American
Heart
Association.

Artificial Intelligence: Practical Primer for Clinical Research in Cardiovascular Disease

Nobuyuki Kagiyama, MD, PhD; Sirish Shrestha, MS; Peter D. Farjo, MD; Partho P. Sengupta, MD, DM

Artificial intelligence (AI) has begun to permeate and reform the field of medicine and cardiovascular medicine. Impacting about 100 million patients in the United States, the burden of cardiovascular disease is felt in a diverse array of demographics.[1,2] Meanwhile, routine mediums such as multimodality images, electronic health records (EHR), and mobile health devices store troves of underutilized data for each patient. AI has the potential to improve and influence the status quo, with capacity to learn from these massive data and apply knowledge from them to distinct circumstances.[3,4] With considerable information in each heart beat, cardiovascular medicine will definitely be one of the fields that embrace AI to move toward personalized and precise care.[5]

AI has already been woven into the fabric of everyday life. From an internet search engine, email spam and malware filtering, to uncovering fraudulent credit card purchases, AI addresses an individual's needs in the realms of business, entertainment, and technology. Unfortunately, medicine, including cardiology, has not fully embraced this revolution, with only a limited number of AI-based clinical applications being available. Nevertheless, there is promise towards routine implementation; machine learning and deep learning have seen an exponential surge of cardiovascular publications in the past decade.[6,7] These methods have proven beneficial in a variety of complex areas including echocardiogram interpretation and diastolic dysfunction grade stratification.[8,9] The US Food and Drug Administration has already approved several devices that utilize AI features.[10] Imagine coming to work finding that your system has analyzed all your patients while you were sleeping: their laboratory data, imaging results, symptoms, and mobile device data to calculate their risk of cardiovascular events, death, hospitalization, whether medications should be adjusted/added/removed, or whether they should be referred for an examination. The system presents you with the reasons for its recommendations, and you are confident that they are as good as those given by the most experienced physicians. This may allow you to spend time in shared decision making with your patients, both objectively and compassionately. Although there are currently several barriers/challenges to adoption of AI in clinical practice, undoubtedly, AI will drive current healthcare practice towards a more individualized and precision-based approach over the next several years. Therefore, general understanding of AI techniques by clinicians and researchers in cardiovascular medicine is paramount. In this review article, we describe the fundamentals of AI that clinicians and researchers should understand, its definition and principles, how to interpret and apply AI in cardiovascular research, limitations, and future perspectives.

Basic Knowledge of AI

Terminologies

Currently, the term "AI" is glamorous in the medical field; however, there are confusion and misunderstanding of the terms and techniques. AI is a broad and ambiguous term that describes any computational programs that simulate and mimic human intelligence such as problem solving and learning. AI can indicate general-purpose AI (general AI), in which the system is self-sufficient and possesses cognition comparable to that of humans. Yet, such general AI has not appeared and an applied AI (specific AI), for a specialized and dedicated purpose, is the AI that is currently available.[11] "Machine learning" is 1 subfield of an applied AI, which automatically discovers patterns of data without using explicit instructions.[12] In machine learning, the machine learns from the data and performs tasks based on the learned model, whereas simple computer programs perform tasks according to the preset rules that are created based on human experience and knowledge. An applied AI and machine learning have been used interchangeably in the medical

---

research context, since machine learning accommodates most AI technologies in the medical research setting. It includes various algorithms for prediction and classification tasks that perform well on complex *big data*. "Deep learning" is a subfield of such machine learning algorithms that use deep (=multiple layered) neural networks originally inspired by the structure of the human brain.[13] The neural networks designed currently, however, work substantially differently from how the human neuron functions. In the past decade, deep learning has been increasingly used and shown to outperform other machine learning techniques in various fields, such as image recognition, voice recognition, and game playing. In a competition of image classification, the emergence of modern deep learning, the convolutional neural network (CNN) in 2012 had a big impact on the scene. Among traditional machine learning methods showing error rates of ≈26%, which were reasonably good at the time, CNN showed an outstanding error rate of 15.3%.[14] Image classification using CNN has been improving and the current error rate is ≈3%, which surpasses the abilities of the human eye.[15] Its successes are attributed to its capability to extract important features from enormous data through iterative data processing. In contrast to traditional machine learning where algorithms require some degree of arbitration from the analysts (eg, feature selection, or feature engineering: the process of selecting and creating features, or variables, which make algorithms perform better), deep learning is generally more self-directed once implemented. For example, in order to deal with chest x-ray images with traditional machine learning, analysts must first collect the measurements and parameters such as the size of the heart and presence of the congestion into a spreadsheet. This process of categorization loses information and is time consuming. Instead, deep learning can feed raw image information and extract key patterns of images by itself.

Supervised, unsupervised, and reinforcement learning is another group of terms that describe the way a machine learns from data (Figure 1). In supervised learning, algorithms learn from the data with information on the outcome, or ground truth to develop a prediction model. Typical tasks handled by supervised learning are classification and regression. Classification is a task for predicting a categorized outcome, such as 1-year mortality (yes or no) and disease diagnosis[3,16] using given parameters, while regression predicts the value (eg, predicting echocardiographic early diastolic left ventricular relaxation velocity value from ECG information). On the contrary, unsupervised learning does not require ground truth and explores the data to find hidden patterns and associations.[8] The most common tasks in unsupervised learning are clustering and dimensionality reduction. Clustering is a task to divide objects into groups with similar characteristics. Dimensionality reduction, which

can also be performed in a supervised manner, is a task to reduce the dimensionality (=the number of variables) of data with keeping principal variables that explain the data. These tasks aim to identify phenotypes by inferring the patterns from the data set without known labeled outcome. Reinforcement learning is a technique in which an algorithm is trained to learn an action that gains maximum reward in the situation. This technique is widely used for decision-making in gaming programs. AlphaGo, the first program that beat professional Go player, also used reinforcement learning to learn the best actions in the game of Go.[17]

Big Data and Machine Power to Deal With It

The concept and methodologies of AI techniques themselves are not brand-new; however, the sudden prosperity of AI in the past decade occurred with the emergence of big data and evolution of computing power as well as the development of deep learning techniques. Previously, the lack of data that are big enough to train AI was one of the bottlenecks of AI development. However, this limitation disappeared as big data became available because of the popularization of the internet. The digital information stored by smart devices with internet connection that can transfer data over a network without requiring human interaction further increased the influx of usable data. Analyzing such big data using AI requires a huge amount of computational power. Graphic Processing Units were originally invented to perform specialized tasks in gaming graphics, but its fast and parallel computing power fitted well in deep learning tasks.[13] In recent years, Graphic Processing Units have become affordable while the computing power continued to grow exponentially. Currently, researchers can also appreciate cloud-based Infrastructure as a Service, or more recently called Machine Learning as a Service such as Amazon Web Services, Microsoft Azure ML, and Google Cloud ML. They provide power of Graphic Processing Units with various AI applications and limitless data storage on the cloud platform. Furthermore, some Machine Learning as a Service provide automated machine learning systems, such as Google Cloud Auto ML and BigML, where various machine learning algorithms that require none-to-minimal coding are available. These systems can be used with simple graphic user interfaces and may be a better choice for researchers who are novices at machine learning. Thus, resources for AI research have become available for general clinicians and researchers.

