Applications of TDA to Analyze Cardiovascular Signals

Subjects: Mathematics, Interdisciplinary Applications

Contributor: Enrique Hernández-Lemus , Pedro Miramontes , Mireya Martínez-García

Cardiovascular diseases (CVDs) have been, for a long time, a major global health problem that has caused more deaths than any form of cancer or respiratory disease combined. The detection and prediction of CVDs is made difficult by the numerous etiological factors, complex disease pathways, and diverse clinical presentations. Topological data analysis (TDA) is an approach for analyzing and interpreting complex data sets based on ideas a branch of mathematics called algebraic topology.

topological data analysis cardiovascular signals alegbraic topology

1. General Features

Aljanobi and Lee [1] recently applied the Mapper algorithm to predict heart disease. They selected nine significant features in each of the two UCI heart disease data sets (Cleveland and Statlog). The authors then used a tri-dimensional SVD filter to improve the filtering process. As a result, they observed an accuracy of 99.32% in the Cleveland data set and 99.62% in the Statlog data set in predicting heart disease.

Though not precisely *signals* but rather contextual string corpora (i.e., *texts*), TDA has also been applied to the analysis of structured and unstructured text in EHRs and clinical notes. This is achieved by first converting textual data into a suitable format for analysis. This can involve techniques such as tokenization, stemming, and removing stop words. Afterwards, one needs to represent the text as numerical vectors using methods like Term Frequency-Inverse Document Frequency (TF-IDF) or word embeddings (Word2Vec, GloVe). Once data are vectorized and tokenized, TDA can be applied. Lopez and coworkers [2], for instance, used the Mapper algorithm to classify distinctive subsets of patients receiving optimal treatments post-acute myocardial infarction (AMI) in order to identify high-risk subgroups of patients for having a future adverse event (AE) such as death, heart failure hospitalization, or recurrent myocardial infarction. A retrospective analysis of 31 clinical variables from the EHR of 798 AMI subjects was conducted at a single center. The subjects were divided into high- and low-risk groups based on their probability of survival without AEs at 1 year. TDA identified six subgroups of patients. Four of these subgroups, totaling 597 subjects, had a probability of survival without AEs that was greater than a one-fold change, and were considered low risk. The other two subgroups, totaling 344 subjects, had a probability of survival without AEs that was less than a one-fold change, and were considered high risk. However, 143 subjects (18% of the total) were classified as intermediate risk because they belonged to both high- and low-risk subgroups. TDA was also able to significantly stratify AMI patients into three subgroups with distinctive rates of AEs up to 3 years after AMI.

This approach to EHR-based risk stratification does not require additional patient interaction and is not dependent on prior knowledge, but more studies are needed before it can be used in clinical practice.

2. Electrocardiographic Data and Heart Rate Signals

In another recent study, Yan and coworkers [3] explored the use of topological data analysis to classify electrocardiographic signals and detect arrhythmias. Cardiac arrhythmias are abnormal heart rhythms or irregularities in the heartbeat that may be too fast (tachycardia), too slow (bradycardia), or irregular. Arrhythmias may originate from problems with the heart's electrical system, damage to the heart muscle, or other medical conditions. The most common types of arrhythmias are *Atrial Fibrillation* (AF) defined as an irregular and often rapid heartbeat that can lead to stroke and other heart-related complications; *Atrial Flutter* which is similar to AF, is characterized by a rapid, regular heartbeat originating in the atria; *Supraventricular Tachycardia* is characterized by episodes of rapid heart rate originating above the heart's ventricles; *Ventricular Tachycardia*, a fast, regular beating of the heart's lower chambers (ventricles) that can be life-threatening; *Ventricular Fibrillation* (VF) which is a chaotic, rapid heartbeat that can be life-threatening and requires immediate medical attention; and finally *Bradycardia*, a slower than normal heart rate, often caused by issues with the heart's natural pacemaker.

