JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY VOL. 71, NO. 23, 2018

JACC FOCUS SEMINAR: FUTURE TECHNOLOGY OF CARDIOVASCULAR CARE

JACC REVIEW TOPIC OF THE WEEK

Artificial Intelligence in Cardiology

Kipp W. Johnson, BS,[a,b] Jessica Torres Soto, MS,[c,d,e] Benjamin S. Glicksberg, PhD,[a,b,f] Khader Shameer, PhD,[g] Riccardo Miotto, PhD,[a,b] Mohsin Ali, MPhil,[a,b] Euan Ashley, MBChB, DPhil,[c,d,e] Joel T. Dudley, PhD[a,b]

ABSTRACT

Artificial intelligence and machine learning are poised to influence nearly every aspect of the human condition, and cardiology is not an exception to this trend. This paper provides a guide for clinicians on relevant aspects of artificial intelligence and machine learning, reviews selected applications of these methods in cardiology to date, and identifies how cardiovascular medicine could incorporate artificial intelligence in the future. In particular, the paper first reviews predictive modeling concepts relevant to cardiology such as feature selection and frequent pitfalls such as improper dichotomization. Second, it discusses common algorithms used in supervised learning and reviews selected applications in cardiology and related disciplines. Third, it describes the advent of deep learning and related methods collectively called unsupervised learning, provides contextual examples both in general medicine and in cardiovascular medicine, and then explains how these methods could be applied to enable precision cardiology and improve patient outcomes. (J Am Coll Cardiol 2018;71:2668–79) © 2018 The Authors. Published by Elsevier on behalf of the American College of Cardiology Foundation. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

The promise of artificial intelligence (AI) and machine learning in cardiology is to provide a set of tools to augment and extend the effectiveness of the cardiologist. This is required for several reasons. The clinical introduction of data-rich technologies such as whole-genome-sequencing and streaming mobile device biometrics will soon require cardiologists to interpret and operationalize information from many disparate fields of biomedicine (1–4). Simultaneously, mounting external pressures in medicine are requiring greater operational efficiency from physicians and health care systems (5). Finally, patients are beginning to demand faster and more personalized care (6,7). In short, physicians are being inundated with data requiring more sophisticated interpretation while being expected to perform more efficiently. The solution is machine learning, which can enhance every stage of patient care—from research and discovery to diagnosis to selection of therapy. As a result, clinical practice will become more efficient, more convenient, more personalized, and more effective. Furthermore, the future's data will not be collected solely within the health care setting. The proliferation of mobile

From the [a]Institute for Next Generation Healthcare, Mount Sinai Health System, New York, New York; [b]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York; [c]Division of Cardiovascular Medicine, Stanford University, Palo Alto, California; [d]Departments of Medicine, Genetics, and Biomedical Data Science, Stanford University, Palo Alto, California; [e]Center for Inherited Cardiovascular Disease, Stanford University, Palo Alto, California; [f]Institute for Computational Health Sciences, University of California, San Francisco, California; and the [g]Department of Information Services, Center for Research Informatics and Innovation, Northwell Health, New Hyde Park, New York. Dr. Dudley is supported by the following grants from the National Institutes of Health: National Institute of Diabetes and Digestive and Kidney Diseases grant R01DK098242; National Cancer Institute grant U54CA189201; Illuminating the Druggable Genome; Knowledge Management Center sponsored by the National Institutes of Health Common Fund; National Cancer Institute grant U54-CA189201-02; and the National Center for Advancing Translational Sciences and Clinical and Translational Science Award UL1TR000067. Dr. Shameer has received consulting fees or honoraria from McKinsey, Google, LEK Consulting, Parthenon-EY, Philips Healthcare, and Kencore Health. Dr. Dudley has received consulting fees or honoraria from Janssen Pharmaceuticals, GlaxoSmithKline, AstraZeneca, and Hoffman-La Roche; is a scientific advisor to LAM Therapeutics, NuMedii, and Ayasdi; and holds equity in NuMedii, Ayasdi, and Ontomics. Dr. Ashley is founder of Personalis Inc. and Deepcell Inc; and is an advisor to Genome Medical and SequenceBio. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose. Mort Naghavi, MD, served as Guest Editor for this paper.

Manuscript received September 12, 2017; revised manuscript received March 1, 2018, accepted March 5, 2018.

Listen to this manuscript's audio summary by *JACC* Editor-in-Chief Dr. Valentin Fuster.

ISSN 0735-1097 https://doi.org/10.1016/j.jacc.2018.03.521

sensors will allow physicians of the future to monitor, interpret, and respond to additional streams of biomedical data collected remotely and automatically. In this technology corner, we introduce common methods for machine learning, review several selected applications in cardiology, and forecast how cardiovascular medicine will incorporate AI in the future (Central Illustration).

HOW DO ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING RELATE TO STATISTICS?

Physicians have long needed to identify, quantify, and interpret relationships among variables to improve patient care. AI and machine learning comprise a variety of methods that allow computers to do just this, by algorithmically learning efficient representations of data. Here, we use the terms "artificial intelligence" and "machine learning" more or less synonymously, although more precisely machine learning can be understood as a set of techniques to enable AI. The difference between classical machine learning and classical statistics is less one of methodology than one of intent and culture. The primary focus of statistics is to conduct inference about sample or population parameters, whereas machine learning focuses on algorithmically representing data structure and making predictions or classifications. These 2 ambitions are often intertwined. Thus, we do not place a definite boundary between classical statistics and machine learning methods and instead view them as analogous but often applied to answer different questions.

WHY DOES CARDIOLOGY NEED ARTIFICIAL INTELLIGENCE?

AI emerged because more familiar algorithms can often be improved on for real-world tasks. Consider the case of logistic regression. To enable statistical inference such as estimation of coefficients and p values, this model requires a number of strong assumptions (e.g., independence of observations and no multicollinearity among variables). When logistic regression is used for other purposes, the assumptions that enable statistical inference may be unrelated to the goal and can hinder the model's performance. In contrast, machine learning algorithms are typically used without making as many assumptions of the underlying data. Although this approach hinders the possibility for traditional statistical inference, it results in algorithms that generally are more accurate for prediction and classification. Thus, cardiovascular medicine can benefit from the incorporation of AI and machine learning.

SUPERVISED LEARNING: CLASSIFICATION AND PREDICTION

Machine learning strategies can be broadly split into either unsupervised or supervised learning. These have different goals. Unsupervised learning focuses on discovering underlying structure or relationships among variables in a dataset, whereas supervised learning often involves classification of an observation into 1 or more categories or outcomes (e.g., "Does this electrocardiogram represent sinus rhythm or ventricular fibrillation?"). Supervised learning thus requires a dataset with predictor variables ("features" in machine learning parlance) and labeled outcomes. In medicine, predictive modeling is often performed when observations have labels such as "cases" or "controls," and these observations are paired to associated features such as age, sex, or clinical variables.

