Editorial Comment

The Mechanics of Machine Learning: From a Concept to Value

Sirish Shrestha, MSc, and Partho P. Sengupta, MD, DM, *Morgantown, West Virginia*

We can only see a short distance ahead, but we can see plenty there that needs to be done.

—Alan Turing

Heart failure with preserved ejection fraction (HFpEF) is a complex syndrome with heterogeneous pathophysiology that occurs in approximately 50% of heart failure cases.[1] Although the presence of left ventricular diastolic dysfunction is an important pathophysiologic contributor, the development of multiple echocardiographic parameters and numerous iterations in consensus-driven guidelines has continuously challenged clinicians' ability to interpret diastolic function.[2] Moreover, with the advent of speckle-tracking techniques for quantification of myocardial motion and deformations, the sheer number of parameters has increased to provide additional information that is generally not available using traditional methods. And yet, despite the trove of information, there remains considerable uncertainty regarding steps toward integrating clinical and echocardiographic parameters in a way that is relevant in the clinical setting. To this end, the erudite work by Tabassian *et al.*[3] in this issue of *JASE* provides important steps and strengthens the available evidence toward addressing this knowledge gap.

Tabassian *et al.*[3] hypothesized that temporal and spatial variations in left ventricular myocardial velocities and deformation may provide salient information in the assessment of HFpEF. These may guide in the quantification of dyssynchronous contraction at rest and during stress to differentiate patients with HFpEF from healthy subjects. In their prospective observational study, they recruited 100 subjects who were healthy, had HFpEF, were breathless, or had hypertension on the basis of the current guidelines of the European Society of Cardiology.[4] They extracted velocity, strain, and strain rate curves from all 18 segments of the left ventricle both at rest and during submaximal exercise testing. Given the dense multidimensional data collected from each subject, the authors resorted to machine-learning approaches, a subset of artificial intelligence, to segregate phenotypic relationships for identifying the HFpEF syndrome. Although the concept of using machine learning for cardiac imaging is not new,[5-7] interest in using such analyses in routine clinical cardiology is growing because of the advent of a new generation of modern, scalable computing systems and algorithms that are capable of processing petabytes of data in real time.

WHAT IS MACHINE LEARNING?

Machine learning is based on algorithms that can learn from data without relying on rule-based programming. Although the concept of intelligent machines was first proposed by Alan Turing in 1947, machine learning came into its own as a scientific discipline only in the late 1990's. This was made possible by steady advances in digitization and inexpensive computing power that enabled data scientists and computer scientists to build sophisticated algorithms and models that can be trained for making predictions. For the medical profession, machine learning is now more than a buzzword; artificial intelligence and machine learning have played a huge role in sequencing large genomic data sets quickly and easily. The past 5 years have witnessed an exponential interest in the use of machine learning techniques for echocardiography, prompting national societies to rethink their basic assumptions about the relationships among imagers, data scientists, and information technology professionals and their respective roles.[8,9] The American College of Cardiology recently convened a think tank on the future of cardiac imaging and emphasized an urgent need for cardiologists to collaborate with big data scientists and bioinformatics experts. This was primarily done with a focus on expediting ways of using imaging data across multiple modalities to validate the value of cardiac imaging in improving health care outcomes.[9]

Machine-learning techniques can be generally categorized as one of four methods—supervised, unsupervised, semisupervised, or reinforcement learning methods—depending on the complexity of the data, heterogeneity of the disease, and the outcome of the analysis (Table 1). Supervised learning problems are usually divided into "regression" and "classification" problems. There is a profusion of algorithms; for example, a free and favored machine-learning library[11] lists dozens of algorithms for each group of classification, regression, and other popular tasks in the machine-learning universe. However, the choice depends on the type of data and the purpose. In a regression problem, the model tries to predict results within a continuous output, meaning that the model tries to map input variables to some continuous function. In a classification problem, instead, the model predicts results in a discrete output. Tabassian *et al.*[3] attempted to predict the placement of cases into one of the four groups (healthy, HFpEF, hypertensive, and breathless) in a supervised way. The authors chose distance-weighted *k*-nearest neighbor (KNN), which is built on a popular KNN algorithm that has been used for both a classification and regression supervised machine-learning techniques. Notably, the distance-weighted KNN algorithm can skillfully produce probabilities that can provide the degree to which data belong to a group. Moreover, distance-weighted KNN can handle smaller sample sizes and is also suitable for multiclass and binary classifications, as evident from the categorization of symptomatic and asymptomatic patients in the study.

