Individualized patient risk stratification using machine learning and topological data analysis

Ng, A.C.T.; Delgado, V.; Bax, J.J.

JACC: CARDIOVASCULAR IMAGING VOL. 13, NO. 5, 2020

PUBLISHED BY ELSEVIER

EDITORIAL COMMENT

Individualized Patient Risk Stratification Using Machine Learning and Topological Data Analysis*

Arnold C.T. Ng, MBBS, PhD,[a,b] Victoria Delgado, MD, PhD,[c] Jeroen J. Bax, MD, PhD[c]

Originally associated with the 18th-century mathematician Leonhard Euler, topology is a discipline in mathematics that studies shapes. One of the key ideas in topology is that the properties of shapes are invariant (i.e., unchanging) when deformed (e.g., a circle, an ellipse, and a hexagon are all topologically identical in that they are all "loops," and any of these shapes can be obtained by deforming and stretching it). In topological data analysis (TDA), data points from large datasets are summarized into nodes, and the relationships between these nodes are connected by a line (also called an "edge"), thereby forming a large network of nodes connected by edges. This network has a particular "shape," and typical networks shapes are either "loops" (continuous circular segments) or "flares" (long linear segments) (1). Groups within these network shapes can be arbitrarily partitioned into segments with similar properties to identify patterns within the data, and they can be compared with other groups within the network using standard statistical techniques. Because these network shapes are invariant under deformations, one of the strengths of TDA is that it is less sensitive to noise and can detect patterns that can be missed by other analysis techniques such as principal-component analysis, multidimensional scaling, and cluster analysis (1).

TDA can be combined with machine learning, which involves computer algorithms that adjust and learn from input data. These algorithms can automatically divide a large dataset into clusters of smaller datasets such that those within the same cluster group are more similar to one another than to those in other clusters (e.g., echocardiograms from normal subjects vs. those of patients with heart failure) (2). In unsupervised machine learning, no a priori knowledge of the expected outcome is incorporated, and the computer algorithm uses only the content of the input data to discover any possible underlying structure. Such combined machine learning and TDA techniques can be used to identify novel classifications systems or offer further insights into the natural history of diseases. In the first ever application of TDA in cardiovascular research, Casaclang-Verzosa et al. (3) described the natural history of aortic stenosis, which has 2 distinct moderate stenosis phenotypic expressions as it progresses from mild to severe stenosis (i.e., moderate aortic stenosis with normal vs. reduced ejection fraction).

In this issue of *iJACC*, the same group (4) report on an evaluation of a large number of patients with a wide range of cardiac diseases at different stages of severity. Using unsupervised machine learning with TDA, 4 subgroups of patients were automatically identified with distinct differences in major adverse cardiac event (MACE) outcomes on the basis of standard echocardiographic parameters such as left ventricular ejection fraction, mass, and so on.

*Editorials published in *JACC: Cardiovascular Imaging* reflect the views of the authors and do not necessarily represent the views of *iJACC* or the American College of Cardiology.

From the [a]Department of Cardiology, Princess Alexandra Hospital, Brisbane, Australia; [b]The Faculty of Medicine, South Western Sydney Clinical School, The University of New South Wales, Sydney, Australia; and the [c]Department of Cardiology, Leiden University Medical Centre, Leiden, the Netherlands. The Department of Cardiology of the Leiden University Medical Center received unrestricted research grants from Abbott Vascular, Bioventrix, Biotronik, Boston Scientific, Edwards Lifesciences, Medtronic, and GE Healthcare. Dr. Bax has received speaker fees from Abbot Vascular. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the *JACC: Cardiovascular Imaging* author instructions page.

ISSN 1936-878X/$36.00 https://doi.org/10.1016/j.jcmg.2020.02.003

Identifying patients within each individual subgroup provided incremental risk stratification compared with traditional risk scoring such as New York Heart Association functional class.

The investigators should be commended as pioneers, bringing machine learning and state-of-the-art TDA into clinical cardiovascular research and developing novel risk stratification tools in the growing field of big data, electronic medical records, and artificial intelligence. The strengths of the present study are the large number of patients from different retrospective and prospective datasets, validation of their topological network shape using the prospective cohort, and including a longitudinal cohort to predict individual patient risk as they move from 1 subgroup to another as their disease severity changes over time. "Traditional" risk stratification presents a hazard ratio or an odds ratio for an "average" patient. In contrast, one of the most important clinical implications of this study is the ability of machine learning and TDA to automatically present individual patient risk profiles on the basis of similarities with other patients within the topological network.

However, several important questions remain unanswered. First, there are nodes that do not fit into the network, as demonstrated in the Central Illustration, and no information is provided on the characteristics of these patients. Therefore, not all patients' data can be automatically compressed into a node to fit the topological network.

Second, the data used to form the network are simple and crude echocardiographic measures. The data do not include newer imaging techniques such as strain imaging, identification of subclinical coronary atherosclerosis on cardiac computed tomography, and scarring or interstitial fibrosis on cardiac magnetic resonance. Many of these techniques are already routinely used in clinical practice. Moreover, the network had limited information on patient diagnosis and therapeutic interventions such as device therapies, revascularizations, and so on. These variables will clearly have an impact on MACE outcomes for individual patients.

Finally, the investigators suggest that heart failure should be viewed as a continuum instead of arbitrary divisions into reduced, midrange, or preserved left ventricular ejection fraction. However, a long line of research had demonstrated the futility of therapies such as beta-blockers, angiotensin-converting enzyme inhibitors, and so on, for patients with heart failure with preserved ejection fraction. This strongly suggests a different pathophysiology compared with those with reduced left ventricular ejection fraction. As such, one should not confuse the association between subgroups of patients (with a particular pattern of echocardiographic parameters) and the corresponding MACE rates with direct causality.

In summary, Tokodi et al. (4) used a "new" mathematical modeling technique (i.e., TDA) with unsupervised machine learning to map the "shape" of a large echocardiographic dataset. The shape of the network can identify patterns in our echocardiographic measurements (e.g., left ventricular ejection fraction, mass, filling pressures) associated with different MACE rates. This study represents an initial stepping-stone in developing more complex and comprehensive cardiac risk stratification models using unsupervised machine learning fed with large datasets. Using novel mathematical modeling techniques such as TDA, clinicians may be able to provide individualized risk stratification for patients at different stages of their disease.

ADDRESS FOR CORRESPONDENCE: Dr. Jeroen J. Bax, Department of Cardiology, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, the Netherlands. E-mail: j.j.bax@lumc.nl.

REFERENCES

1. Lum PY, Singh G, Lehman A, et al. Extracting insights from the shape of complex data using topology. Sci Rep 2013;3:1236.

2. Seetharam K, Shrestha S, Sengupta PP. Artificial intelligence in cardiovascular medicine. Curr Treat Options Cardiovasc Med 2019;21:25.

3. Casaclang-Verzosa G, Shrestha S, Khalil MJ, et al. Network tomography for understanding phenotypic presentations in aortic stenosis. J Am Coll Cardiol Img 2019;12:236–48.

4. Tokodi M, Shrestha S, Bianco C, et al. Interpatient similarities in cardiac function: a platform for personalized cardiovascular medicine. J Am Coll Cardiol Img 2020;13:1119–32.

KEY WORDS echocardiography, heart failure, machine learning