FEATURE SELECTION

Feature selection is essential for predictive modeling, and machine learning is particularly useful for it. Consider the example of a physician who wishes to predict whether a patient with congestive heart failure will be readmitted to the hospital within 30 days of the index admission. This is a difficult problem where machine learning techniques have been shown to improve on traditional statistical methods (8,9). Our hypothetical clinician possesses a large but "messy" electronic health record (EHR) dataset (Figure 1). Typically, EHRs include variables such as International Classification of Diseases-ninth revision and tenth revision billing codes, medication prescriptions, laboratory values, physiological measurements, imaging studies, and encounter notes. It is difficult to decide a priori which variables should be included in a predictive model. Fitting a logistic regression model is, in fact, algebraically impossible when there are more independent variables than observations. Techniques such as univariate significance screening (i.e., including independent variables only if each is associated with the outcome in univariate analyses) or forward step-wise regression are commonly used. Unfortunately, these methods lead to models that do not tend to validate in other datasets and are poorly generalizable to patients (10,11). Furthermore, there are often complex interactions between variables. For example, 1 drug may significantly interact with another drug only if

ABBREVIATIONS AND ACRONYMS

AI = artificial intelligence

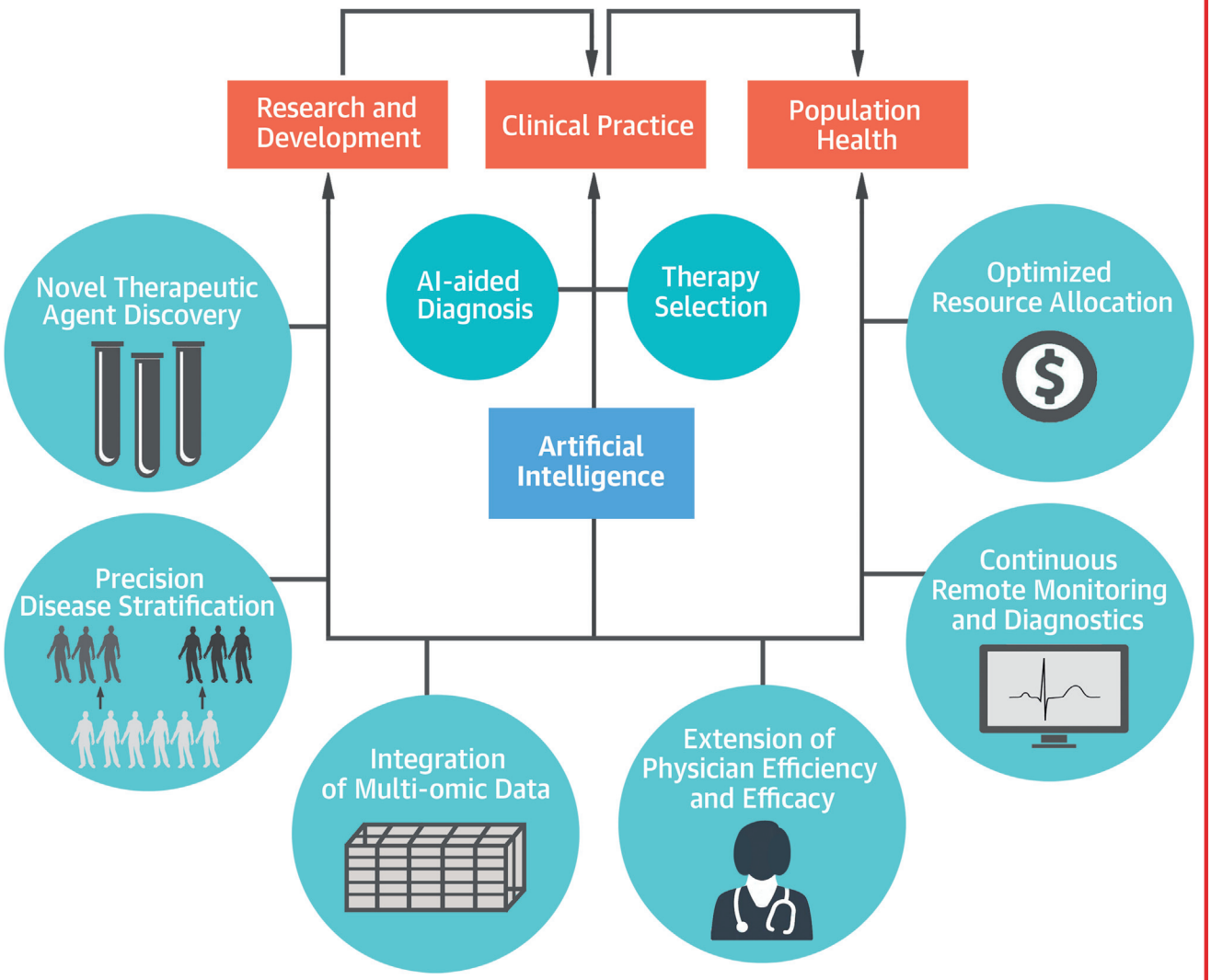
CNN = convolutional neural network

EHR = electronic health record

RNN = recurrent neural network

SVM = support vector machine

CENTRAL ILLUSTRATION **Role of Artificial Intelligence in Cardiovascular Medicine**



Johnson, K.W. et al. J Am Coll Cardiol. 2018;71(23):2668–79.