To assess model performance, data are typically divided into training and testing sets for the algorithm to learn and validate the performance and accuracy of the model. The training set contains echocardiographic parameters and outcomes to build a model, whereas the testing set (for which the model is blinded to outcomes) is used to assess its predictive power. The algorithms learn better as they are exposed to larger data sets, but for investigations in which there

are limited data, other strategies may be required. For example, the present study by Tabassian *et al.*[3] had a modest sample size, and traditional splitting into training and testing sets would have led to suboptimal group divisions. In such situations, sophisticated techniques for creating training and testing splits are needed. Cross-validation makes more efficient use of data by dividing the data into, say, 10 equal parts. Then the model is run separately for each of the component parts, and the results of the 10 estimates of predictive accuracy can be averaged into an overall measure. There are other advanced methods that repeatedly and randomly divide multiple subsets to yield training and testing sets—typically several orders of magnitude times more than the conventional cross-validation method—to learn the pattern in the data. For example, the authors split the training and testing sets with 64 and 36 randomly selected cases, respectively, and repeated the process 100 times to ensure that the model did not overfit or was not biased toward a subset of the subjects.[3]

One must be careful, however, in applying supervised learning in heterogeneous diseases, such as HFpEF, because the interobserver variability may reduce the accuracy of the classification or regression. This can happen because of the intensive requirement of interaction with the data for collection and identification of the outcomes before the application of machine learning. Furthermore, the researchers' bias can also seep into the very data machine learning uses to train on, producing inaccurate predictions.

CAN LEARNING OCCUR WITHOUT SUPERVISION?

In contrast to supervised learning, unsupervised machine learning does not seek labeled outcomes in the data to assess accuracy. It instead comports to naturally occurring patterns within seemingly heterogeneous data[11] to identify similarities among patients across many parameters. Neither are there expected response data to predict or training and testing sets to split (although it can be done), but rather only an underlying pattern to discern. Therefore, the performance of the model can be subjective and specific to disease or phenotypes. Consequently, unsupervised learning can be used as a stand-alone method to cluster patients or diseases, compress image quality, or identify phenotypic relationship from the topology of the network. However, often the algorithm is used before the use of supervised learning as a data-preprocessing step for its aptness to distill noise and large-dimensional data in producing conspicuous conception.

Techniques such as clustering and dimensionality reduction methods have been widely used in exploration and identifying phenotypic variations in cardiac diseases that are highly heterogeneous. In cluster analysis, the aim is to segregate a data set into nonoverlapping groups that have some distinct defining properties or conceptual coherence. When patient data split naturally into such distinct groups, the clustering techniques can be powerful for finding phenotypic presentations. For example, Omar *et al.*[12] applied a clustering algorithm for integrating high-dimensional information from speckle-tracking strain data to obtain unique clusters of patients with varying severities of diastolic dysfunction and left ventricular filling pressures. Similar other studies have also been successful in applying clustering and dimensionality reduction technique in high-dimensional echocardiographic data to group patients with their homogenous counterparts.[13,14] Similarly, Tabassian *et al.*[3] relied on an unsupervised learning method known as principal-component analysis to manage the variable spectrum. Principal-component analysis is a popular unsupervised technique that combines similar parameters to form a set of distinct features that are independent of one another and arranged in a rank that ex-

Table 1 Types of learning methods[10]

Method	Definition
Supervised	Most common learning method, applied to tasks such as regression, classification, predictive modeling, and survival analysis. These have labels and outcomes defined and infers from training data. Examples include logistic regression, support vector machine, and neural networks.
Unsupervised	A learning method to observe similarities and relationships among groups and variables. Most prominent applications include clustering and dimensionality reduction. Generally not divided into training and testing sets (though possible), and response data are not available. Mostly used algorithms include hierarchical clustering, k-means clustering, and principal-component analysis.
Semisupervised	A learning method in which the data input has a mixture of labeled and unlabeled outcomes or classes. Data that are not completely classified are generally used. Examples include image and speech recognition.
Reinforcement	Learning method based on behavioral psychology. It updates on the basis of the feedback it receives from the choices it makes until the exploration of the environment produces "reward." It is currently being used in medical imaging analytics, disease screening, and personalized prescription selection.