In the particular case of reference [3], phase space reconstruction was used to convert the signals into point clouds, which were then analyzed using topological techniques to extract persistence landscapes as features for the classification task. The authors found that the proposed method was effective, with a normal heartbeat class recognition rate of 100% when using just 20% of the training set, and recognition rates of 97.13% for ventricular beats, 94.27% for supraventricular beats, and 94.27% for fusion beats. This ability to maintain high performance with a small training sample space makes TDA particularly suitable for personalized analysis.

One particularly difficult problem in cardiovascular disease diagnostics with important implications for therapy is the evolution of atrial fibrillation [4]. Indeed, the progression of AF from paroxysmal to persistent or permanent forms has become a significant issue in cardiovascular disorders. Information about the pattern of presentation of AF (paroxysmal, persistent, or permanent) is useful in the management of algorithms for each category, which aims to reduce symptoms and prevent severe problems associated with AF. Until now, AF classification has been based on the duration and number of episodes. In particular, changes in complexity of Heart Rate Variation (HRV) may contain clinically relevant signals of impending systemic dysregulation. HRV measures the fluctuations in the time intervals between consecutive heartbeats, providing insights into the autonomic nervous system's activity, particularly the balance between its sympathetic and parasympathetic branches. A number of nonlinear methods based on phase space and topological properties can provide further insight into HRV abnormalities such as fibrillation. In an effort to provide a tool for the qualitative classification of AF stages, Safarbaly and Golpayegani [5] proposed two geometrical indices (fractal dimension and persistent homology) based on HRV phase space, which were able to successfully replicate the changes in AF progression.

Their studied population included 38 lone AF patients and 20 normal subjects, whose data were collected from the Physio-Bank database [6]. "Time of Life (TOL)" was proposed as a new feature based on the initial and final Čech radius in the persistent homology diagram. A neural network was implemented to demonstrate the effectiveness of both TOL and fractal dimension as classification features, resulting in a classification accuracy of 93%. The proposed indices thus provide a signal representation framework useful for understanding the dynamic changes in AF cardiac patterns but also for classifying normal and pathological rhythms.

PH was also used by Graff et al. to study HRV [7]; the authors suggested the use of persistent homology for the analysis of HRV, relying on some topological descriptors previously used in the literature and introducing new ones that are specific to HRV, later discussing their relationship to standard HRV measures. The authors showed that this novel approach produces a set of indices that may be as useful as classical parameters in distinguishing between series of beat-to-beat intervals (RR-intervals) in healthy individuals as well as in patients who have experienced a stroke.

Also in the context of fibrillation (this time for the prediction of early ventricular fibrillation), Ling and coworkers [8] proposed a novel feature based on topological data analysis to increase the accuracy of early VF prediction. In their work, the heart activity was first treated as a cardiac dynamical system, which was described through phase space reconstruction. The topological structure of the phase space was then characterized using persistent homology, and statistical features of the topological structure were extracted and defined as TDA features. To validate the prediction performance of the proposed method, 60 subjects (30 VF, 30 healthy) from three public ECG databases were used. The TDA features showed a superior accuracy of 91.7% compared to heart rate variability features and box-counting features. When all three types of features were combined as fusion features, the optimal accuracy of 95.0% was achieved. The fusion features were then ranked, and the first seven components were all from TDA features. The proposed features may have a significant effect on improving the predictive performance of early VF.

A similar approach was taken by Mjahad, et al. [9], who applied TDA to generate novel features contributing to improve both detection and classification performance of cardiac arrhythmias such as Ventricular Fibrillation (VF) and Ventricular Tachycardia (VT). The electrocardiographic (ECG) signals used for this evaluation were obtained from the standard MIT-BIH and AHA databases. The authors evaluated two types of input data for classification: TDA features and the so-called Persistence Diagram Image (PDI). When using the reduced TDA-derived features, a high average accuracy of nearly 99% was observed to discriminate between four types of rhythms (98.68% for VF; 99.05% for VT; 98.76% for normal sinus; and 99.09% for other rhythms), with specificity values higher than 97.16% in all cases. In addition, a higher accuracy of 99.51% was obtained when discriminating between shockable (VT/VF) and non-shockable rhythms (99.03% sensitivity and 99.67% specificity). These results show that the use of TDA-derived geometric features, combined with the k-Nearest Neighbor (kNN) classifier, significantly improves classification performance compared to previous works. These results were achieved without pre-selection of ECG episodes, suggesting that these features may be successfully used in Automated External Defibrillation (AED) [10][11] and Implantable Cardioverter Defibrillation (ICD) [12][13] therapies.