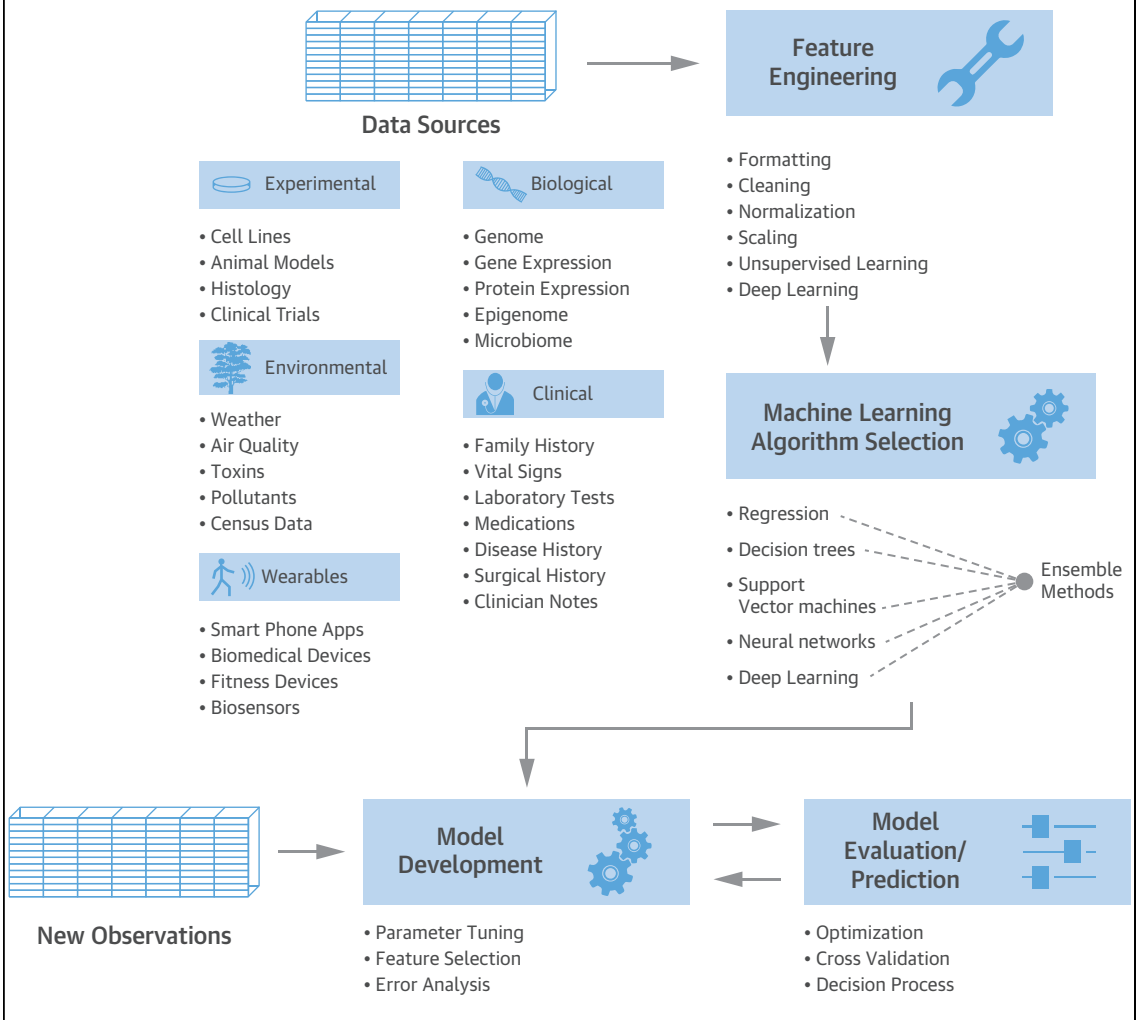
The incorporation of artificial intelligence (AI) into cardiovascular medicine will affect all aspects of cardiology, from research and development to clinical practice to population health. This illustration demonstrates selected applications within all 3 domains of cardiovascular care.

other conditions are present. The quantity and quality of such interactions are difficult to describe using traditional methods. With machine learning, we can capture and use these complex relationships. Features engineered by unsupervised learning are also often incorporated into supervised learning models. Churpek et al. (12) demonstrated the utility of machine learning feature selection in their paper comparing methods for prediction of clinical deterioration on the wards. Using demographics, laboratory values, and vital signs, these investigators found that a variety of different algorithms outperformed logistic regression to a clinically significant extent. An overview of several machine learning algorithms is presented later.

PROBLEMS IN BIOMEDICAL MACHINE LEARNING

Of course, supervised machine learning is not a panacea for prediction tasks. Even a perfect model is limited by the quality and magnitude of signal in the dataset from which it is trained. This is an important idea—even with a perfect algorithm, the model can

FIGURE 1 Overview of the Machine Learning Workflow



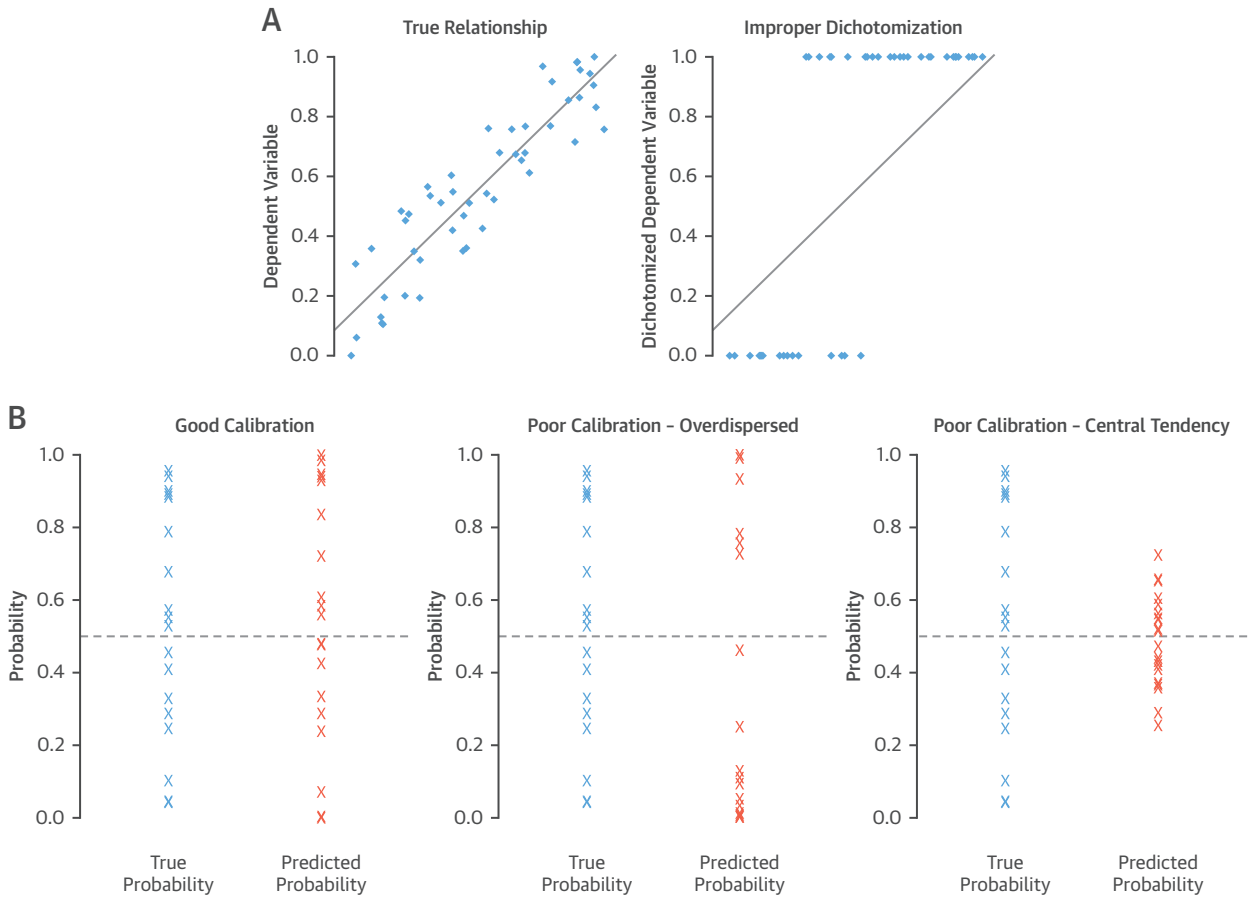
The central promise of machine learning is to incorporate data from a variety of sources (clinical measurements and observations, biological -omics, experimental results, environmental information, wearable devices) into sensible models for describing and predicting human disease. The typical machine learning workflow begins with data acquisition, proceeds to feature engineering and then to algorithm selection and model development, and finally results in model evaluation and application.

