



OPEN

## A novel multi-task machine learning classifier for rare disease patterning using cardiac strain imaging data

Nanda K. Siva<sup>1,2</sup>, Yashbir Singh<sup>2,3</sup>, Quincy A. Hathaway<sup>1,2</sup>, Partho P. Sengupta<sup>4</sup>✉ & Naveena Yanamala<sup>4,5</sup>✉

To provide accurate predictions, current machine learning-based solutions require large, manually labeled training datasets. We implement persistent homology (PH), a topological tool for studying the pattern of data, to analyze echocardiography-based strain data and differentiate between rare diseases like constrictive pericarditis (CP) and restrictive cardiomyopathy (RCM). Patient population (retrospectively registered) included those presenting with heart failure due to CP (n = 51), RCM (n = 47), and patients without heart failure symptoms (n = 53). Longitudinal, radial, and circumferential strains/strain rates for left ventricular segments were processed into topological feature vectors using Machine learning PH workflow. In differentiating CP and RCM, the PH workflow model had a ROCAUC of 0.94 (Sensitivity = 92%, Specificity = 81%), compared with the GLS model AUC of 0.69 (Sensitivity = 65%, Specificity = 66%). In differentiating between all three conditions, the PH workflow model had an AUC of 0.83 (Sensitivity = 68%, Specificity = 84%), compared with the GLS model AUC of 0.68 (Sensitivity = 52% and Specificity = 76%). By employing persistent homology to differentiate the “pattern” of cardiac deformations, our machine-learning approach provides reasonable accuracy when evaluating small datasets and aids in understanding and visualizing patterns of cardiac imaging data in clinically challenging disease states.

**Keywords** Echocardiography, Machine learning, Constrictive pericarditis, Restrictive cardiomyopathy, Rare disease

### Abbreviations

4ch	4-Chamber
AUC	Area under curve
CP	Constrictive pericarditis
GLS	Global longitudinal strain
PH	Persistent homology
PI	Persistent image
RCM	Restrictive cardiomyopathy
Sn	Sensitivity
Sp	Specificity
STE	Speckle-tracking echocardiography
TDA	Topological data analysis

Effectively interpreting the vast amount of medical data generated daily in healthcare is paramount to improving clinical decision-making. Machine learning has been employed in the healthcare field to discover meaningful trends. For cardiac imaging, advances in artificial intelligence have improved both the speed and accuracy of

<sup>1</sup>School of Medicine, West Virginia University, Morgantown, WV, USA. <sup>2</sup>Division of Cardiology, Heart and Vascular Institute, West Virginia University, Morgantown, WV, USA. <sup>3</sup>Department of Radiology, Mayo Clinic, Rochester, MN, USA. <sup>4</sup>Division of Cardiovascular Disease and Hypertension, Rutgers Robert Wood Johnson Medical School, 125 Patterson St, New Brunswick, NJ 08901, USA. <sup>5</sup>Institute for Software Research, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA. ✉email: partho.sengupta@rutgers.edu; ny128@rwjms.rutgers.edu

image interpretation as well as have facilitated detection of subtle changes in cardiac structure and function. Many studies have demonstrated the value of artificial intelligence in cardiac imaging, deep learning to classify left ventricular hypertrophy in echocardiography images<sup>1</sup>, structured random forests to trace myocardium borders in 3D echocardiograph volumes<sup>2</sup>, and convolution neural network model to develop high resolution images from 2D magnetic resonance image stacks<sup>3</sup>.

However, the large number of attributes present in many datasets must be reduced to prevent overwhelming the machine learning algorithms. For example, data derived from cardiac deformation analyses capture left ventricular wall motion in both 2- and 3-dimensional planes, generating a significant amount of data for machine learning applications<sup>4</sup>. Although we have a vast amount of data, today's clinical evaluations usually depend on just one data point. The methods we use to gain insights from all the data haven't shown to be more helpful in guiding patient care<sup>5</sup>.

Dimensionality reduction techniques, such as Principal Component Analysis and Linear Discriminant Analysis, are commonly applied techniques to better aggregate large feature sets. Another dimensionality reduction technique growing in prominence is Topological Data Analysis (TDA), which extracts data features based on local geometry and global topology encoded in the distribution of data points.

Many studies have incorporated general TDA principles in identifying patterns in nature, e.g., zebrafish pattern variability<sup>6</sup>, stability of protein folding<sup>7</sup>, and outcomes in preclinical traumatic brain injury and spinal cord injury<sup>8</sup>. Other groups have specifically utilized Persistent Homology from the TDA toolbox to understand human anatomy and physiology, e.g., to study brain artery branching and looping<sup>9</sup>, gait signals in patients with neurodegenerative diseases<sup>10</sup>, and MRI liver images in patients with Primary Sclerosing Cholangitis<sup>5</sup>. In this study, we propose a workflow for applying persistent homology to study high-dimensional data where the ratio of sample to features is very low.

The purpose of this study is to determine if topological data analysis can help in identifying patterns in echocardiography images to improve diagnostic accuracy for rare cardiac conditions. We provide a use-case scenario of our workflow that harnesses both deformation patterns and global topological information of the left ventricle to characterize cardiovascular diseases (Figs. 1, 2, 3).