Difference Between AI and Traditional Statistics

Statistics has been the standard method for medical research for the purpose of showing the benefit of new therapies,
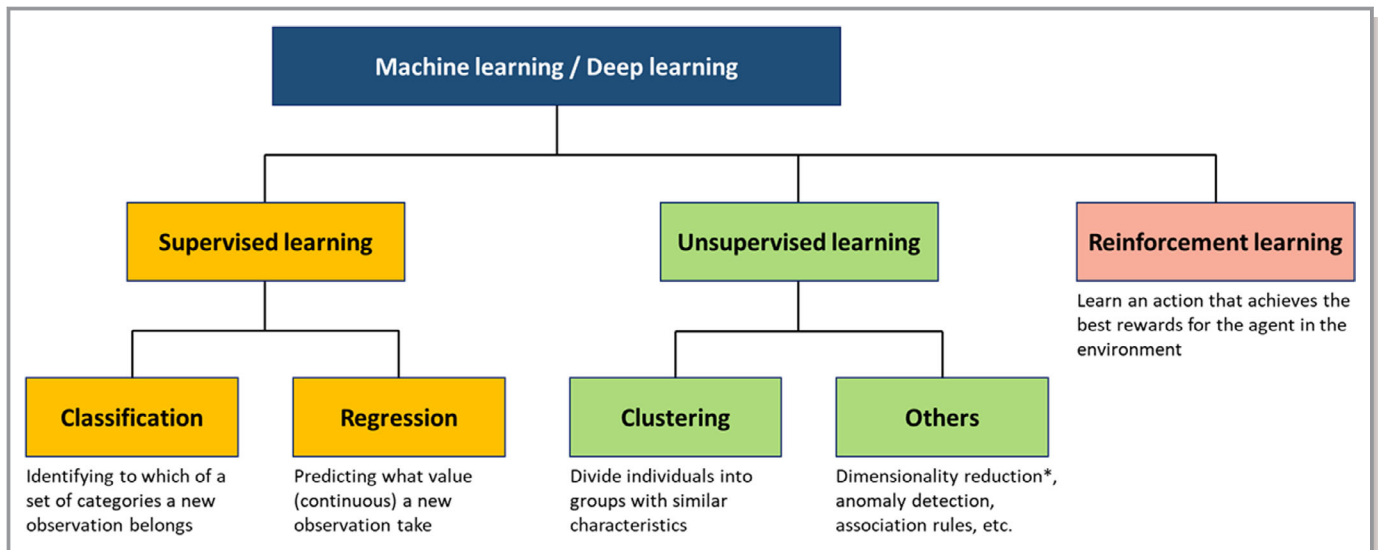
Figure 1. Supervised, unsupervised, and reinforcement learning. Machine learning tasks are categorized into supervised, unsupervised, or reinforcement learning. Supervised learning is used for prediction (classification or regression), whereas unsupervised learning aims to reveal hidden patterns in data. Reinforcement learning is another way of learning where an algorithm learns the best action based on its consequences, and is well suited for game theory and control theory. However, it has not had a significant role in clinical research because it requires simulating many "wrong actions" to learn. *Dimensionality reduction can also be performed in a supervised manner.

predicting prognoses, identifying risk factors, and revealing disease mechanisms. Interestingly, there are significant overlaps in techniques and methodologies in the domain of traditional statistics and AI. For example, logistic and linear regression models, which most medical researchers are familiar with, are also techniques in machine learning. Perhaps the fundamental difference is in their philosophies; statistics is a science that estimates and explains data, whereas AI, or machine learning, aims to achieve practical prediction from data at hand. For example, in linear regression models, the most important parameters of interest in statistics are coefficients (weights) of each term and the goodness-of-fit, both of which explain the data. On the other hand, AI focuses on prediction of unknown data. Accordingly, the primary concern of AI research is model performance in a test set, which is not used in the process of model training, and it is presented with different terms from statistics, such as recall (=sensitivity), F-measure (a harmony of sensitivity and positive predicted value), and confusion matrix (a type of cross-tabulation table). In fact, the calculation process is of less interest in AI, and some complex AI models do not even provide coefficient or other metrics for interpretability. Therefore, usually AI requires fewer assumptions of data and often uses very complex nonparametric models, which requires much more data than simple parametric models that are frequently used in traditional statistics. While traditional statistics perform well for a certain hypothesis with suitable-sized data set, AI generally outperforms statistical methods for prediction in large and complex data.

Another important consideration, especially important for medicine, is that AI techniques, unlike some sophisticated statistical analysis, have been suggested not to provide causal inferences. Actually, there is renewed enthusiasm in using machine learning for this very purpose. This evolving paradigm, however, should be verified cautiously with the existing domain knowledge about pathophysiology and disease mechanisms to support the results of AI analysis. Capability for working on various data structures is also an important strength of AI compared with traditional research, as discussed below. Figure 2 summarizes differences between traditional medical research with statistics and research using AI.

Representative Algorithms of AI

Machine learning and deep learning consist of a multitude of algorithms. Table 1 summarizes brief descriptions of basic machine learning algorithms used in different tasks. Currently, ensemble learning and deep learning can be described as the mainstay of algorithms of AI. Ensemble learning is a machine learning method that combines multiple "weak" learners (algorithms) such as decision tree and logistic regression (Table 1) to obtain a good prediction. Boosting, bagging, and stacking are the 3 main methods of ensemble learning.[18] In boosting, multiple weak learners are combined in series and trained subsequently with considering errors of preceding algorithms to reduced bias. Bagging is a method in which multiple weak learners are trained in parallel, and the results

of each algorithm are combined to make a final output with small variance. Stacking is the other method in which the results of weak learners are used as input of another machine learning algorithm (meta learner). These ensemble learning methods work very well by combining various types of simple algorithms and generally outperform any single machine learning algorithm.