Jiang and collaborators studied non-invasive atrial fibrillation using TDA on ballistocardiographic (BCG) data [14]. BCG refers to the measurement and recording of the ballistic forces generated by the ejection of blood from the heart during each cardiac cycle. These forces produce subtle vibrations and movements in the body, particularly in the chest and torso. Ballistocardiography is a non-invasive technique used to capture and analyze these mechanical movements, providing valuable information about cardiac function and hemodynamics. In this research, BCG series was transformed into a high-dimensional point cloud in order to capture more rhythmic information. These point clouds were then analyzed using TDA, resulting in persistent homology barcodes. The statistics of these barcodes were extracted as nine persistent homology features to quantitatively describe the barcodes. In order to test the effectiveness of this method for detecting atrial fibrillation (AF), the researchers collected BCG data from 73 subjects with both AF and non-AF segments, and applied six machine learning classifiers. The combination of these 9 features with 17 previously proposed features resulted in a 6.17% increase in accuracy compared to using 17 features alone ($p<0.001$�<0.001), with an overall accuracy of 94.50%. By selecting the most effective features using feature selection, the researchers were able to achieve a classification accuracy of 93.50%. According to the authors, these results suggest that the proposed features can improve AF detection performance when applied to a large amount of BCG data with diverse pathological information and individual differences.

TDA can also be combined with other data analytics/Ml approaches such as random forests (RF). Ignacio and collaborators developed a topological method to *inform* RF-based ECG classifiers [15]. In brief, in this approach, a two-level random forest model is trained to classify 12-lead ECGs using mathematically computable topological signatures as proxy for features informed by medical expertise. ECGs are treated as multivariate time series data and transformed into point cloud embeddings that capture both local and global structures and encode periodic information as attractor cycles in a high-dimensional space. Topological data analysis is then used to extract topological features from these embeddings, and these features are combined with demographic data and statistical moments of RR intervals calculated using the Pan–Tompkins algorithm for each lead to train the classifier. This multi-class classification task aims to leverage the medical expertise represented in the topological signatures to accurately classify ECGs.

The same group combined TDA and ML approaches to study atrial fibrillation in ECGs [16], showing that topological features can be used to accurately classify single-lead ECGs by applying delay embeddings to map ECGs onto high-dimensional point clouds, which convert periodic signals into algebraically computable topological signatures, thus allowing them the use of these topological features to classify ECGs.

A similar approach coupling TDA with Deep Learning to study ECGs has been recently developed [17]. Training deep learning models on time series data such as ECG poses some challenges such as lack of labeled data and class imbalance [18][19]. The authors used TDA to improve the performance of deep learning models on this type of data. The authors found that using TDA as a time-series embedding method for input to deep learning models resulted more effective than training these models directly on raw data. TDA in this context serves as a generic, low-level feature extractor that can capture common signal patterns and improve performance with limited training data. Experiments on public human physiological biosignal data sets showed that this approach leads to improved accuracy, particularly for imbalanced classes with only a few training instances compared to the full data set.

Conventional TDA of ECG time series can be further improved by considering the time delay structure to generate higher dimensional mappings of the original series [20] able to be analyzed via TDA. This was the approach taken by Fraser and coworkers [21], who found that TDA visualizations are able top unveil ectopic and other abnormal occurrences in long signals, indicating a promising direction for the study of longitudinal physiological signals.

A different approach to deal with ECG time series using TDA makes use of optimal representative cycles. In reference [22], the authors applied a topological data-analytic method to identify parts of an electrocardiogram (ECG) signal that are representative of specific topological features and proposed that these parts correspond to the P, Q, S, and T-waves in the ECG signal. They then used information about these parts of the signal, identified as P, Q, S, and T-waves, to measure PR-interval, QT-interval, ST-segment, QRS-duration, P-wave duration, and T-wave duration. The method was tested on simulated and real Lead II ECG data, demonstrating its potential use in analyzing the morphology of the signal over time and in arrhythmia detection algorithms.