## Results

### Average strain pattern motifs

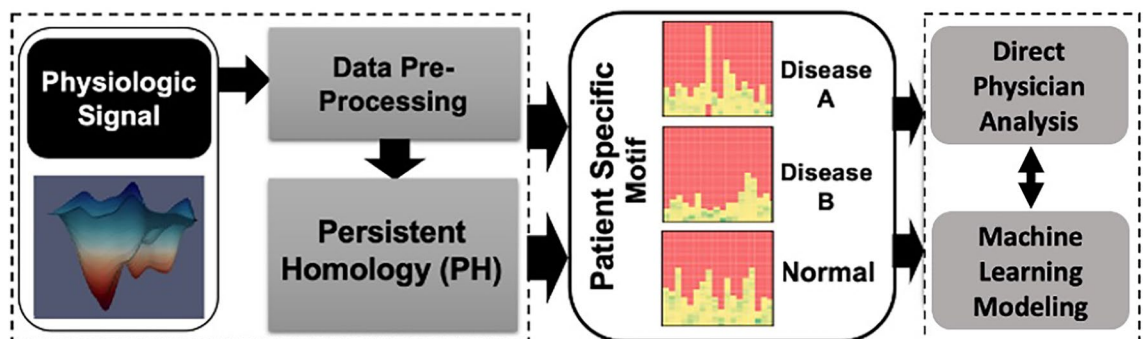
The average strain patterns motifs for each cardiac condition are shown (Fig. 4). From visual inspection, general trends can be identified, e.g., RCM patients generally have higher intensity values limited to lower persistence pixels while having lower intensity values at higher persistence pixels. This indicates a restrictive pattern within these patients as it suggests that a fully connected component in the H0 dimension is formed at a lower scale parameter. Alternatively, for both CP and normal groups, the average motifs showed a much wider spread in their strain persistence pixel intensities, indicating more spacing in these patients' data points than the general constraint seen in RCM.

### Feature selection

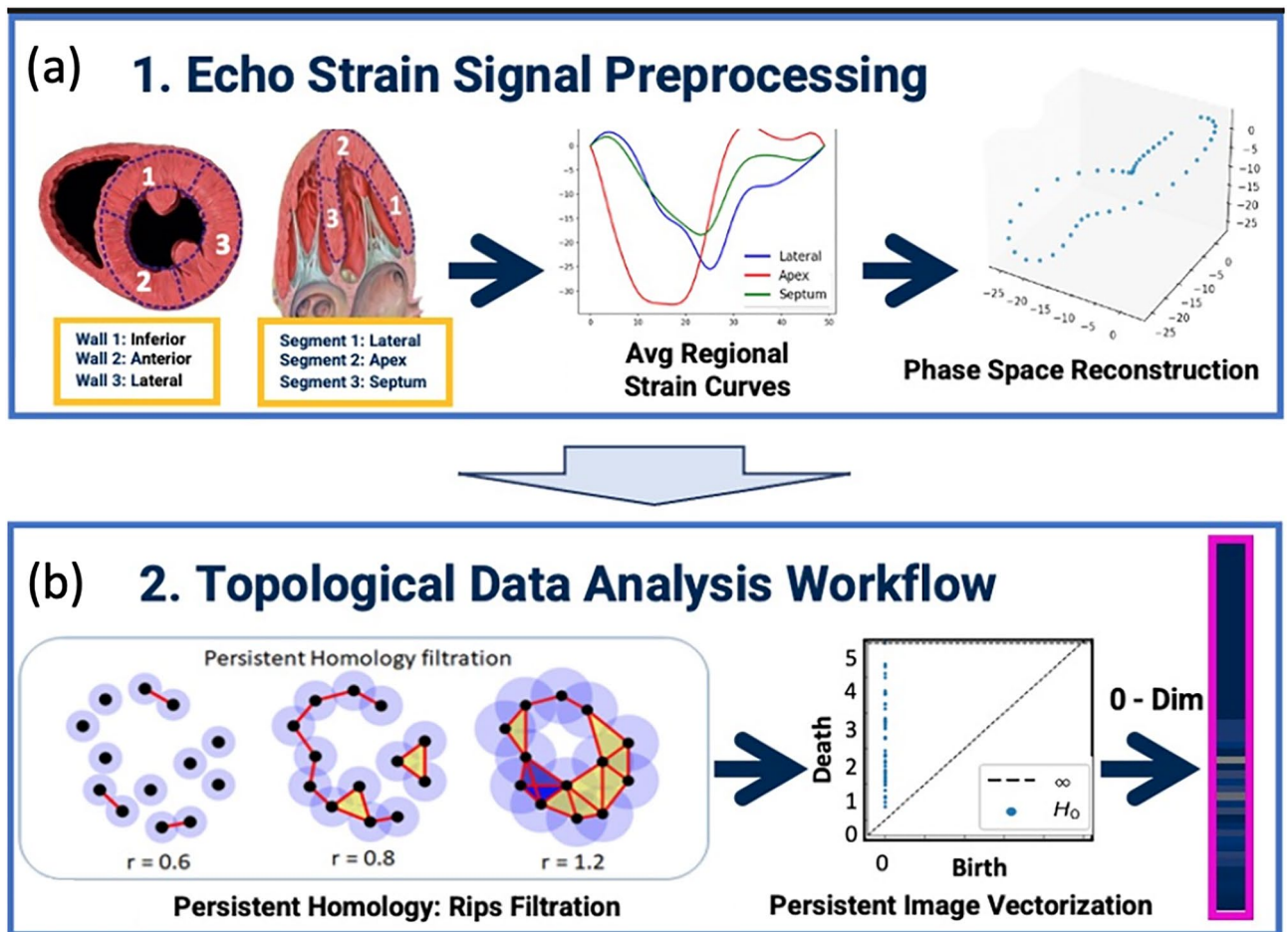
While the number of features was considerably reduced from 14,700 to 900 (18 by 50), this data still exhibited a high feature to sample ratio. The selected features through Boruta feature selection were used to develop predictive models for distinguishing between CP, RCM, and normal patients.