Deep learning outperforms other traditional machine learning methods in analyzing complex data such as images, texts, and other unstructured data. In general, deep learning consists of an input layer, hidden layers, and an output layer, where input and output layers indicate original data and output of the algorithm, respectively. Through multiple hidden layers, raw input is gradually converted into more abstract and essential features that represent the original data (Figure 3). In image recognition, the input layer indicates raw pixels of the image, then first layers identify simple features of the image such as edges and lines. Succeeding layers identify somewhat more complex features such as ears, eyes, and tails. Finally, last layers recognize features of cats and dogs. As such, deep learning extracts key features from raw unstructured data and returns outputs as classification or regression.

Examples of AI on Cardiovascular Data

Structured Data

Examples of the AI studies in various data sources are summarized in Table 2. Currently, most medical research is done using structured data, which are labeled properly and organized into formatted fields in tabular form.
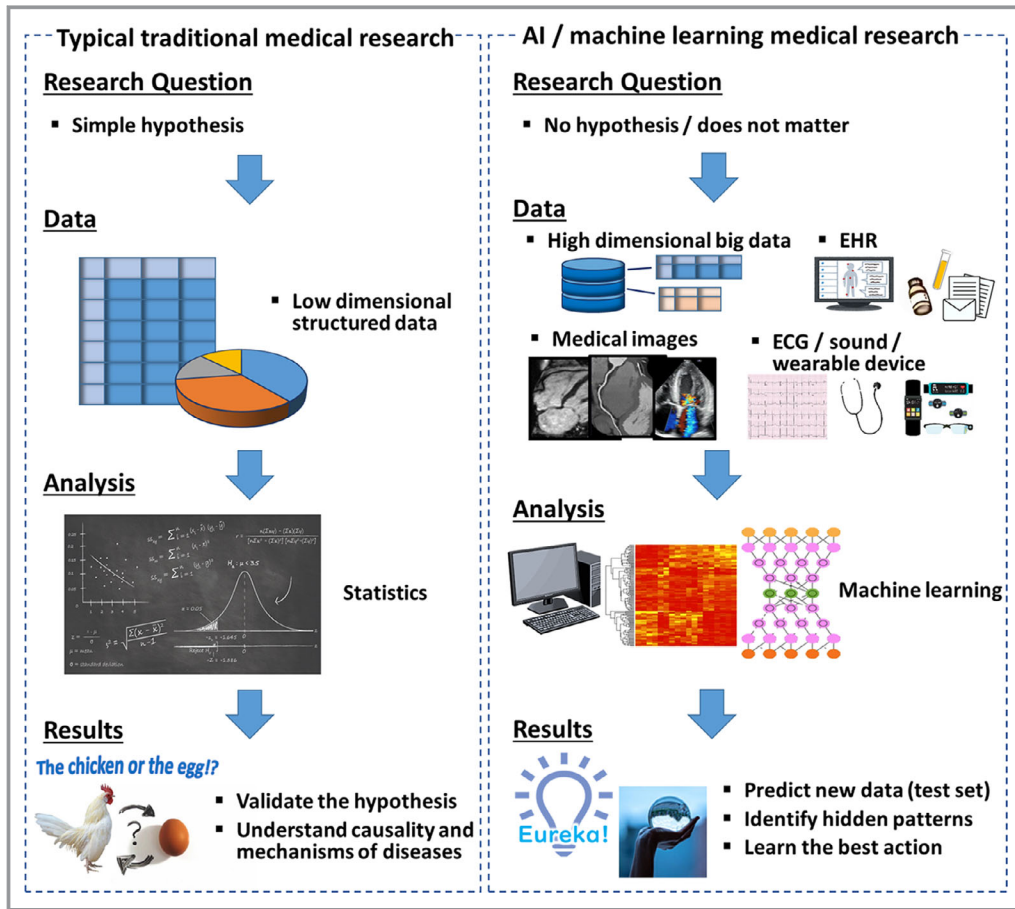

Figure 2. Pipelines of medical research using traditional statistics and AI. Traditional medical research formulates a hypothesis first, and tests it using statistical analysis. Medical research using AI can be hypothesis-free and data-driven. Compared with traditional statistics, AI can deal with various types of data, including unstructured data such as images, signals, and EHR. In contrast to traditional medical research that focuses on validation of hypotheses and understanding causality and mechanisms, the main goal of research using AI is to predict new data and identify a hidden pattern in the data. AI indicates artificial intelligence; EHR, electronic health record.

Table 1. Representative Machine Learning Algorithms

Algorithm	Description	Use
Logistic regression	An algorithm that estimates probability of dichotomized outcome from multiple covariates using logistic function.	Classification
Decision tree	A flow chart–like algorithm that divides data into branches by considering information gain. The final branches represent output of the algorithm (class or value).	Classification/regression
(simple) Neural network	An algorithm inspired by human brain architecture. Layers consisting of nodes are connected to one another with edges weighted as per training results.	Classification/regression
K nearest neighbor	A simple algorithm that classifies observations by comparing k examples that exist in the nearest locations (=examples with the most similar features).	Classification/regression
Support vector machine	Support vector machine draws a boundary line that maximizes margins from each class. New observations are classified using this line.	Classification/regression
K means	A clustering method that makes k clusters in which each observation belongs to the cluster that has its mean in the nearest locations from the observation.	Clustering
Hierarchical clustering	A type of cluster analysis that builds a dendrogram with a hierarchy of clusters. Pairs of clusters are merged to form clusters as they move up the hierarchy (agglomerative approach).	Clustering
Principal component analysis	An algorithm that converts high dimensional data into lower dimensional data with keeping important information as much as possible by orthogonal transformation	Dimensionality reduction

Motwani and colleagues[16] investigated the predictive ability of machine learning algorithms on structured data to predict 5-year mortality in a large population of 10 030 patients with coronary artery disease enrolled in the CONFIRM (Coronary CT Angiography Evaluation for Clinical Outcome) study. They used ensemble learning to analyze 69 parameters in the data set. Using 10-fold cross-validation (explained later), they trained the machine learning algorithm to predict mortality (=supervised learning) in each training set and tested the predictive ability using each validation set. Features were selected in each fold using information gain. They found that the area under the receiver-operating-characteristic curves for prediction of 5-year mortality was significantly better with the machine learning algorithm (0.79) than other traditional risk scores (0.61–0.64).

Structured data can also be applied in the unsupervised machine learning algorithm effortlessly and efficiently. Recently, our group[19] applied a novel data analytics technique called topological data analysis to build a patient–patient similarity network that utilized the underpinning of mathematics and an underlying unsupervised machine learning. Topological data analysis is a framework for machine learning; it borrows and amalgamates various machine learning algorithms to understand the fundamental properties and the shape of complex data. The study applied the technique to understand the phenotypic representation of the pattern of left ventricular responses in the progression of aortic stenosis. Topological data analysis, along with dimensionality reduction, formed a loop segregating mild and severe disease in opposite ends, while linking them through moderate disease over the routes of preserved and reduced left ventricular ejection fraction

(Figure 4). Interestingly, upon supplementing the data with the follow-up succeeding the aortic valve replacement therapy, the patients were accurately captured to have traversed from severe to mild or moderate aortic stenosis. A similar model was then applied on a murine model as a reverse-translational study that showed a similar distribution separating mice with high peak aortic velocities in 1 end to low velocities in the other, connecting via moderately severe peak aortic velocities in the top and the bottom of the loop.