3. Stenosis and Vascular Data

TDA has also been used to study stenosis, an abnormal narrowing or constriction of blood vessels. In the cardiovascular system, arterial stenosis can be particularly significant. Atherosclerosis, a condition characterized by the buildup of plaque on the inner walls of arteries, is a common cause of arterial stenosis. Nicponski and collaborators [23] demonstrated the use of persistent homology to assess the severity of stenosis in different types of stenotic vessels. They introduced the concept of *critical failure value*, which applies one-dimensional homology to these vessels as a way to quantify the degree of stenosis. They also presented the spherical projection method, which could potentially be used to classify various types and levels of stenosis, and showed that the two-dimensional homology of the spherical projection can serve as a new index for characterizing blood vessels. It is worth noticing that, as in many other instances in data analytics, data pre-processing often represents a crucial stage [24].

4. Topological Data Analysis in Echocardiography

Applications of TDA to echocardiographic data have also been developed. Such is the case of the work the group of Tokodi [25][26] who analyzed a cohort of 1334 patients to identify similarities among patients based on several echocardiographic measures of left ventricular function. Echocardiography, a medical imaging technique that uses sound waves to create detailed images of the heart, generates 2D and 3D maps of vascular structure. However, accurate image reconstruction is far from trivial, in particular for small convoluted cavities. A network was developed in reference [25] to represent these similarities, and a group classifier was used to predict the location of 96 patients with two consecutive echocardiograms in this network. The analysis revealed four distinct regions in the network, each with significant differences in the rate of major adverse cardiovascular event (MACE) rehospitalization. Patients in the fourth region had more than two times the risk of MACE rehospitalization compared to those in the other regions. Improvement or stability in Regions I and II was associated with lower MACE rehospitalization rates compared to worsening or stability Regions III and IV. The authors concluded that

TDA-driven patient similarity analysis may improve the precision of phenotyping and aid in the prognosis of patients by tracking changes in cardiac function over time.

References

1. Aljanobi, F.A.; Lee, J. Topological Data Analysis for Classification of Heart Disease Data. In Proceedings of the 2021 IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju Island, Republic of Korea, 17–20 January 2021; pp. 210–213.
2. Lopez, J.E.; Datta, E.; Ballal, A.; Izu, L.T. Topological Data Analysis of Electronic Health Record Features Predicts Major Cardiovascular Outcomes After Revascularization for Acute Myocardial Infarction. Circulation 2022, 146, A14875.
3. Yan, Y.; Ivanov, K.; Cen, J.; Liu, Q.H.; Wang, L. Persistence landscape based topological data analysis for personalized arrhythmia classification. 2019; preprints.
4. Falsetti, L.; Rucco, M.; Proietti, M.; Viticchi, G.; Zaccone, V.; Scarponi, M.; Giovenali, L.; Moroncini, G.; Nitti, C.; Salvi, A. Risk prediction of clinical adverse outcomes with machine learning in a cohort of critically ill patients with atrial fibrillation. Sci. Rep. 2021, 11, 18925.
5. Safarbali, B.; Hashemi Golpayegani, S.M.R. Nonlinear dynamic approaches to identify atrial fibrillation progression based on topological methods. Biomed. Signal Process. Control 2019, 53, 101563.
6. Goldberger, A.L.; Amaral, L.A.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000, 101, e215–e220.
7. 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, e0253851.
8. Ling, T.; Zhu, Z.; Zhang, Y.; Jiang, F. Early Ventricular Fibrillation Prediction Based on Topological Data Analysis of ECG Signal. Appl. Sci. 2022, 12, 10370.
9. Mjahad, A.; Frances-Villora, J.V.; Bataller-Mompean, M.; Rosado-Muñoz, A. Ventricular Fibrillation and Tachycardia Detection Using Features Derived from Topological Data Analysis. Appl. Sci. 2022, 12, 7248.
10. Caffrey, S.L.; Willoughby, P.J.; Pepe, P.E.; Becker, L.B. Public use of automated external defibrillators. N. Engl. J. Med. 2002, 347, 1242–1247.
11. Delhomme, C.; Njeim, M.; Varlet, E.; Pechmajou, L.; Benameur, N.; Cassan, P.; Derkenne, C.; Jost, D.; Lamhaut, L.; Marijon, E.; et al. Automated external defibrillator use in out-of-hospital cardiac arrest: Current limitations and solutions. Arch. Cardiovasc. Dis. 2019, 112, 217–222.