### Machine learning classifiers

To determine if the features extracted through our pipeline helped distinguish the cardiac conditions, we developed three binary class classifiers for CP vs. RCM, CP vs. normal, and RCM vs. normal. The combined dataset from Amaki et al.<sup>11</sup> and Sengupta et al.<sup>12</sup> was evaluated using tenfold cross-validation. Finally, we compared the performance of these models with a baseline performance achieved by logistic regression models using average



**Figure 1.** Outline of the general proposed workflow. Physiological signal is pre-processed to a suitable input for persistent homology (PH) feature extraction. The resulting topological features are represented visually as an individualized patient motif produced by concatenating multiple persistent images. This motif can be directly interpreted by physician or used to develop machine learning models.



**Figure 2.** Schematic overview of use case scenario. Three regions were defined in the mid short axis and apical four chamber view; the average regional curve for each strain parameter was calculated and transformed using phase space reconstruction (a). This served as the input for persistent homology filtration; the resulting birth and death coordinates, based on radius ( $r$ ) values, for dimension 0 were converted to a feature vector form for machine learning using persistent image methodology (b).

peak longitudinal strain from the 4Ch view. This peak value approximates the global longitudinal strain that clinicians typically extract from cardiac strain imaging data (Fig. 5).

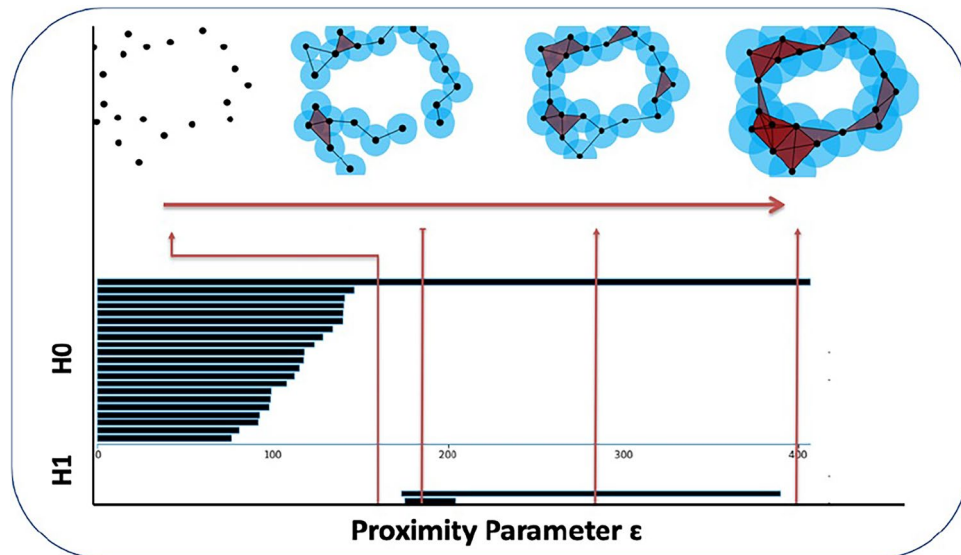
Our CP vs RCM logistic regression classifier showed a statistically significant improvement compared to the GLS model (PH AUC = 0.94; GLS AUC = 0.69;  $p = 1.4 \times 10^{-6}$ ). Our CP vs normal logistic regression classifier demonstrated improvement as well (PH AUC = 0.83; GLS AUC = 0.66;  $p = 0.019$ ). Our RCM vs normal random forest classifier showed a statistically significant improvement compared to the GLS model (PH AUC = 0.91; GLS AUC = 0.82;  $p = 0.028$ ). We created a multi-class random forest classifier to discriminate between all conditions; the average across all classes AUC, sensitivity (Sn), and specificity (Sp) were improved in comparison to the baseline model. Our PH model achieved AUC = 0.83 (Sn = 68% and Sp = 84%) whereas GLS model achieved AUC = 0.68 (Sn = 52% and Sp = 76%).

### Interpretable artificial intelligence

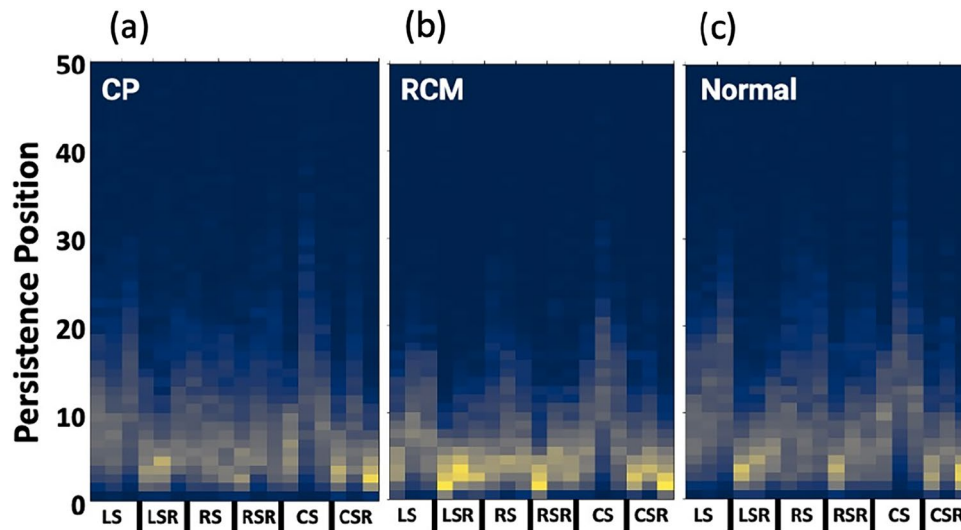
We show the interpretable artificial intelligence results as Shapley additive explanation plots indicating the top ten features integral in distinguishing each class from the others (Fig. 6). Thus, a combination of feature trends is responsible for the model to output a particular prediction. Moreover, these results allow better comprehension of the average strain motifs produced (Fig. 4). To understand these patterns, we can refer to the original phase space reconstruction point clouds for septal longitudinal strain; for convenience, a few example patients from each disease group are depicted in Supplementary Figs. 1 and 2.