only be as good as the relevant information in the dataset. For example, using approximately 250 variables representing demographics, socioeconomic status, medical history, clinical symptoms, vital signs, laboratory values, and discharge interventions, Frizzell et al. (13) found that machine learning algorithms were unable to predict 30-day readmission better than logistic regression. In fact, all models performed only marginally better than a random classifier. Although the study by Frizzell et al. (13) is methodologically excellent, the findings are potentially limited by a dataset that does not contain many strong predictors of heart failure readmission. For example, although these investigators did include variables to describe socioeconomic status, it remains difficult to code and quantify social determinants of health, which seem to be highly important for hospital readmission. This limitation applies to both classical statistical modeling and machine learning methods.

DICHOTOMANIA

Clinicians generally work with dichotomized outcomes (e.g., "Should we give this patient a statin or not?") (14). However, framing clinical and scientific questions like this in some cases is imprecise and is called "improper dichotomization" (15,16). Two cases

FIGURE 2 Two Common Pitfalls in Predictive Modeling



(A) Visual demonstration of the concept of improper dichotomization on a dependent variable. Improper dichotomization obfuscates continuous relationships between predictors and response variables. **(B)** Concept of "calibration" in predicted probabilities in a supervised learning model. Because many machine learning tasks are framed as binary classification, the calibration of predicted probabilities is often underappreciated. Proper calibration of predicted probabilities is often just as important as accurate binary classification because reduction of probabilities to binary classifications can be understood as a form of improper dichotomization.

illustrate this point. First, consider a treatment paradigm such as the U.S. Preventive Services Task Force primary prevention statin recommendation guidelines, which incorporate a 10% threshold for 10-year cardiovascular risk as 1 of 3 criteria when grading the evidence for whether patients should be advised to take statins for primary cardiovascular disease prevention. Creating hard cutoffs for continuous outcomes (e.g., 10-year cardiovascular risk) leads to problems for individuals on the boundaries of the classification rule. To continue with the previous example, there may be only small differences for patients with a 9.5% predicted risk versus those with a predicted 10.5% predicted risk, yet a hypothetical dichotomous clinical recommendation machine using a 10% threshold as its basis could lead to different clinical plans for these "medium-risk" and "high-risk" patients that may or may not be appropriate. We instead recommend treating patients on the basis of their personalized 10-year risk, as many physicians already intuitively do, instead of considering them part of discretized blocks of patients who are at different risks according to a dichotomized categorization.

Furthermore, improper dichotomization reduces the precision of predictive models. Consider the example of a clinical trial of a disease biomarker with normally distributed values. A treatment changes biomarker values, and the amount of change in the biomarker is measured. Next, the researchers choose to dichotomize those patients with changed biomarker levels in the top half of change as "responders" and the bottom half of change as "nonresponders." This decision reduces the precision of the biomarker study to

TABLE 1 Brief Overview of 3 Common Supervised and Unsupervised Learning Algorithm Classes*

Example Algorithm Class	Advantages	Disadvantages	Example Application (Ref. #)
Supervised Learning Goals: Prediction of outcome, classification of observation, estimation of a parameter			
Regularized regression	Straightforward and automatic solution to high-dimensional problems Familiar interpretations for relationship of variables to outcomes	For groups of correlated features, arbitrary selection of single feature (LASSO)	Construction of a predictive model for acute myocardial infarction by using proteomic measurements and clinical variables (18)
Ensembles of decision trees	Often best "off-the-shelf" algorithm for prediction or classification Feature selection and variable importance assessment are built in	More useful for prediction than for descriptive analysis of dataset and variables Tendency to overfit data	Prediction of cardiovascular event risk (19)
Support vector machines	Transforms linear classifiers into nonlinear classifiers with the "kernel trick" Often makes highly accurate predictions	Performs nonprobabilistic classification by default Computation can be difficult in high-dimensional space	Prediction of in-stent restenosis from plasma metabolites (22)
Unsupervised Learning Goals: Discovery of hidden structure in a data, exploration of relationships between variables. Features discovered by unsupervised learning can often be incorporated into supervised learning models			
Deep learning algorithms	Current state-of-the art method for feature engineering; features are often used as input for supervised learning model Wide interest across industry and academia; rapidly developing software ecosystems	Computationally expensive to train Requires a large dataset to train the model Model interpretability can be difficult	Construction of predictive representations of patients in an unsupervised fashion from electronic health records (36)
Tensor factorization	Natural incorporation of multimodal and multidimensional data	Modest number of applications thus far in published cardiovascular reports Choice of factorization algorithm is crucial for results	Subtyping of congestive heart failure with preserved ejection fraction (34)
Topological data analysis	Interpretable clustering and discovery of variable relationships	Software ecosystem less mature than for other methods Commercial offerings require licensing agreement	Subtyping of type 2 diabetes mellitus from electronic medical records (35)

*Deep learning is included as an unsupervised learning method; however, many of the most notable applications of deep learning are those that use features learned using deep neural networks as inputs to supervised learning models. In fact, the final neural network layer in a deep learning model is often simply a classification layer, and in such a case deep learning models may be considered to be an example of supervised learning.