However, machine learning is not always superior to traditional statistics. Frizzell et al[20] studied 56 477 patients who were admitted to hospitals and were older than 65 years of age. The data set included about 100 candidate variables, and the performance for prediction of 30-day rehospitalization rate was compared among several machine learning methods including ensemble models. Despite the large data set, a logistic regression model achieved better performance than other complex machine learning models. Importantly, even the performance of this logistic model was modest (C statistics 0.624), suggesting that there are many unmeasured/unknown important variables that contribute to the outcome. As such, complex machine learning models can fail to outperform simple models in the absence of several contributing variables.

Unstructured Data

Unstructured data are the data stored without a well-organized structure that is applied on traditional statistical and tabular databases. In the medical field, textual information in EHR, medical images, and audio and visual clips are
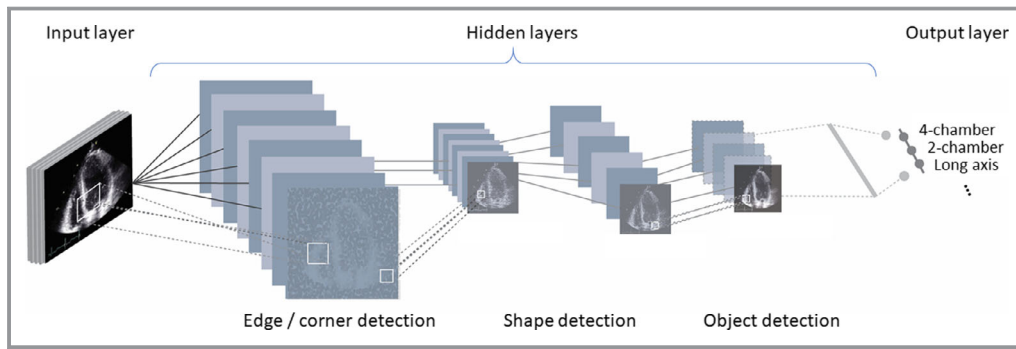
Figure 3. Structure of deep learning. Deep learning consists of input layers, hidden layers, and output layers. Through multiple hidden layers, raw input is gradually converted into more abstract and essential features that represent the original data. In image recognition, the input layer indicates raw pixels of the image, then first layers identify simple features of the image such as edges and lines. Succeeding layers identify somewhat more complex features such as ears, eyes, and tails. Finally, last layers recognize features of cats and dogs. As such, deep learning extracts key features from raw unstructured data and returns outputs as classification or regression.

examples of data that are considered unstructured. The emergence of machine learning has opened avenues to effectively analyze such data, which are thought to contain 80% to 90% of all potentially usable information[21] and to be a huge resource for medical research.

Medical Images

Traditionally, it was difficult for computers to automatically deal with medical images, as they have enormous varieties of disease patterns. Physicians have had to exclusively read, interpret, and analyze a variety of medical images. These processes are time-consuming and can be one of the bottlenecks of clinical practice. Deep learning, especially CNN and other derivative neural networks, are becoming a game changer in the process of medical image analysis, with its capability to learn features from pixels and classify and segment objects in images. [22,23]

Zhang et al[24] used a CNN to develop a pipeline for automation of the following echocardiographic tasks: (1) view classification, (2) image segmentation, (3) measurements of cardiac structure and function, and (4) discrimination of diseases. In the study, this deep learning algorithm was trained in a supervised manner to classify images into 1 of 23 types of views (normal parasternal long axis, remote parasternal long axis, etc.) using ≈70 000 preprocessed still images generated from 277 echocardiograms. In a 5-fold cross-validation, the algorithm could excellently distinguish broad subclasses from one another with an overall classification accuracy of 84%. For image segmentation, another algorithm was trained to segment cardiac chambers in 5 views (apical 2-, 3-, and 4-chambers and parasternal long- and short-axis views), respectively. After supervised learning using 124 to 214 images per single view, the algorithms were able to segment areas of individual cardiac chambers with excellent

overlap to human-annotated areas of chambers. Using these auto-annotated chambers, they calculated chamber volumes, left ventricular ejection fraction, and left ventricular mass, which agreed with manually measured values (median absolute deviations were 15−17%). Finally, they trained algorithms for disease classification of hypertrophic cardiomyopathy (n=260) from matched normal controls (n=2064), cardiac amyloidosis (n=81) from controls (n=771), and pulmonary artery hypertension (n=104) from controls (n=2180). The trained algorithms showed excellent discrimination of the diseases (area under the receiver-operating-characteristic curves 0.93, 0.87, and 0.85, respectively). In the article, the authors implied the possibility of auto-analysis and direct diagnosis from echocardiographic images using deep learning.

Signal Data Including ECG and Phonocardiograms

Signal modalities include ECG, sound, phonocardiograms, oscillometric devices, and some wearable devices. An ECG signal is one of the best-studied signals in cardiovascular medicine. Deep learning techniques have also pervaded this field. Recently, Hannun and colleagues[25] developed a deep learning algorithm that classifies single-lead ECG into 12 classes of rhythms, such as sinus rhythm, junctional rhythm, atrioventricular block, and atrial fibrillation. They used 30-second-long raw ECG signals from single lead of 91 232 ambulatory ECGs, which were labeled by certified ECG technicians (supervised learning) to train the algorithm. After training, the algorithm identified arrhythmias in 328 test sets with better accuracy (F-measure 0.84) than cardiologists achieved (averaged F-measure 0.78).