### Discussion

Compared with traditional echocardiography approaches, our methodology uses subclinical features identified using topological data extraction of segmental strain analysis to more clearly delineate clinically similar cardiac phenogroups, such as restrictive cardiomyopathy (RCM) and constrictive pericarditis (CP). Global and local structural deformations of the cardiac myocardium were captured with persistent homology, enabling us to

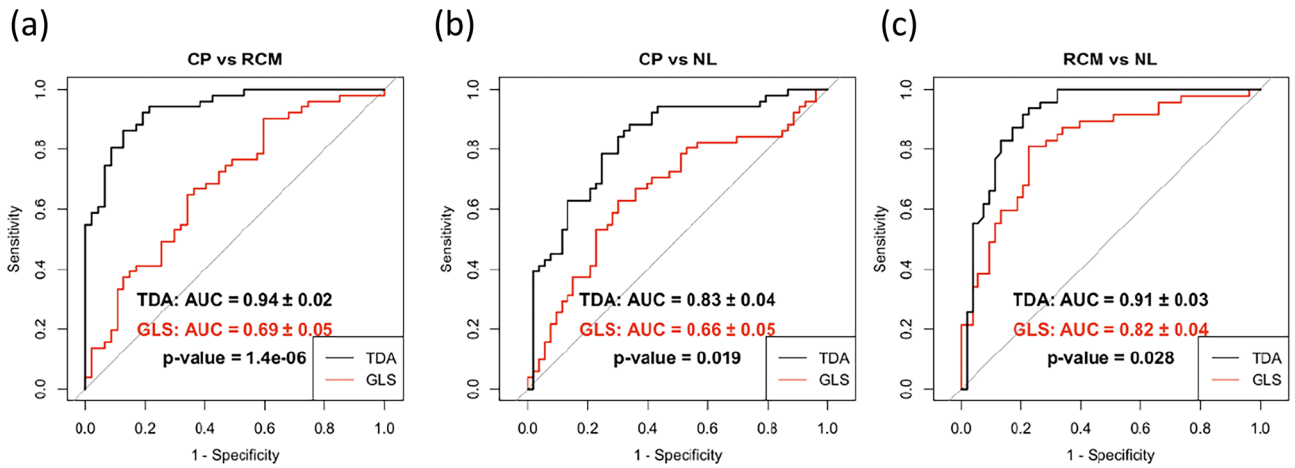


**Figure 3.** Persistent Homology Explanation. For a given proximity parameter  $\epsilon$ , a circle with radius  $= \epsilon$  is drawn around each data point. The intersections of these circles guide the construction of a set of simplicial complexes. All possible values of  $\epsilon$  are tested to detect variations in topology at different scales. The appearance and disappearance of connected components and open loops is measured by H0 and H1, respectively, and subsequently visualized as a persistent barcode. H0 represents dimension 0 persistent homology, and H1 represents dimension 1 persistent homology.