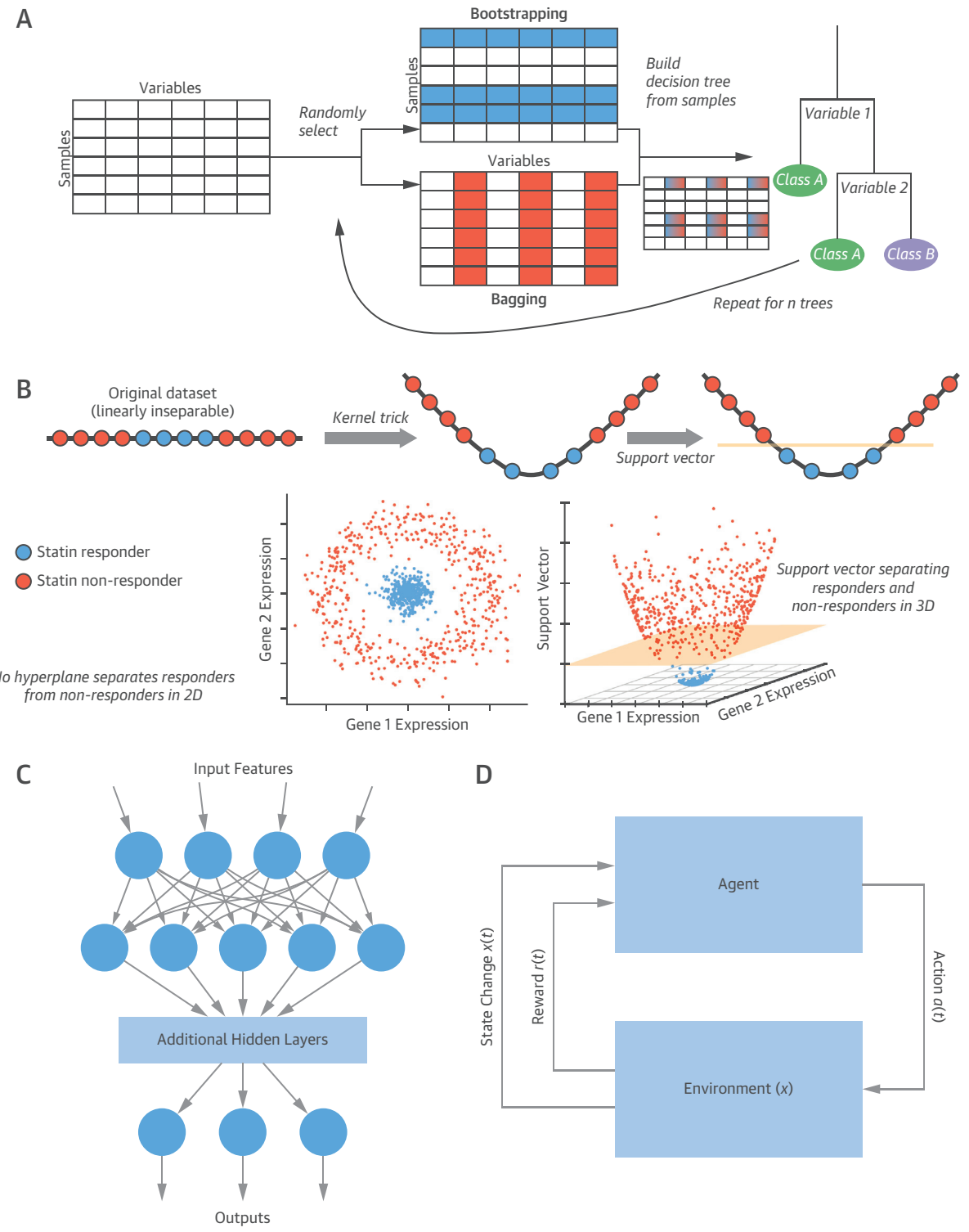
LASSO = least absolute shrinkage and selection operator.

64% of the value attainable by using raw, nondichotomized numbers instead (17). This issue is so widespread in biomedical publications that it is sometimes facetiously referred to as "dichotomania" by statisticians (17). Essentially, dichotomizing continuous data leads to loss of useful information about the strength of relationships and thus leads to a loss of power (**Figure 2A**). Instead, it is preferable to predict individual patient probabilities instead of making binary classifications. However, probabilities are useful only when they are accurate—consider the models validated by Kolek et al. (18) to predict atrial fibrillation from electronic health records. Although the models performed moderately well at classifying patients into low-risk or high-risk groups, predicted probabilities in each group were respectively too low or too high. This is called poor model calibration, which often occurs when standard regression techniques are used to model rare events (19) (**Figure 2B**). Better-calibrated prediction of outcome probabilities is an area where machine learning algorithms could provide clinical benefit.

A BRIEF SURVEY OF SUPERVISED MACHINE LEARNING ALGORITHMS IN CARDIOLOGY

Ultimately, supervised machine learning is the attempt to model how independent variables relate to a dependent variable (**Table 1**). In machine learning, one must choose a strategy (by selecting a particular

FIGURE 3 Visual Representation of Some Common Algorithms in Machine Learning



(A) Random forests incorporate both bootstrapping (selection of a subset of samples) and bagging (selection of a subset of predictive variables) for each individual tree. **(B)** Support vector machines. In binary classification, a support vector machine finds a hyperplane that separates classes. The "kernel trick" projects input data to a higher dimension before an ensuing support vector is computed. **(C)** Deep learning models comprise layers of stacked neurons that can be used to learn complex functions. **(D)** Reinforcement learning algorithms are used to train the action of an agent on an environment.

algorithm) to discover these relationships. This section highlights several algorithms that may be used in cardiovascular settings and provides a summary of supervised and unsupervised algorithms.

REGULARIZED REGRESSION. Imagine a situation where you have dozens, hundreds, or thousands of variables collected on just a few patients. You wish to decipher which variables are actually of relevance. As noted earlier, in this "big p, small n" (i.e., large number of features relative to a small number of samples) situation, a potential solution is a class of techniques called regularized regression. In this context, regularization means the introduction of additional constraints to decrease model complexity, thus allowing the model to generalize better to other data. Some common forms of regularized regression are called LASSO (least absolute shrinkage and selection operator) regression, ridge regression, and elastic net regression. To give an example of the benefits of regularization, Halim et al. (20) used logistic elastic net regression when combining proteomic measurements with clinical variables to predict the incidence of myocardial infarction or death. As these investigators noted, elastic net regression allowed them to evaluate many independent features in a novel way and ultimately helped them find candidate proteomic biomarkers for cardiovascular event risk.

TREE-BASED METHODS. Tree-based methods are a widely applied set of powerful but deceptively simple algorithms. Clinicians will find these useful because they are often referred to as the best "off-the-shelf" machine learning algorithm, and they are often among the first algorithms that should be used. In contrast to regularized regression, tree methods are especially useful when the data are "tall," that is, when there are many observations for a few variables. Tree-based methods use different subsets of the data repeatedly to build a final, complex, nonlinear model. In a clinical example, Weng et al. (21) used trees and another technique called gradient boosting to predict cardiovascular event risk in a sample of approximately 380,00 patients in the United Kingdom. These investigators found that the 2 machine learning algorithms outperformed the American College of Cardiology and American Heart Association risk algorithm by 1.7% and 3.6%, respectively.

The major problem with simple decision trees in practice is that they tend to overfit data—simple decision trees are high-variance learners and do not generalize well to other datasets. Two methods to address this issue are called bootstrapping and bagging (**Figure 3A**). These methods create many different decision trees, each of which is a weaker model than a single decision tree. Bootstrapping involves only taking a random sample of the observations before each decision tree is built. Random forests are a popular modification of trees: at each if-then step of the tree building process, only a randomly sampled subset of the variables is shown to the building algorithm. This is called bagging. When the outputs of the many weak individual learners (individual trees) in a random forest are aggregated together, they tend to perform very well. This is analogous to a team of medical providers, each with a different area of training (e.g., cardiology, gastroenterology, surgery) consulting together to treat a complex patient. Each physician will notice different features of the patient's presentation, and their combined treatment decision would often be better than a single physician's decision alone. We have previously used such "ensemble" methods to automate analysis of echocardiography imaging (22,23).

