515-518 (1986).

[6] García-González, M.A.; Argelagós-Palau, A.; Fernández-Chimeno, M.; Ramos-Castro, J., "A comparison of heartbeat detectors for the seismocardiogram," Computing in Cardiology Conference (CinC), 2013

[7] Munkres, J. R. (1993). Elements of Algebraic Topology. Boston, MA: Westview Press.

[8] Edelsbrunner, H., and Harer, J. (2010). Computational Topology. An Introduction. Rhode Island: American Mathematical Society. doi: 10.1090/mbk/069

[9] Edelsbrunner, H., Letscher, D., and Zomorodian, A. (2000). "Topological persistence and simplification," in Proceedings 41st Annual Symposium on Foundations of Computer Science (Redondo Beach, CA), 454–463. doi: 10.1109/SFCS.2000.892133

[10] giotto-tda: A Topological Data Analysis Toolkit for Machine Learning and Data Exploration, Tauzin et al, arXiv:2004.02551, 2020.

[11] Yu - Min Chung, et al. *A Persistent Homology Approach to Heart Rate Variability Analysis With an Application to Sleep-Wake Classification* URL : https://www.frontiersin.org/article/10.3389/fphys.2021.637684