12. Kamp, N.J.; Al-Khatib, S.M. The subcutaneous implantable cardioverter-defibrillator in review. Am. Heart J. 2019, 217, 131–139.
13. Friedman, P.; Murgatroyd, F.; Boersma, L.V.; Manlucu, J.; O'Donnell, D.; Knight, B.P.; Clémenty, N.; Leclercq, C.; Amin, A.; Merkely, B.P.; et al. Efficacy and safety of an extravascular implantable cardioverter–defibrillator. N. Engl. J. Med. 2022, 387, 1292–1302.
14. Jiang, F.; Xu, B.; Zhu, Z.; Zhang, B. Topological Data Analysis Approach to Extract the Persistent Homology Features of Ballistocardiogram Signal in Unobstructive Atrial Fibrillation Detection. IEEE Sens. J. 2022, 22, 6920–6930.
15. Ignacio, P.S.; Bulauan, J.A.; Manzanares, J.R. A Topology Informed Random Forest Classifier for ECG Classification. In Proceedings of the 2020 Computing in Cardiology, Rimini, Italy, 13–16 September 2020; pp. 1–4.
16. Ignacio, P.S.; Dunstan, C.; Escobar, E.; Trujillo, L.; Uminsky, D. Classification of single-lead electrocardiograms: TDA informed machine learning. In Proceedings of the 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA), Boca Raton, FL, USA, 16–19 December 2019; pp. 1241–1246.
17. Byers, M.; Hinkle, L.B.; Metsis, V. Topological Data Analysis of Time-Series as an Input Embedding for Deep Learning Models. In Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations, Crete, Greece, 17–20 June 2022; Springer: Cham, Switzerland, 2022; pp. 402–413.
18. Seversky, L.M.; Davis, S.; Berger, M. On time-series topological data analysis: New data and opportunities. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 27–30 June 2016; pp. 59–67.
19. Karan, A.; Kaygun, A. Time series classification via topological data analysis. Expert Syst. Appl. 2021, 183, 115326.
20. Sun, F.; Ni, Y.; Luo, Y.; Sun, H. ECG Classification Based on Wasserstein Scalar Curvature. Entropy 2022, 24, 1450.
21. Fraser, B.A.; Wachowiak, M.P.; Wachowiak-Smolíková, R. Time-delay lifts for physiological signal exploration: An application to ECG analysis. In Proceedings of the 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Windsor, ON, Canada, 30 April–3 May 2017; pp. 1–4.
22. Dlugas, H. Electrocardiogram feature extraction and interval measurements using optimal representative cycles from persistent homology. bioRxiv 2022.
23. Nicponski, J.; Jung, J.H. Topological data analysis of vascular disease: A theoretical framework. Front. Appl. Math. Stat. 2020, 6, 34.

24. Bresten, C.L.; Kweon, J.; Chen, X.; Kim, Y.H.; Jung, J.H. Preprocessing of general stenotic vascular flow data for topological data analysis. bioRxiv 2021.
25. Tokodi, M.; Shrestha, S.; Ashraf, M.; Casaclang-Verzosa, G.; Sengupta, P. Topological Data Analysis for quantifying inter-patient similarities in cardiac function. J. Am. Coll. Cardiol. 2019, 73, 751.
26. Tokodi, M.; Shrestha, S.; Bianco, C.; Kagiyama, N.; Casaclang-Verzosa, G.; Narula, J.; Sengupta, P.P. Interpatient similarities in cardiac function: A platform for personalized cardiovascular medicine. Cardiovasc. Imaging 2020, 13, 1119–1132.