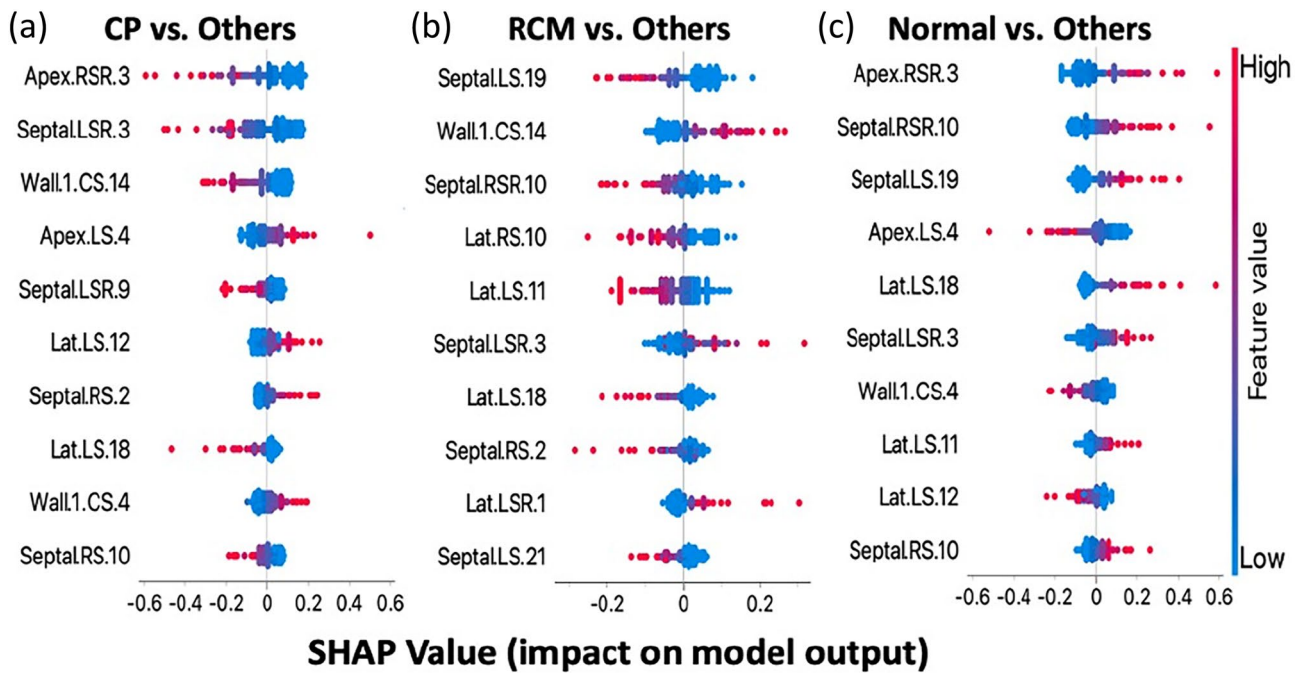


**Figure 4.** Cardiac condition average strain motifs. The output visual signature for persistent homology workflow was a heatmap-like motif with x-axis corresponding to strain or strain rate in longitudinal, radial, and circumferential directions; y-axis corresponding to persistence pixel position in the resulting 1 by 50 vector for each combination of wall region and strain measurement; and z-color corresponding to the pixel intensity calculated through persistent image vectorization. The average strain motifs for constrictive pericarditis, restrictive cardiomyopathy, and normal/control patients are shown to demonstrate group defining patterns. CS circumferential strain, CSR circumferential strain rate, CP constrictive pericarditis, LS longitudinal strain, LSR longitudinal strain rate, RCM restrictive cardiomyopathy, RS radial strain, and RSR radial strain rate.

predict the presence of CP (AUC: 0.83) and RCM (AUC: 0.91) from normal patients. Additionally, when directly differentiating between the two types of heart failure, CP and RCM (AUC: 0.94), our model demonstrated a comparable combination of sensitivity (92%) and specificity (81%) compared to the sensitivity (87%) and specificity (91%) shown in a study by Welch et al. at the Mayo Clinic investigating the conventional evaluation of constrictive pericarditis from restrictive myocardial disease or severe tricuspid regurgitation based on five principal echocardiographic features, including respiration related ventricular septal shift, maintained or greater



**Figure 5.** Binary classifier receiver operating curves. Cross-validation (tenfold) area under the curve (AUC) and respective p values for the models classifying binary conditions, restrictive cardiomyopathy (RCM), Constrictive pericarditis (CP), and normal (NL) is shown. Across all binary class distinctions, the persistent homology workflow outperformed the performance metrics of the GLS model. *GLS* Global Longitudinal Strain.



**Figure 6.** Shapley additive explanations for multi-class model. Shapley plot presents the top ten important features responsible for the multi-class machine learning model to output its predictions that discriminated each cardiac condition from the other two, i.e., constrictive pericarditis (CP) from restrictive cardiomyopathy (RCM) and Normal. The features identified from this interpretable artificial intelligence tool can be corresponded to specific regions in the patient specific motif that contribute to disease stratification visually. *CS* circumferential strain, *CSR* circumferential strain rate, *lat* lateral wall, *LS* longitudinal strain, *LSR* longitudinal strain rate, *RS* radial strain, and *RSR* radial strain rate.

medial mitral annular *e'* velocity, and expiratory diastolic reversal ratio of the hepatic vein<sup>13</sup>. To provide context of our workflow's comparison to another non-invasive method assessing clinically similar cardiac phenogroups, in a study by Masui et al. cardiac MRI was used to differentiate CP from RCM, with a sensitivity of 88% and specificity of 100%; our model performed worse in terms of specificity but better in terms of sensitivity<sup>14</sup>. While our TDA/echocardiography model performs similarly, it is important to note that ultrasound-based techniques offer a wider range of accessibility to patients than either CT or MRI.

In this small cohort, we highlight the ability of our TDA model to make accurate predictions of rare disease presentations as a use-case example for future cardiovascular applications. In diseases with low prevalence, this presents an obvious advantage to traditional approaches that may require a specific threshold of cases before allowing appropriate stratification. Focal involvement, either decompensation or compensation, of the

myocardium will be captured in our current workflow in the appropriate wall region which would be useful in both common and rare cardiac conditions, such as myocardial infarction and apical hypertrophic cardiomyopathy, respectively. However, in order to gain more granularity to exactly which portion the deviation originates from, our workflow could subgroup the left ventricular wall into the American Heart Association 18–19 segments<sup>15</sup>, which would also require expanding our patient specific motif.

Previous investigations have endeavored to apply segmental strain analysis to predict structural and/or functional outcomes in the heart. Tabassian et al. used principal component analysis to represent the complex spatio-temporal nature of stress–strain curves and utilized machine learning to classify patients with myocardial infarctions<sup>16</sup>. Senapati et al. proposed a relative regional strain ratio (a metric of relative longitudinal strain sparing in the apex) to provide prognostic information in cardiac amyloidosis patients<sup>17</sup>. While these applications provide meaningful interpretation of segmental strain data, their translation to diverse cardiac phenotypes is likely limited by the lack of integration of the ultrastructural component of the left ventricle (LV). Another limitation of segmental strain data is high variability in measurements between vendors<sup>18</sup>; we attempt to reduce the effects of this issue in our workflow through standardizing patients' cardiac cycles with spline interpolation and by aggregating individual segments into larger functional wall regions. The EACVI-ASE Strain Standardization Task Force recommends utilizing segmental strain pattern analysis rather than single segmental strain values<sup>19</sup>. We believe application of our protocol inherently accomplishes this as persistent homology is a topological data analysis tool that describes the shape of data by extracting its topological invariants<sup>20</sup>.