Another signal of interest in cardiovascular medicine is the phonocardiogram (ie, the heart sound information). PhysioNet Resources, a research resource established for the purpose of studies on biomedical and physiologic signals,

Table 2. Examples of ML in Cardiovascular Research

Data Structure	Year	First Author	Journal/Conference	Task	Model	Summary
Structured data						
	2016	Motowani	Eur Heart J	Classification: Prognostic prediction	Ensemble	Using 69 clinical and CT parameters of 10 030 CAD patients, a ML model predicted mortality better than traditional statistics
	2018	Kakadiaris	JAHA	Classification: Prognostic prediction	SVM	Using 9 parameters that consist of ACC/AHA risk calculator, a ML model showed better prediction than original ACC/AHA risk score.
	2016	Narula	JACC	Classification: Diagnosis of HCM	Ensemble (SVM, RF, and ANN)	Using clinical and echocardiographic parameters, ML algorithms discriminated HCM from ATH with 87% sensitivity and 82% specificity.
	2019	Lancaster	JACC CV Imaging	Clustering	Hierarchical clustering	Using echocardiographic parameters that guidelines recommend for assessment of LVDD, hierarchical clustering identified clusters that discriminate patient prognosis better than guidelines-based classification
	2019	Casaclang-Verzosa	JACC CV Imaging	Clustering with dimensionality reduction	Topological data analysis	Topological data analysis was able to visualize patient-patient similarity network that is created from 4 parameters. Relative location of patients in the network were associated with disease phenotypes and prognosis.
Unstructured data						
Echocardiographic images	2018	Zhang	Circulation	Classification: Automatic interpretation of echocardiography	CNN	Using 14 035 echocardiograms, CNN enabled automatic classification of views, identification of chambers, measurements of cardiac volumes, and discrimination of diseases from healthy controls (see text for details)
MRI images	2019	Zhang	Radiology	Classification: Prediction of MI from non-enhanced MRI	LSTM+Optical flow	In 212 patients and 87 controls, algorithms were able to detect chronic MI (validated by LGE) with 90% sensitivity and 99% specificity using nonenhanced cine MRI.
CT images	2016	Shandmi	Med Image Anal	Classification: Coronary artery calcium in a voxel	CNN	Using 3D CTA of 250 patients, after localization of volume of interest using 3 CNNs, 2 CNNs were used to classify voxels to calcium or noncalcium. Agatston score calculated based on the voxel classification showed excellent agreement with reference standard (accuracy 83%).
ECG signals	2019	Hannun	Nat Med	Classification: Arrhythmia detection	DNN	Using 91 232 single-lead ECG, trained algorithm showed better prediction of 12 types of heart rhythm than cardiologist (F-measure 0.84 vs 0.78).
Heart sound signals	2016	Potes	2016 CinC	Classification: Normal and abnormal heart sound	AdaBoost+CNN	Combination of AdaBoost and CNN showed 94.2% sensitivity and 77.8% specificity for identifying abnormal heart sound in PhysioNet/CinC data set.
EHR	2019	Mallya	arXiv	Classification: Prognostic prediction	LSTM	Using >23 000 patients time-series data, LSTM algorithm successfully predicted the onset of heart failure 15 mo in advance (AUC 0.91)
EHR: medical letters (text)	2019	Diller	Eur Heart J	Classification: Diagnosis, symptoms and prognosis	CNN+LSTM	Using natural language processing, diagnosis (accuracy 91%) and symptoms (90.6%) were extracted from medical letters. Also, prognostic prediction using the same data was useful (HR 34.0)

ACC indicates American College of Cardiology; AHA, American Heart Association; ANN, artificial neural network; ATH, athlete; AUC, area under the receiver-operating-characteristic curves; CAD, coronary artery disease; CNN, convolutional neural network; CT, computed tomography; CTA, computed tomography angiography; DNN, deep neural network; HCM, hypertrophic cardiomyopathy; HR, hazard ratio; LGE, late gadolinium enhancement; LSTM, long short time memory; LVDD, left ventricular diastolic dysfunction; MI, myocardial infarction; ML, machine learning; MRI, magnetic resonance imaging; RF, random forest; SVM, support vector machine.
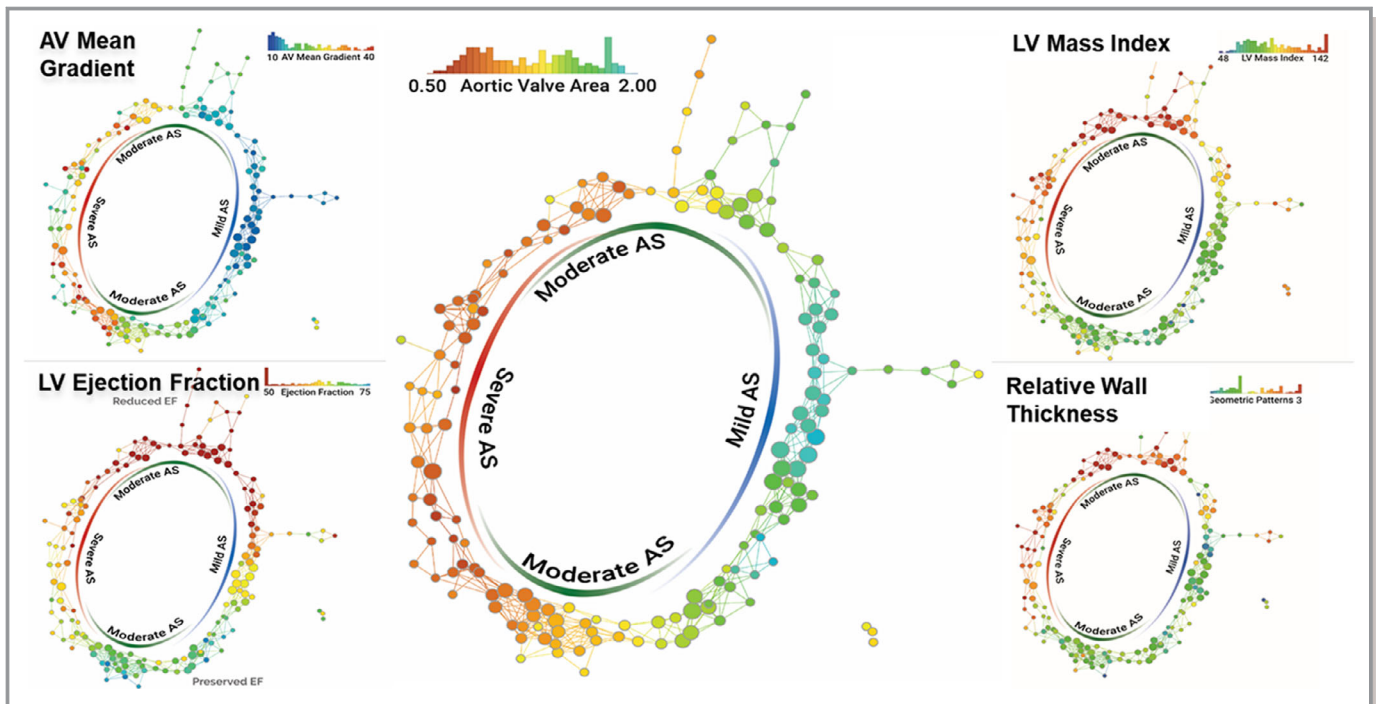
Figure 4. Topological data analysis in patients with AS. Topological data analysis enables integration of multidimensional complex data and visualization of hidden patterns in the data. Each node represents 1 or more patients with similar echocardiographic parameters of AS, and nodes including similar patients are connected by edges. Each panel is colored by 1 parameter written on the top right and color of the nodes represents the mean value of the parameter in the nodes. Although the network was created only from the parameters of aortic stenosis, preserved and reduced LV function (systolic and diastolic) were segregated in different regions. AS indicates aortic stenosis; AV, aortic valve; LV, left ventricle. Reprinted from Casaclang-Verzosa et al[19] with permission. Copyright ©2019, Elsevier.