plains the most variance in the data in a reduced spectrum of variable. This is an effective way to drop less informative parameters while retaining parameters that validate the motif in the data set. After all, managing a large number of features can be a complex task that can impede the way an algorithm learns to be generalizable. Specifically, after a certain point, the performance of the model will decrease with increasing number of features. This phenomenon is often referred to as "the curse of dimensionality." Each feature adds to a newer dimension of the disease. However, without increasing the number of patient samples, the growth in dimensions is associated with more separation in data points. Because of this separation with increasing dimensions, the solutions offered by the machine-learning model are often spurious and not generalizable to other data samples, a phenomenon referred to as overfitting. Thus, as well exemplified by Tabassian *et al.*, reducing the number of dimensions using methods such as principal-component analysis can curtail overfitting.

CAN UNSUPERVISED AND SUPERVISED LEARNING BE COMBINED?

Unsupervised learning algorithms are designed with the anticipation of capturing some useful unknown structure in data. This can aid improved performance on subsequent supervised learning tasks, without requiring more labels from experts. For example, Tabassian *et al.*[3] experienced a conundrum with their data set, which contained 36 segmental curves for each subject with 208 and 123 time points at rest and during exercise, respectively. They reduced their high-dimensional data to smaller dimensions using unsupervised learning

and applied a supervised classification algorithm to classify patients to one of the four groups. As evident from the analysis, reducing the data to manageable dimensions produced a modest overall average accuracy of 0.85 (sensitivity, 0.86; specificity, 0.82) for the data with strain rate. Although the authors elegantly addressed the challenging problem, several factors may affect the real-world generalizability of the model accuracy. First, a smaller sample size may not account for all patterns of heterogeneous presentations in HFpEF. Second, data imputation, a process to substitute missing data with mean, median, or predicted values, may introduce error and reduce the ability of the model to learn. It would have been interesting to assess the area under the curve had Tabassian *et al.* used regularization methods given the dimension and the sample size. Thus, future studies with a larger sample size may be required to assess the accuracy and the quality of the predictors.

WHAT IS THE SCOPE OF MACHINE LEARNING IN HFₚEF?

Diverse diseases such as HFpEF with varying etiologies and pathophysiology may benefit from automatic data collection from electronic health records and cardiac images using machine-learning tools, supplementing it with clinical and factor data, genomics, proteomics,[15] and metrics from wearable devices. Applying deep learning, a novel artificial intelligence technique based on the neural network (a computer network inspired by the layered architecture of neurons in the human brain), also may provide ample opportunities for cardiologists to use such data for classification and phenotypic identification of the disease; however, the need for large data sets is critical for deep-learning algorithms to identify hidden signals and insights. Provided sufficient data, deep learning may provide diagnostic and predictive capabilities in many complex heterogenous cardiovascular diseases and has already shown promising results in drug discovery and digital pathology.[16] Making it more interesting, transfer learning is making headway in deep learning, whereby pretrained models used in one task can be applied in other cardiovascular disease prediction and classification tasks. Recently machine-learning techniques have also been combined with visualization techniques for making sense of data. A novel method called topological data analysis has flourished and can help cardiologists view data in a feature or patient space to ascertain pattern and similarity networks that were previously considered intricate.[17-19] It effectively manages high-dimensional and noisy data to find salient patterns within and has been successfully illustrated to differentiate heart failure phenotypes.[20,21]

Regardless of the algorithms or machine-learning methods used, this ubiquitous technology is gradually expanding in cardiology, which could help automate mundane tasks via image classification or parameter quantization. As machine-learning techniques become more routine in cardiology, more advanced techniques such as deep learning can improve work flow and ensure veracity and integrity in scientific data. Thus, as our observational and health care data grow, computational prowess increases, and ideas continue to germinate, we may not achieve solutions to every aspect of cardiovascular disease, but we have gained a powerful ally to explain, explore, and strive toward better patient outcomes. The study by Tabassian *et al.*[3] is thus symbolic of such a paradigm shift that emphasizes integration, analysis, and representation of data for knowledge discovery and patient care.

REFERENCES

1. Bhatia RS, Tu JV, Lee DS, Austin PC, Fang J, Haouzi A, et al. Outcome of heart failure with preserved ejection fraction in a population-based study. N Engl J Med 2006;355:260-9.