SUPPORT VECTOR MACHINES. Support vector machines (SVMs) comprise another widely used machine learning algorithm in the cardiovascular domain (**Figure 3B**). Clinicians may find SVMs useful because although relatively straightforward, they can capture complex nonlinear relationships. In a binary classification problem, SVMs map input observations into a higher-dimensional space and then attempt to construct a "hyperplane" that linearly separates the 2 classes. Cui et al. (24) demonstrated the usefulness of SVMs, by predicting in-stent restenosis with 90% accuracy from plasma metabolite levels. SVMs have 2 major downsides. First, they perform non-probabilistic classification (25). This means that, by default, SVMs work on dichotomized outcomes. As noted earlier, this is sometimes a problem; however, secondary methods to compute probabilistic outcomes (called Platt scaling or isotonic regression) from SVMs do exist. Second, similar to linear regression, computation of the input observations in a very high-dimensional space (i.e., when there are many variables) can be difficult or impossible.

UNSUPERVISED LEARNING, NEURAL NETWORKS, AND DEEP LEARNING

NEURAL NETWORKS AND DEEP LEARNING. Neural networks are machine learning models inspired by the organization of the human brain. The earliest application of neural networks in cardiology dates to at least 1995 (26,27). These models consist of nodes called neurons arranged in a network layout (28). The first level of nodes points into another layer of nodes

in the network called a "hidden layer," and there may be 1 or multiple hidden layers. A neuron in the hidden layer is activated when input neurons pass a large enough value to trigger the neuron, much like a biological neuron. Activated neurons continue to pass a value to the next layer of neurons until the final "output layer" of neurons is reached.

Deep learning is a powerful method premised on learning complex hierarchical representations from the data that constitute multiple levels of abstraction. Clinicians should understand that deep learning models are quickly becoming the state-of-the-art method and will enable the coming future applications of AI. In fact, deep learning models already underpin many of the features of modern technology we currently use, from automatic facial recognition in images in photographs uploaded to Facebook to the technology that allows Amazon's Alexa and Apple's Siri to perform high-quality voice recognition. Deep learning models are essentially neural networks with many layers of intermediate "hidden" neurons. Practical deep learning emerged only in past few years, in part because of the advent of graphics processing unit (GPU)-based parallel processing. Interestingly, a driving force behind this hardware technology is the 3-dimensional graphics company NVidia, which makes GPUs that are often used for deep learning.

WHAT MAKES DEEP LEARNING COMPELLING? Deep learning models use many hidden layers of neurons to produce increasingly abstracted, nonlinear representations of the underlying data (**Figure 3C**). This so-called "representation learning" is perhaps the most important part of a deep neural network—after the representations are learned, final output nodes are often used as inputs to a logistic regression model or SVM for the final regression or classification. Two of the most common forms of deep learning models for supervised learning are called convolutional neural networks (CNNs) and recurrent neural networks (RNNs). The difference between CNNs and RNNs is chiefly how layers of nodes are designed. There exists an enormous variety of deep neural network architectures in addition to these 2 methods. LeCun, Bengio, and Hinton (29) provide an excellent introduction to deep learning.

CNNs are similar to fully connected neural networks, also constructed of neurons that have learnable weights and biases. What makes them powerful is the ability to create local connectivity across an image or a signal. These simple local connections have non-linear activation functions that transform the representation into a higher, slightly more abstract representation. In addition, shared weights across layers, layer pooling, and the ability to use many hidden layers allow for learning of very complex functions. Conversely, RNNs are well suited for sequential data such as speech and language. RNNs are composed of an additional hidden state vector that contains "memory" about the history of data previously observed.

DEEP LEARNING IN CARDIOLOGY. In contrast to other technological fields, deep learning in health care is still developing, and its applications thus far to cardiology are rather limited (30,31). The earliest commercial applications of deep learning were for computer vision, or the computational analysis of images. Similarly, many of the initial biomedical applications of AI have been in the domain of image processing. For example, Gulshan et al. (32) harnessed a CNN to detect diabetic retinopathy from a database of 128,00 retinal images. These investigators obtained a sensitivity of 97.5% and specificity of 96.1% when compared with a gold standard classification by 7 to 8 ophthalmologists. Esteva et al. (33) used a CNN on 129,000 of dermatological lesions to classify whether the lesion was a benign seborrheic keratosis versus a keratinocyte carcinoma or a benign nevus versus a malignant melanoma. This group found that their CNN performed about as well as a panel of 21 board-certified dermatologists. Importantly, these 2 papers demonstrate an important drawback of deep learning: it takes an enormous amount of data to train a deep learning model because of the vast number of parameters that must be estimated. The expense and difficulty of acquiring biomedical data compared with other fields are limiting factors for the application of AI in some circumstances.

Despite its nascence, deep learning applied to the domain of cardiology shows great potential. For example, in 2016, citizen-scientists participated in the Second Annual Kaggle Data Science Bowl, "Transforming How We Diagnose Heart Disease." The bowl challenged scientists to create a method to measure end-systolic and end-diastolic volumes in cardiac magnetic resonance images from more than 1,000 patients automatically. The top-performing team had no prior background in medicine. In fact, they were data scientists who worked for a financial institution. In addition, at the beginning of 2016, the first paper was published applying CNNs for electrocardiographic anomaly detection (34). The method consisted of a 2-stage learning process, first finding an appropriate feature representation per patient and then using

the first learned features for anomaly detection at later time points for the same patient.

Abdolmanafi et al. (35) used a CNN called AlexNet to classify coronary artery optical coherence tomography images in Kawasaki disease automatically. Notably, these investigators found that using the convolutional network as a feature extractor in combination with a random forest was as accurate as fine-tuning the CNN itself for prediction, but it took much less time. In an example of noncomputer vision-based neural network, Choi et al. (36) used an RNN to predict heart failure diagnosis from EHRs. Their RNN only modestly outperformed other machine learning algorithms when using 12 months of EHR data. When these investigators expanded their dataset to include another 6 months of data, their model outperformed other machine learning algorithms. Notably, as part of this work Choi et al. (36) developed an innovative method to include temporal sequencing as part of the neural network.

UNSUPERVISED LEARNING. Although we have thus largely focused on supervised machine learning, an equally important concept is unsupervised learning (**Table 1**). Whereas supervised learning focuses on prediction of outcomes and requires labeled cases, unsupervised learning focuses on uncovering underlying structure and relationships in a dataset. Unsupervised learning does not require labeled observations. Like supervised learning, unsupervised machine learning methods exist on a continuum with more traditional statistical methods such as principal components analysis, mixture modeling, and various methods of clustering. However, in recent years some new techniques that require fewer assumptions about the dataset have emerged, such as advanced algorithms for matrix or tensor factorization (37), topological data analysis (38), and deep learning (39).