holds yearly competitions in cooperation with Computing in Cardiology. In 2016, their challenge was to distinguish normal heart sound recorded from healthy subjects and abnormal ones from patients with heart disease.[26] They provided 4430 recordings including 233 512 heartbeats taken from 1072 subjects, among which 3153 recordings were annotated with labels. The training set included 18.1% and 8.8% of abnormal and unclear (poor quality recording) data, respectively. In total, 348 open-source entries by 48 teams were submitted to the challenge and the top score team reached 94.2% sensitivity and 77.8% specificity.[27] Interestingly, the top 5 teams all used different kinds of machine learning algorithms. This type of competition is also held by other societies such as Kaggle and MICCAI (The Medical Image Computing and Computer Assisted Intervention Society), where researchers contend for high performance of prediction using provided data.

EHR and Other Unstructured Text Data

EHR, the largest resource where most clinical information is stored, is generally not well structured, which has been a major burden for clinicians and researchers who have to read unstructured texts and manually extract information. Mallya et al[28] from Amazon conducted a nested case–control study using 21 405 patients with heart failure and 194 989 controls in a cohort of >600 000 patients' data. They extracted 1840 parameters per single patient over a 12-month time period and predicted the onset of heart failure 15 months in advance by analyzing the data using long short-term memory, a deep learning algorithm that considers time-series, with an area under the receiver-operating-characteristic curve of 0.91.

Natural language processing is a subfield of AI that is concerned with understanding and analysis of human (natural) languages by computer, and is one of the best tools to extract information from raw and unstructured text data stored in EHR. Diller et al[29] developed deep learning algorithms to automatically yield diagnosis and prognosis of 10 019 patients with adult congenital heart disease. The data during an 18-year period including 63 326 medical letters written in natural language were separated into training (80%) and test set (20%). After training and validation, deep learning algorithms (combination of CNN and long short-term memory) automatically extracted a diagnosis from the test set with an accuracy of 91.1%, and New York Heart Association class with an accuracy of 90.6%. Furthermore, an algorithm trained to predict all-cause mortality showed significant value after it was adjusted by ECG, laboratory, and exercise data.
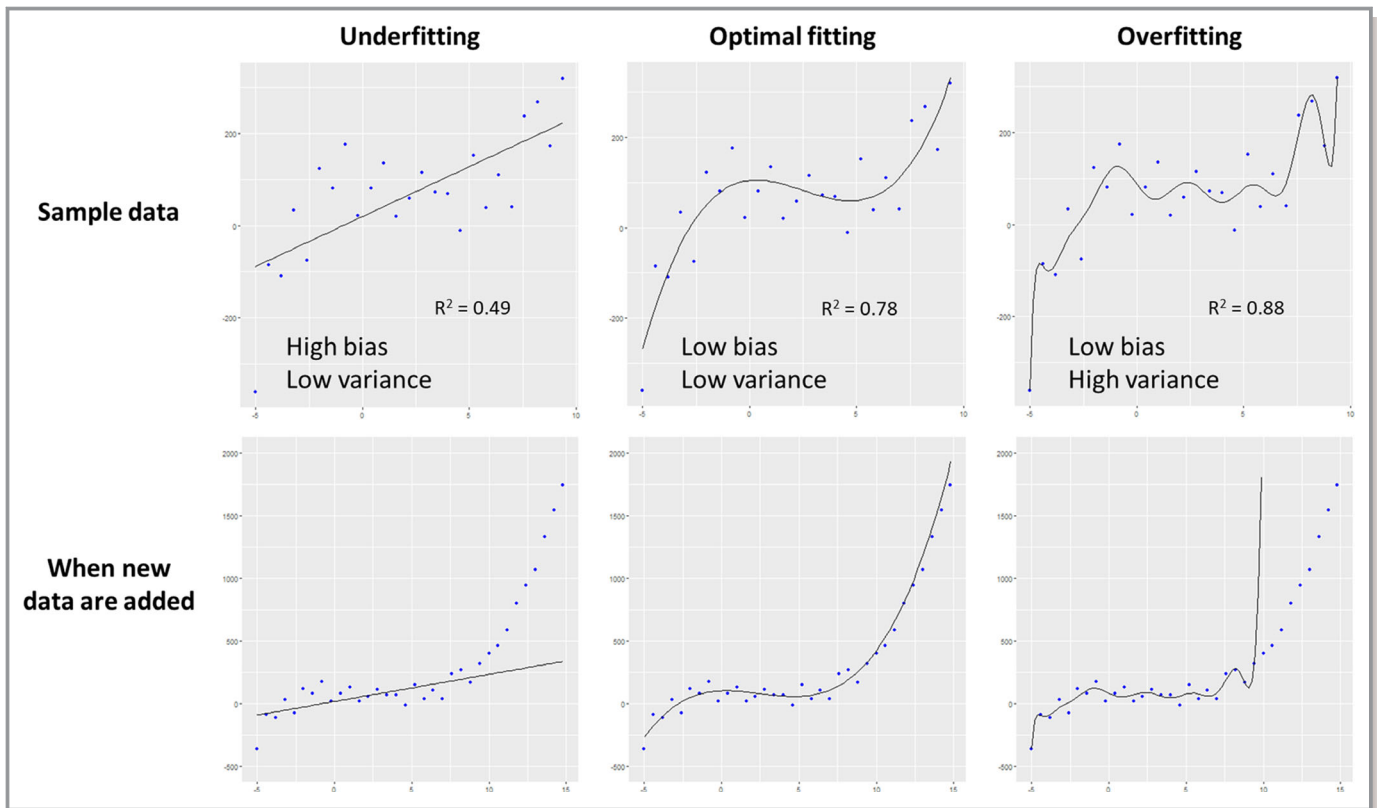
Figure 5. Underfitting, optimal fitting, and overfitting. The upper row shows regression models created in sample data (=training data), and new data from the same population were added in the bottom row. A simple linear model on the left panel was underfitted to the data, with low variance (ie, fluctuations in predicted value) but high bias (ie, difference between predicted and true value). In contrast, a complex model on the right panel was overfitted with low bias but high variance, because it also modeled random noise in the sample data. As a model becomes more complex, goodness-of-fit increases and bias decreases. However, overfitted models do not capture real association in data and cannot work well for new data.