2. Nagueh SF, Smiseth OA, Appleton CP, Byrd BF, Dokainish H, Edvardsen T, et al. Recommendations for the evaluation of left ventricular diastolic function by echocardiography: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr 2016;29:277-314.

3. Tabassian M, Sunderji I, Erdei T, Sanchez-Martinez S, Degiovanni A, Trino P, et al. Diagnosis of heart failure with preserved ejection fraction: machine learning of spatiotemporal variations in left ventricular deformation. J Am Soc Echocardiogr 2018;31:1272-84.

4. Windecker S. The European Society of Cardiology clinical practice guidelines. Eur Heart J 2017;38:2696-7.

5. Mahmoud A, Bansal M, Sengupta PP. New cardiac imaging algorithms to diagnose constrictive pericarditis versus restrictive cardiomyopathy. Curr Cardiol Rep 2017;19:43.

6. Narula S, Shameer K, Salem Omar AM, Dudley JT, Sengupta PP. Machine-learning algorithms to automate morphological and functional assessments in 2D echocardiography. J Am Coll Cardiol 2016;68:2287-95.

7. Sengupta PP, Huang YM, Bansal M, Ashrafi A, Fisher M, Shameer K, et al. Cognitive machine-learning algorithm for cardiac imaging; a pilot study for differentiating constrictive pericarditis from restrictive cardiomyopathy. Circ Cardiovasc Imaging 2016;9:e004330.

8. Sengupta PP. Intelligent platforms for disease assessment: novel approaches in functional echocardiography. JACC Cardiovasc Imaging 2013;6:1206-11.

9. Douglas PS, Cerqueira MD, Berman DS, Chinnaiyan K, Cohen MS, Lundbye JB, et al. The Future of Cardiac Imaging: Report of a Think Tank Convened by the American College of Cardiology. JACC Cardiovasc Imaging 2016;9:1211-23. Elsevier.

10. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: are we there yet? Heart 2018;104:1156-64.

11. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in {P}ython. J Mach Learn Res 2011;12:2825-30.

12. Omar AMS, Narula S, Abdel Rahman MA, Pedrizzetti G, Raslan H, Rifaie O, et al. Precision phenotyping in heart failure and pattern clustering of ultrasound data for the assessment of diastolic dysfunction. JACC Cardiovasc Imaging 2017;10:1291-303.

13. Lancaster MC, Salem Omar AM, Narula S, Kulkarni H, Narula J, Sengupta PP. Phenotypic clustering of left ventricular diastolic function parameters: patterns and prognostic relevance. JACC Cardiovasc Imaging 2018;11:935-7.

14. Sanchez-Martinez S, Duchateau N, Erdei T, Kunszt G, Aakhus S, Degiovanni A, et al. Machine learning analysis of left ventricular function to characterize heart failure with preserved ejection fraction. Circ Cardiovasc Imaging 2018;11:e007138.

15. Shah SJ. Innovative clinical trial designs for precision medicine in heart failure with preserved ejection fraction. J Cardiovasc Transl Res 2017;10:322-36.

16. Johnson KW, Shameer K, Glicksberg BS, Readhead B, Sengupta PP, Björkegren JLM, et al. Enabling precision cardiology through multiscale biology and systems medicine. JACC Basic Transl Sci 2017;2:311-27.

17. Li L, Cheng W-Y, Glicksberg BS, Gottesman O, Tamler R, Chen R, et al. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. Sci Transl Med 2015;7:311ra174.

18. Nielson JL, Paquette J, Liu AW, Guandique CF, Tovar CA, Inoue T, et al. Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury. Nat Commun 2015;6:8581.

19. Lum PY, Singh G, Lehman A, Ishkanov T, Vejdemo-Johansson M, Alagappan M, et al. Extracting insights from the shape of complex data using topology. Sci Rep 2013;3:1236.

20. Tokodi M, Salem Omar AM, Cummins M, Shrestha S, Bianco C, Amin AH, et al. 2018 ASE 29th Annual scientific Sessions. J Am Soc Echocardiogr 2018;31:B3.

21. Tokodi M, Omar A, Cummins M, Shrestha S, Bianco C, Amin AH, et al. Extracting knowledge from geometric shape of echocardiography data: isolating phenotypic traits within the systolic-diastolic dysfunction continuum. J Am Soc Echocardiogr 2018;31:B3 (abstract).