One of the most promising uses of unsupervised learning methods for cardiology is subtyping or "precision phenotyping" of cardiovascular disease (40,41). We use *precision medicine* as a term describing the synthesis of multiple sources of evidence to refine monolithic disease categories into more stratified and ultimately more personal disease concepts. Precision medicine in cardiology exists in contrast to precision medicine as understood in other fields such as cancer, where a series of somatic genetic mutations can clearly define a before and after state (40,41). In cardiology, most diseases are slow, heterogeneous, multimorbid, chronic processes where pathogenesis may begin decades before any ultimate disease manifestation. This is compounded by the issue that many disease concepts in cardiology such as heart failure or coronary artery disease are somewhat broadly defined and may be arrived at by different pathophysiological mechanisms. Unsupervised learning allows to us enable precision cardiology by learning subtypes of monolithic disease concepts, and we envision ultimately it will help to treat these subtypes differently and thus lead to improved outcomes.

In this context, cardiology is ripe for the application of unsupervised learning. For example, Li et al. (38) combined EHR with genetic data from a health system biobank to study type 2 diabetes mellitus. An unsupervised learning technique called topological data analysis revealed the presence of 3 distinct subtypes of type 2 diabetes. These subtypes may reflect differing etiologies and enable subtype-based therapies. In another example, Miotto et al. (39) used a type of deep learning network with stacked denoising autoencoders (a type of neural network) to build representations of all patients within a single hospital's EHR in an entirely unsupervised fashion. Katz et al. (42) and Shah and Ho et al. (43-46) published a series of papers using various clustering techniques to identify disease subtypes of heart failure with preserved ejection fraction. More recently, Luo, Ahmad, and Shah (37) proposed the application of tensor factorization for subtyping of heart failure with preserved ejection fraction. Tensor factorization is especially attractive because tensors (multidimensional matrices) naturally lend themselves to representation of multimodal data. In the context of precision medicine, factorization of tensors should enable the identification of disease subtypes by using very high-dimensional data.

REINFORCEMENT LEARNING

Reinforcement learning algorithms learn behavior through trial and error given only input data and an outcome to optimize (**Figure 3D**). In popular culture, a breakthrough example from 2015 highlights the power of this technique. A group of researchers trained a reinforcement learning model on a variety of classic Atari 2600 video games and provided only video input and the game's final score (47). The model "learned" the optimal method to maximize the final score. More recently, a research group at Google trained a reinforcement learning model to beat a world champion at the Chinese board game Go, a task once believed too difficult for computers (48). Specifically, reinforcement learning algorithms consist of an agent at a particular time interacting with an environment. An action is selected for each time point according to some selection policy. Transitions

to the next state are then performed, and a reward is received depending on the result of the transition. The restricted learning model aims to maximize the expectation of long-term rewards from each state visited.

Application of reinforcement learning to health care and cardiology thus far has been scarce. So-called dynamic treatment regimens that tailor treatment decisions to a patient's characteristics are potential applications for reinforcement learning algorithms because of their inherent sequential decision-making structure, although statistical causal inference approaches also show promise when applied to this problem (49). Work from Shortreed et al. (50) demonstrated that reinforcement learning can work for optimization of treatment policies in chronic illnesses. Importantly, these investigators showed that reinforcement learning can overcome the problem of missing data and quantify the uncertainty of recommended policy. More recent work using restricted learning to manage weaning of mechanical ventilation in intensive care units shows great promise in minimizing rates of reintubation and regulating physiological stability (51). We envision that reinforcement learning models will eventually be commonplace and function as physician extenders in day-to-day clinical practice, either built into the EHRs or as part of devices worn by the clinician.

WHAT WILL CARDIOVASCULAR MEDICINE GAIN FROM MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE?

Cardiologists make decisions for patient care from data, and they tend to have access to richer quantitative data on patients compared with many other specialties. Despite some potential pitfalls, it is becoming evident that the best way to make decisions on the basis of data is through the application of techniques drawn from AI. Cardiologists will thus need to incorporate AI and machine learning into the clinic. Indeed, as the amount of available patient-level data continues to increase and we continue to incorporate new streams of complex biomedical data into the clinic, it is likely that AI will become essential to the practice of clinical medicine. This will probably happen sooner rather than later, as exemplified by the rapid adoption of automated algorithms for computer vision in radiology and pathology (52).

However, the incorporation of AI into cardiology is not something that clinicians should fear, but is instead a change that should be embraced. AI will drive improved patient care because physicians will be able to interpret more data in greater depth than ever before. Reinforcement learning algorithms will become companion physician aids, unobtrusively assisting physicians and streamlining clinical care. Advances in unsupervised learning will enable far greater characterization of patients' disorders and ultimately lead to better treatment selection and improved outcomes. Indeed, AI may obviate much of the tedium of modern-day clinical practice, such as interacting with EHRs and billing, which will likely soon be intelligently automated to a much greater extent. Although currently machine learning is often performed by personnel with specialized training, in the future deploying these methods will become increasingly easy and commoditized. The expert knowledge of pathophysiology and clinical presentation that physicians acquire over their training and career will remain vital. Physicians should therefore take a lead role in deciding where to apply and how to interpret these models.

ADDRESS FOR CORRESPONDENCE: Dr. Joel T. Dudley, Icahn School of Medicine at Mount Sinai University, 770 Lexington Avenue, 15th Floor, New York, New York 10065. E-mail: joel.dudley@mssm.edu. Twitter: @jdudley.

REFERENCES

1. Kuo FC, Mar BG, Lindsley RC, Lindeman NI. The relative utilities of genome-wide, gene panel, and individual gene sequencing in clinical practice. Blood 2017;130:433–9.

2. Muse ED, Barrett PM, Steinhubl SR, Topol EJ. Towards a smart medical home. Lancet 2017;389:358.

3. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015;7:283rv3.

4. Shameer K, Badgeley MA, Miotto R, Glicksberg BS, Morgan JW, Dudley JT. Translational bioinformatics in the era of real-time biomedical, health care and wellness data streams. Briefings in Bioinformatics 2017;18:105–24.

5. Konstam MA, Hill JA, Kovacs RJ, et al. The academic medical system: reinvention to survive the revolution in health care. J Am Coll Cardiol 2017;69:1305–12.

6. Steinhubl SR, Topol EJ. Moving from digitalization to digitization in cardiovascular care: why is it important, and what could it mean for patients and providers? J Am Coll Cardiol 2015;66:1489–96.

7. Boeldt DL, Wineinger NE, Waalen J, et al. How consumers and physicians view new medical technology: comparative survey. J Med Internet Res 2015;17:e215.