Cautions and Limitations

As described so far, AI has tremendous possibilities in cardiovascular medicine. Yet, these techniques are not a panacea and there are several situations where AI does not work well, or even causes misleading results. First, AI can easily overfit the data set because it uses complex models with several parameters, although there are techniques to avoid overfitting in many algorithms as discussed before. An overfitted model shows very high performance in the training data set but fails to generalize in the new data set because the model also captures noise that interferes with identifying a true pattern in the data (Figure 5). Testing an established model in a test data that is completely new to the model is, therefore, mandatory for AI research and there are several techniques to overcome overfitting (Figure 6). Cross-validation is one of the preferred methods to reduce the variance in prediction error and maximize the use of data compared with a simple holdout method.[30] In a typical cross-validation (k-fold cross-validation), the data set is partitioned into several (k) bins of the data set where 1 bin is used for evaluation while the remaining bins are used for training the model. The iterative learning experiment is run k times.

As mentioned above, causal inference is one of the limitations of the current AI approach. In other words, most current AI approaches do not consider confounders. Results should be interpreted carefully in a sense of medical knowledge, when they are applied to clinical practice, especially to interventions beyond simple prediction.

Quality of data is another key important aspect of AI training. Incorrect data selection and inaccurate measurements may cause incorrect results and predictions. In 2014, Amazon developed an automated algorithm that reviewed job applications and scored candidates. The algorithm was trained using data from the previous decade, where the majority of hired personnel were male. Then, the algorithm started penalizing applications including the word "women" and ended up being scrapped later.[31] Too noisy data or data without important variables will not work either.

With these cautions and limitations, standardization of conducting and reporting AI research in medicine is mandatory. Since inaccurate analysis and insufficient reporting
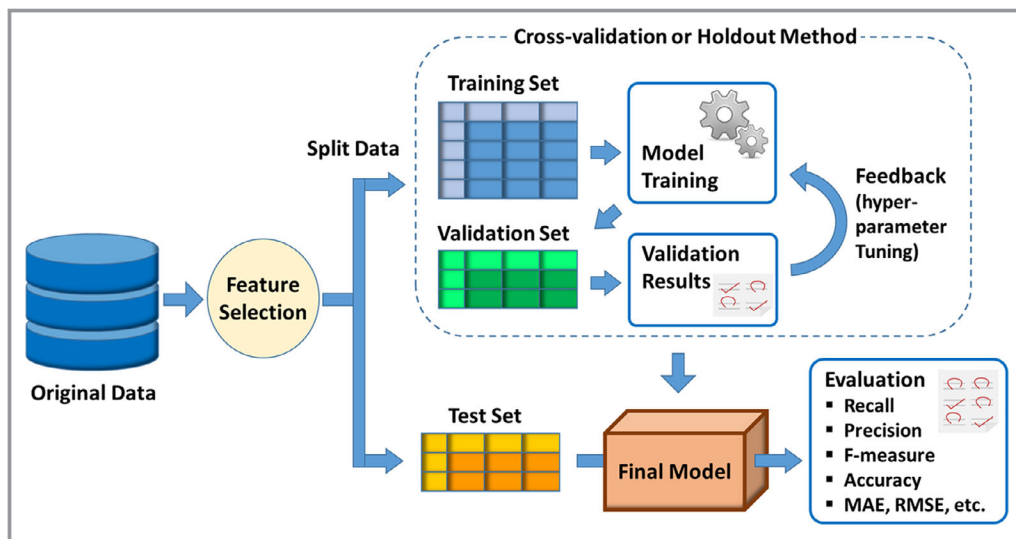
Figure 6. Development and evaluation of machine learning model. Since machine learning aims to predict new data in supervised learning, the test set is always preserved during when the machine learning model is built in order to guarantee generalizability. Ordinarily, the remaining data are further split into the training set, which is used to build models (calculate weights), and the validation set, which is used to validate the generated models and to tune hyperparameters. This training-validation process is performed using a cross-validation or holdout method. Finally, performance of the created model is evaluated using a test set that is not used in the model-building process. MAE indicates mean absolute error; RMSE, root mean square error.

contribute to the impediment in reliable assessment and causes misleading interpretation, guidelines and recommendation statements should be required for reporting consistent and reproducible results.

Challenges to Implementation

The number of research and clinical applications using AI will further increase paralleled by continuous evolution of computing power and prevailing AI platforms. Clinicians and researchers will be more likely to be involved. Thus, learning terminologies and understanding their possibilities and limitations will be more important. Cardiovascular disease is one of the fields that AI can effectively contribute to, because of its complex and multifactorial nature. Current barriers to adoption of AI involve several issues regarding infrastructures rather than AI techniques. First, with privacy issues, open data availability is limited compared with other fields. Scientific organizations and companies will need to establish data infrastructure with sufficient privacy policy. In addition, data are usually stored in multiple servers and sometimes in an analogue paper format. Even if AI establishes excellent prognostic models, it may be worthless if the hundreds of parameters for prediction are scattered in several systems and require manual input. Development of seamless data structures will be necessary. Regularization for legal and ethical issues is also important. Since AI devices can change by learning from real-world data even after deployment, the

traditional paradigm of medical device regularization is not sufficient. In addition, in cases where AI devices lead to adverse clinical outcomes, current laws may not clarify the responsibility.

However, technologies and systems are continuously improving. The US Food and Drug Administration has already issued a statement on a new, tailored review framework for AI devices, which includes modifications of devices after deployment.[32] Also, leading researchers and business leaders have already signed the Asilomar AI statement of 23 cautionary principles. The French radiology community published a white paper that better defines the clinical provider's role in AI research and its ethical implementation in their field.[33] Furthermore, multiple guidelines are now being established for the standardization of medical AI research. Survey data have shown that there is a willingness to embrace these changes, especially in younger physicians. A recent study showed that almost 95% of radiology residents would attend AI information courses if offered and 70% stated they would like advanced training in the field.[34] Hopefully, in the coming decades, a number of well-designed studies on synthesized big data will usher in a new paradigm in medicine.