TDA has been increasingly applied to areas of biomedical research<sup>10,21,22</sup> but only recently been evaluated in cardiovascular medicine. Specifically, diagnostic tests such as the ECG have provided the first applications of TDA in converting simple waveforms to numeric data<sup>23,24</sup>. More recently, TDA has been proposed as a method for the assessment of vascular diseases<sup>25</sup> and has even provided improved predictive capacity for detecting acute coronary syndrome or revascularization in patients with coronary plaques than through the use of more commonly used clinical markers, such as risk factors, stenosis, and high-risk plaque features<sup>26</sup>.

Our work can be translated into clinical applications such as a medical decision support system for physicians, AI virtual assistant for patients, or an automated image analysis software. To facilitate accessibility for physicians and scientists with varying levels of expertise/understanding, we plan to provide the visual motifs with annotations and labels; providing comparison images of disease states and normal motifs will also enhance comprehension for a broader audience.

The limitation inherent to many studies investigating rare diseases is the relatively small sample size, which can be due to practical and resource constraints. However, a key strength of topological data analysis is its ability to find patterns in small groups of data<sup>27</sup>. For further studies, increasing the sample size collected will help address most data/performance errors. The inclusion of other cardiac pathologies, such as dilated cardiomyopathy, ischemic cardiomyopathy, and valvular heart diseases, along with the integration of other input data types, such as ratios of regional strains<sup>28</sup>, cardiac MRI, or patient specific demographics, can enhance the versatility of the workflow in the clinical setting. A limitation to the current study is the analysis of only RCM and CP; without consideration of other cardiac pathologies and how their strain parameters may help, or interfere, with correct classification it is unclear how this will generalize to other uncommon pathologies.

The current application of our use-case scenario highlights the ability of TDA, and more specifically persistent homology, to correctly stratify unique cardiovascular anomalies from segmental stress strain analysis.

## Materials and methods

### Study population

In this retrospective case study, we utilized a merged cohort from two previously published datasets<sup>11,12</sup>, with a total of 54 constrictive pericarditis (CP), 49 restrictive cardiomyopathies (RCM), and 55 no structural heart failure control patients (normal).

The institutional review board at the Mayo Foundation approved the protocol outlined by Amaki et al.<sup>11</sup>. Between July 2005 and January 2007, 37 consecutive patients with CP that were scheduled for pericardiectomy treatment and 22 heart failure patients diagnosed with RCM through transthoracic echocardiography; due to suboptimal 2D-echocardiography image quality, seven patients with CP were excluded. Of the remaining patients, 26 with CP and 19 with RCM provided informed consent for participation in the study. Additionally, there was recruitment of 21 control subjects without cardiovascular disease and no evidence of left ventricular dysfunction or significant valvular heart disease observed with echocardiography.

The institutional review board at the Mount Sinai Medical Center approved the protocol by Sengupta et al.<sup>12</sup>. 92 patients (28 with CP, 30 with RCM, and 34 control with no structural/functional abnormality) who underwent transthoracic echocardiography imaging were retrospectively identified.

Three CP patients, 2 RCM patients, and 2 normal patients were excluded from analysis due to incomplete data of all investigated strain values; the remaining dataset utilized in this study had a total of 51 CP, 47 RCM, 53 and control (normal) patients.

### Data security

To maintain confidentiality and integrity of study data, all data was generated, stored, and transmitted using protected encryption measures. Specifically, stored on encrypted drives with industry standard encryption algorithms to prevent unauthorized access. All transmission of data both onsite and offsite was performed using secure protocols to prevent interception and tampering.

### Proposed framework

We propose a persistent homology workflow based on topological data analysis techniques to identify disease patterns from functional physiologic signals. The pipeline is outlined as follows (Fig. 1):

1. Data preprocessing—data is converted to an n-dimensional point cloud.
2. Persistent homology
  - a. TDA filtration—simplicial complexes are built upon the point cloud, from which topological invariants are extracted.
  - b. Persistent image—birth, and death features are transformed into a persistent image to develop feature vectors
3. Patient-specific motif—features are stored in a visual representation that can be directly interpreted by physicians/scientists or serve as input for machine learning
  - a. Direct physician analysis—doctors develop general understanding of underlying patterns visually apparent in motif
  - b. Machine learning modeling—various techniques applied for feature selection and classification of patients

### Speckle tracking echocardiography

Grayscale images from the apical 4-chamber (4ch) and midventricular parasternal short-axis views were evaluated with 2-dimensional speckle-tracking echocardiography (STE) by a licensed professional as described previously<sup>11,12</sup>. Stress-strain analysis can be segmentally divided into anatomically unique locations, 48 points for short-axis view and 49 points for apical four chamber view, that comprise the entirety of the left ventricular myocardium<sup>29</sup>. At each spatial location, various features, including longitudinal, circumferential, and radial strains and strain rates, were measured over one cardiac cycle. These measurements were stored in two text files, one for the 4ch view and one for the mid view, which serve as the raw data for this proof-of-concept study (Fig. 2).

### Pre-processing

For our purposes, we combined the 48 (short-axis view) and 49 (4ch) segmental strain locations into three functional groupings. The short axis was grouped into the anterior septal, inferior septal, and lateral wall; the apical four chamber view was grouped into the lateral wall, apex, and interventricular septum. Cubic spline interpolation of all strain tracings was performed to standardize patients' time points within one cardiac cycle. Analysis of segmental strain waveforms as aggregates instead of individual segments has been previously shown<sup>30</sup>. This approach attempts to remove the stochastic nature that analysis of each separate segment would precipitate. Instead, grouping by functional domains allows for averaging curves through a more physiologically relevant manner, specifically regarding the contractile nature and ultrastructural properties of cardiomyocytes within the myocardium<sup>31</sup>. Mean strain curves for each region were created by averaging the corresponding ventricular regions. Each mean strain curve was transformed using phase space reconstruction<sup>32</sup> using Python library `pypsr` (v0.0.1) (<https://github.com/hsharrison/pypsr>) with embedding dimension  $d = 3$  and time delay  $\tau = 2$ . This reconstruction transformed the echocardiography strain time series into a point cloud in a higher dimensional space (phase space) for subsequent TDA processing, representing a more complete picture of the dynamic system's linear time signal as a geometric shape, shown in Supplementary Figs. 1 and 2.