8. Mortazavi BJ, Downing NS, Bucholz EM, et al. Analysis of machine learning techniques for heart failure readmissions. Circ Cardiovasc Qual Outcomes 2016;9:629–40.

9. Shameer K, Johnson KW, Yahi A, et al. Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: a case-study using Mount Sinai heart failure cohort. Pac Symp Biocomput 2016;22:276–87.

10. Steyerberg EW, Eijkemans MJ, Harrell FE Jr., Habbema JD. Prognostic modelling with logistic

regression analysis: a comparison of selection and estimation methods in small data sets. Stat Med 2000;19:1059–79.

11. Janssen KJ, Moons KG, Harrell FE Jr. Prediction rules must be developed according to methodological guidelines. Ann Intern Med 2010;152:263. author reply 263–4.

12. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. Crit Care Med 2016;44:368–74.

13. Frizzell JD, Liang L, Schulte PJ, et al. Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches. JAMA Cardiol 2017;2:204–9.

14. Johnston BC, Alonso-Coello P, Friedrich JO, et al. Do clinicians understand the size of treatment effects? A randomized survey across 8 countries. CMAJ 2016;188:25–32.

15. de Caestecker M, Humphreys BD, Liu KD, et al. Bridging translation by improving preclinical study design in AKI. J Am Soc Nephrol 2015;26:2905–16.

16. Harrell FE. How To Do Bad Biomarker Research. Vanderbilt Center For Quantitative Sciences Workshop, Nashville, Tennessee, 2015.

17. Senn S. Dichotomania: an obsessive compulsive disorder that is badly affecting the quality of analysis of pharmaceutical trials. Presented at: 55th Session of the International Statistical Institute; 2005, Sydney, Australia.

18. Kolek MJ, Graves AJ, Xu M, et al. Evaluation of a prediction model for the development of atrial fibrillation in a repository of electronic medical records. JAMA Cardiol 2016;1:1007–13.

19. Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction model when there are few events. BMJ 2015;351:h3868.

20. Halim SA, Neely ML, Pieper KS, et al. Simultaneous consideration of multiple candidate protein biomarkers for long-term risk for cardiovascular events. Circ Cardiovasc Genet 2015;8:168–77.

21. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS One 2017;12:e0174944.

22. 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.

23. Salem Omar AM, Shameer K, Narula S, et al. Artificial intelligence-based assessment of left ventricular filling pressures from 2-dimensional cardiac ultrasound images. J Am Coll Cardiol Img 2017. S1936–878X.

24. Cui S, Li K, Ang L, et al. Plasma phospholipids and sphingolipids identify stent restenosis after percutaneous coronary intervention. J Am Coll Cardiol Intv 2017;10:1307–16.

25. Alexandru Niculescu-Mizil RC. Predicting good probabilities with supervised learning. In: Proceedings of the 22nd International Conference on Machine Learning. Bonn, Germany, 2005:625–32.

26. Ortiz J, Ghefter CG, Silva CE, Sabbatini RM. One-year mortality prognosis in heart failure: a neural network approach based on echocardiographic data. J Am Coll Cardiol 1995;26:1586–93.

27. Atienza F, Martinez-Alzamora N, De Velasco JA, Dreiseitl S, Ohno-Machado L. Risk stratification in heart failure using artificial neural networks. Proc AMIA Symp 2000:32–6.

28. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 1958;65:386–408.

29. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.

30. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 2017 May 6 [E-pub ahead of print].

31. Narula S, Shameer K, Salem Omar AM, Dudley JT, Sengupta PP. Reply: Deep learning with unsupervised feature in echocardiographic imaging. J Am Coll Cardiol 2017;69:2101–2.

32. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016;316:2402–10.

33. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542:115–8.

34. Kiranyaz S, Ince T, Gabbouj M. Real-Time Patient-specific ECG classification by 1-D convolutional neural networks. IEEE Trans Biomed Eng 2016;63:664–75.

35. Abdolmanafi A, Duong L, Dahdah N, Cheriet F. Deep feature learning for automatic tissue classification of coronary artery using optical coherence tomography. Biomed Opt Express 2017;8:1203–20.

36. Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. J Am Med Inform Assoc 2017;24:361–70.

37. Luo Y, Ahmad FS, Shah SJ. Tensor factorization for precision medicine in heart failure with preserved ejection fraction. J Cardiovasc Transl Res 2017;10:305–12.

38. Li L, Cheng WY, Glicksberg BS, et al. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. Sci Transl Med 2015;7:311ra174.

39. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 2016;6:26094.

40. Antman EM, Loscalzo J. Precision medicine in cardiology. Nat Rev Cardiol 2016;13:591–602.

41. Johnson KW, Shameer K, Glicksberg BS, et al. Enabling precision cardiology through multiscale biology and systems medicine. J Am Coll Cardiol Basic Trans Science 2017;2:311–27.

42. Katz DH, Deo RC, Aguilar FG, et al. Phenomapping for the identification of hypertensive patients with the myocardial substrate for heart failure with preserved ejection fraction. J Cardiovasc Transl Res 2017;10:275–84.

43. Shah SJ, Kitzman DW, Borlaug BA, et al. Phenotype-specific treatment of heart failure with preserved ejection fraction: a multiorgan roadmap. Circulation 2016;134:73–90.

44. Shah SJ, Katz DH, Deo RC. Phenotypic spectrum of heart failure with preserved ejection fraction. Heart Fail Clin 2014;10:407–18.

45. Ho JE, Enserro D, Brouwers FP, et al. Predicting heart failure with preserved and reduced ejection fraction: the International Collaboration on Heart Failure Subtypes. Circ Heart Fail 2016;9.

46. Shah SJ, Katz DH, Selvaraj S, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. Circulation 2015;131:269–79.

47. Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. Nature 2015;518:529–33.

48. Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of Go without human knowledge. Nature 2017;550:354–9.

49. Johnson KW, Glicksberg BS, Hodos RA, Shameer K, Dudley JT. Causal inference on electronic health records to assess blood pressure treatment targets: an application of the parametric g formula. Pac Symp Biocomput 2018;23:180–91.

50. Shortreed SM, Laber E, Lizotte DJ, Stroup TS, Pineau J, Murphy SA. Informing sequential clinical decision-making through reinforcement learning: an empirical study. Mach Learn 2011;84:109–36.

51. Prasad N, Cheng LF, Chivers C, Draugelis M, Engelhardt BE. A reinforcement learning approach to weaning of mechanical ventilation in intensive care units. ArXiv e-prints 2017. Available at: http://auai.org/uai2017/proceedings/papers/209.pdf. Accessed April 17, 2018.

52. Shameer K, Johnson KW, Glicksberg BS, et al. Machine learning in cardiovascular medicine: are we there yet? Heart 2018 Jan 19 [E-pub ahead of print].

KEY WORDS artificial intelligence, cardiology, machine learning, precision medicine