### TDA filtration and persistent image

The topological data analysis technique utilized in this study was persistent homology (PH), which is a TDA tool that describes the shape of data by extracting its topological invariants; the mathematical basis of PH is shown in works by Zomorodian and Carlsson<sup>33</sup>, Edelsbrunner et al.<sup>34</sup>, Ghrist et al.<sup>35</sup>, Bubenik et al.<sup>36</sup> and Adams et al.<sup>37</sup>. The PH concepts relevant to our workflow (Fig. 3) are described briefly in Supplementary File 1.

To accomplish the TDA filtration and conversion to persistent image, we performed this experiment on Python using TDA libraries, `rips` (v0.6.0)<sup>38</sup> and `persim` (v0.2.0), which are both freely available. To analyze the point clouds, simplicial complex filtration was built on the data points using `rips`. The PH of the filtration was extracted as birth and death values for dimension 0. To utilize this information in downstream machine learning tasks, a linear persistent image was created with pixel resolution of [1, 50] and variance of 0.005; the image bounds were automatically selected by the `persim` algorithm. A total of 18 persistent images each with 50 pixels were created, one for each combination of strain type and wall region; stacking these 18 linear images together generated a specific motif for each patient with a total of 18 by 50 pixels. The feature vector utilized in this study contained the intensities of these 900 persistent image pixels. The computational time for this workflow was approximately 1 min per patient data performed on a MacBook Pro laptop with 2.7 GHz with 16 GB ram without distributed computing.

### Patient-specific motif

Our workflow produced a visual representation indicative of the initial input that can be interpreted directly by physicians/scientists while also being capable of feeding into downstream machine learning tasks. The patient-specific motifs showcase the general trends of the disease conditions while maintaining individual patient characteristics, allowing the patients to be monitored for cardiac function changes.

## Statistical analysis

Orange data mining software (v3.28.0) was used for statistical analyses. The models were trained and evaluated using tenfold cross validation. Each model's receiver operating characteristic (ROC) curve, the area under the curve (AUC), sensitivity, and specificity were calculated to evaluate discriminatory power. To determine the significance of the AUC, a p-value less than 0.05 was considered statistically significant when using the pROC package in R (v4.0.3). To avoid an overfitting problem, Boruta<sup>39</sup> feature selection was performed through R (v4.0.3) statistical suite that applies a random forest algorithm to determine meaningful features to retain.

## Ethics declarations

All studies were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration. For one patient cohort, the institutional review board at Mount Sinai Medical Center approved the protocol, and for second patient cohort, the institutional review board at Mayo Foundation approved the protocol. Participants were included regardless of gender, race, ethnicity, or other demographic factors.

## Data availability

The datasets and computer code produced in this study can be provided upon request to the corresponding author(s). A link to the data is provided: [https://github.com/qahathaway/TDA\\_Persistent\\_Homology](https://github.com/qahathaway/TDA_Persistent_Homology).

Received: 21 June 2023; Accepted: 2 May 2024

Published online: 09 May 2024

## References

- Madani, A., Ong, J. R., Tibrewal, A. & Mofrad, M. R. K. Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. *NPJ Digit. Med.* **1**, 59 (2018).
- Domingos, J. S., Stebbing, R. V., Leeson, P. & Noble, J. A. *Structured Random Forests for Myocardium Delineation in 3D Echocardiography in Machine Learning in Medical Imaging: 5th International Workshop* 215–222 (Springer, 2014).
- Oktay, O. *et al.* Multi-input cardiac image super-resolution using convolutional neural networks. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016* (eds Ourselin, S. *et al.*) 246–254 (Springer, 2016).
- Voigt, J. U. & Cvijic, M. 2- and 3-Dimensional myocardial strain in cardiac health and disease. *JACC Cardiovasc. Imaging* **12**, 1849–1863 (2019).
- Singh, Y. *et al.* Algebraic topology-based machine learning using MRI predicts outcomes in primary sclerosing cholangitis. *Eur. Radiol. Exp.* **6**, 58 (2022).
- McGuirl, M. R., Volkening, A. & Sandstede, B. Topological data analysis of zebrafish patterns. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 5113–5124 (2020).
- Xia, K. & Wei, G. W. Persistent homology analysis of protein structure, flexibility, and folding. *Int. J. Numer. Methods Biomed. Eng.* **30**, 814–844 (2014).
- Nielson, J. L. *et al.* Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury. *Nat. Commun.* **6**, 8581 (2015).
- Bendich, P., Marron, J. S., Miller, E., Pieloch, A. & Skwerer, S. Persistent homology analysis of brain artery trees. *Ann. Appl. Stat.* **10**, 198–218 (2016).
- Yan, Y. *et al.* Gait rhythm dynamics for neuro-degenerative disease classification via persistence landscape-based topological representation. *Sensors (Basel)*. <https://doi.org/10.3390/s20072006> (2020).
- Amaki, M. *et al.* Diagnostic concordance of echocardiography and cardiac magnetic resonance-based tissue tracking for differentiating constrictive pericarditis from restrictive cardiomyopathy. *Circ. Cardiovasc. Imaging* **7**, 819–827 (2014).
- Sengupta, P. P. *et al.* Disparate patterns of left ventricular mechanics differentiate constrictive pericarditis from restrictive cardiomyopathy. *JACC Cardiovasc. Imaging* **1**, 29–38 (2008).
- Welch, T. D. *et al.* Echocardiographic diagnosis of constrictive pericarditis: Mayo Clinic criteria. *Circ. Cardiovasc. Imaging* **7**, 526–534 (2014).
- Masui, T., Finck, S. & Higgins, C. B. Constrictive pericarditis and restrictive cardiomyopathy: Evaluation with MR imaging. *Radiology* **182**, 369–373 (1992).
- Cerqueira, M. D. *et al.* Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. A statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. *Circulation* **105**, 539–542 (2002).
- Tabassian, M. *et al.* Machine learning of the spatio-temporal characteristics of echocardiographic deformation curves for infarct classification. *Int. J. Cardiovasc. Imaging* **33**, 1159–1167 (2017).
- Senapati, A. *et al.* Prognostic implication of relative regional strain ratio in cardiac amyloidosis. *Heart* **102**, 748–754 (2016).
- Negishi, K. *et al.* What is the primary source of discordance in strain measurement between vendors: Imaging or analysis? *Ultrasound Med. Biol.* **39**, 714–720 (2013).
- Mirea, O. *et al.* Variability and reproducibility of segmental longitudinal strain measurement: A report from the EACVI-ASE strain standardization task force. *JACC Cardiovasc. Imaging* **11**, 15–24 (2018).
- Edelsbrunner, Letscher, & Zomorodian., Topological persistence and simplification. *Discret. Comput. Geom.* **28**, 511–533 (2002).
- Chung, Y. M., Hu, C. S., Lo, Y. L. & Wu, H. T. A persistent homology approach to heart rate variability analysis with an application to sleep-wake classification. *Front. Physiol.* **12**, 637684 (2021).
- Saggar, M. *et al.* Towards a new approach to reveal dynamical organization of the brain using topological data analysis. *Nat. Commun.* **9**, 1399 (2018).
- Dindin, M., Umeda, Y. & Chazal, F. Topological data analysis for arrhythmia detection through modular neural networks. In *Canadian AI 2020—33rd Canadian Conference on Artificial Intelligence, Ottawa, Canada* (2020).
- Ignacio, P. S., Dunstan, C., Escobar, E., Trujillo, L. & Uminsky, D. Classification of single-lead electrocardiograms: TDA informed machine learning. In *18th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Boca Raton 1241–1246 (2019).
- Nicponski, J. & Jung, J.-H. Topological data analysis of vascular disease: A theoretical framework. *Front. Appl. Math. Stat.* <https://doi.org/10.3389/fams.2020.00034> (2020).
- Hwang, D. *et al.* Topological data analysis of coronary plaques demonstrates the natural history of coronary atherosclerosis. *JACC Cardiovasc. Imaging*. <https://doi.org/10.1016/j.jcmg.2020.11.009> (2021).
- Singh, Y. *et al.* Topological data analysis in medical imaging: Current state of the art. *Insights Imaging* **14**, 58 (2023).



28. Yang, Z. *et al.* Left ventricular strain-curve morphology to distinguish between constrictive pericarditis and restrictive cardiomyopathy. *ESC Heart Fail.* **8**, 4863–4872 (2021).
29. Dandel, M., Lehmkuhl, H., Knosalla, C., Suramclashvili, N. & Hetzer, R. Strain and strain rate imaging by echocardiography—Basic concepts and clinical applicability. *Curr. Cardiol. Rev.* **5**, 133–148 (2009).
30. Hensel, F., Moor, M. & Rieck, B. A survey of topological machine learning methods. *Front. Artif. Intell.* **4**, 681108 (2021).
31. Sengupta, P. P. *et al.* Left ventricular structure and function: Basic science for cardiac imaging. *J. Am. Coll. Cardiol.* **48**, 1988–2001 (2006).
32. Kennel, M. B., Brown, R. & Abarbanel, H. D. I. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **45**, 3403 (1992).
33. Zomorodian, A. & Carlsson, G. Computing persistent homology. *Discret. Comput. Geom.* **33**, 249–274 (2005).
34. Edelsbrunner, H., Letscher, D. & Zomorodian, A. Topological persistence and simplification. *Discret. Comput. Geom.* **28**, 511–533 (2000).
35. Ghrist, R. Barcodes: The persistent topology of data. *Bull. Am. Math. Soc.* **45**, 61–75 (2008).
36. Bubenik, P. Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.* **16**, 77–102 (2015).
37. Adams, H. *et al.* Persistence images: A stable vector representation of persistent homology. *J. Mach. Learn. Res.* **18**, 1–35 (2017).
38. Tralie, C., Saul, N. & Bar-On, R. Ripser py: A lean persistent homology library for python. *J. Open Source Softw.* <https://doi.org/10.21105/joss.00925> (2018).
39. Kursu, M. B. & Rudnicki, W. R. Feature selection with the boruta package. *J. Stat. Softw.* **36**, 1–13 (2010).

### Author contributions

Designing research studies (NKS, YS, QAH, PPS, NY), conducting experiments (NKS, YS, NY), acquiring data (NKS, YS, NY), analyzing data (NKS, YS, QAH, PPS, NY), writing the manuscript (NKS, YS, QAH, PPS, NY). The author Nanda K. Siva had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors read and approved the final manuscript.

### Competing interests

Partho P. Sengupta is a consultant to Heart Sciences, Ultromics, and Kencor Health. Quincy A. Hathaway is the Chief Science Officer for Aspirations LLC. The other authors have nothing to disclose.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-61201-4>.

**Correspondence** and requests for materials should be addressed to P.P.S. or N.Y.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024