Future Direction

AI contributes to the development of innovative areas including computational modeling, generation of synthetic

data and patients, and mobile health technologies. Computational modeling in medicine uses computers to simulate and study the behavior of the human body. It enables simulation of a personalized heart by integrating multiple diagnostic data obtained from clinical modalities and provides a platform for virtual evaluation and optimization of a therapy.[35] Although computer modeling is grounded on theories rather than data-driven patterns, and is usually deterministic, the concept that predicting unknown results using data at hand is common to both computer modeling and machine learning. Recent studies have reported the usefulness of implementing computational modeling using machine learning techniques such as fluid dynamic simulations[36] and adverse drug reaction.[37] Synthetic patients and data are artificially manufactured, allowing the ability to track a disease course. These are realistic, but not real, data created by analyzing existing data using machine learning techniques.[38,39] Since there is no concern regarding privacy and costs for using synthetic data, it will be a powerful tool for clinical studies that require a large number of patients and also can be an effective alternative to prepare training data for machine learning algorithms. These areas will be further enriched by AI in the near future and will contribute to realization of personalized precision medicine.

Mobile health, telemedicine, and other smart devices with internet connection are becoming another choice for collecting enormous amounts of individual-level information.[40,41] Advancement of technologies has enabled ubiquitous computers including smartphones, wearable devices, and miniaturized healthcare devices such as handheld echocardiography. These devices allow gathering of an individual's healthcare information at small clinics and even at a patient's home. The data from these devices are going to be huge, and by integrating such enormous data using AI, more detailed phenotyping of disease and more personalized medicine will be realistic.

Conclusions

AI has emerged as a promising tool in cardiovascular medicine. With the popularization of big data and machine power, the fundamentals of healthcare practice and research are bound to change. Traditional statistics remain highly effective in a simple data set and in assessing causal relationship; however, many areas in clinical practice and research will be led by powerful prediction and exploration of big data using AI. Particularly, the capability of AI to analyze unstructured data expands the field of cardiovascular research. In addition, AI will further increase its contribution to mobile health, computational modeling, and synthetic data generation, with new regularizations for its legal and ethical issues. In this paradigm shift, deep understanding of

physiology and disease mechanisms remains paramount to interpret the results of AI. Meanwhile, AI literacy will become essential to understand the latest medical knowledge and to use novel medical devices.

Disclosures

Sengupta is a consultant to Heart Sciences, Ultromics, and Kencor Health. The remaining authors have no disclosures to report.

References

1. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, Das SR, de Ferranti S, Despres JP, Fullerton HJ, Howard VJ, Huffman MD, Isasi CR, Jimenez MC, Judd SE, Kissela BM, Lichtman JH, Lisabeth LD, Liu S, Mackey RH, Magid DJ, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Nichol G, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Rosamond W, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Woo D, Yeh RW, Turner MB. Executive summary: heart disease and stroke statistics—2016 update: a report from the American Heart Association. *Circulation*. 2016;133:447–454.

2. Pilkerton CS, Singh SS, Bias TK, Frisbee SJ. Changes in cardiovascular health in the United States, 2003–2011. *J Am Heart Assoc*. 2015;4:e001650. DOI: 10.1161/JAHA.114.001650.

3. 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;68:2287–2295.

4. Mayr A, Binder H, Gefeller O, Schmid M. The evolution of boosting algorithms. From machine learning to statistical modelling. *Methods Inf Med*. 2014;53:419–427.

5. Johnson KW, Shameer K, Glicksberg BS, Readhead B, Sengupta PP, Bjorkegren JLM, Kovacic JC, Dudley JT. Enabling precision cardiology through multiscale biology and systems medicine. *JACC Basic Transl Sci*. 2017;2:311–327.

6. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: are we there yet? *Heart*. 2018;104:1156–1164.

7. Sengupta PP, Shrestha S. Machine learning for data-driven discovery: the rise and relevance. *JACC Cardiovasc Imaging*. 2019;12:690–692.

8. 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*. 2019;12:1149–1161.

9. Gandhi S, Mosleh W, Shen J, Chow CM. Automation, machine learning, and artificial intelligence in echocardiography: a brave new world. *Echocardiography*. 2018;35:1402–1418.

10. U.S. Food and Drug Administration. FDA news release: FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. 2018. Available at: https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye. Accessed May 24, 2019.

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

12. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380:1347–1358.

13. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436.

14. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. 2012;1097–1105. Available at: https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

15. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. Thirty-First AAAI Conference on Artificial Intelligence. 2017. Available at: www.cs.cmu.edu/~jeanoh/16-785/papers/szegedy-aaai2017-inception-v4.pdf.

16. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, Andreini D, Budoff MJ, Cademartiri F, Callister TQ, Chang HJ, Chinnaiyan K, Chow BJ, Cury RC, Delago A, Gomez M, Gransar H, Hadamitzky M, Hausleiter J, Hindoyan N, Feuchtner G, Kaufmann PA, Kim YJ, Leipsic J, Lin FY, Maffei E, Marques H, Pontone G, Raff G, Rubinshtein R, Shaw LJ, Stehli J, Villines TC, Dunning A, Min JK, Slomka PJ. Machine learning for prediction of all-cause

mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur Heart J*. 2017;38:500–507.

17. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, Dieleman S, Grewe D, Nham J, Kalchbrenner N, Sutskever I, Lillicrap T, Leach M, Kavukcuoglu K, Graepel T, Hassabis D. Mastering the game of go with deep neural networks and tree search. *Nature*. 2016;529:484.

18. Wang G, Hao J, Ma J, Jiang H. A comparative assessment of ensemble learning for credit scoring. *Expert Syst Appl*. 2011;38:223–230.

19. Casaclang-Verzosa G, Shrestha S, Khalil MJ, Cho JS, Tokodi M, Balla S, Alkhouli M, Badhwar V, Narula J, Miller JD, Sengupta PP. Network tomography for understanding phenotypic presentations in aortic stenosis. *JACC Cardiovasc Imaging*. 2019;12:236–248.

20. Frizzell JD, Liang L, Schulte PJ, Yancy CW, Heidenreich PA, Hernandez AF, Bhatt DL, Fonarow GC, Laskey WK. Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches. *JAMA Cardiol*. 2017;2:204–209.

21. Gandomi A, Haider M. Beyond the hype: big data concepts, methods, and analytics. *Int J Inf Manage*. 2015;35:137–144.

22. Zhang N, Yang G, Gao Z, Xu C, Zhang Y, Shi R, Keegan J, Xu L, Zhang H, Fan Z, Firmin D. Deep learning for diagnosis of chronic myocardial infarction on nonenhanced cardiac cine MRI. *Radiology*. 2019;291:606–617.

23. Wolterink JM, Leiner T, de Vos BD, van Hamersvelt RW, Viergever MA, Isgum I. Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks. *Med Image Anal*. 2016;34:123–136.

24. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, Lassen MH, Fan E, Aras MA, Jordan C, Fleischmann KE, Melisko M, Qasim A, Shah SJ, Bajcsy R, Deo RC. Fully automated echocardiogram interpretation in clinical practice. *Circulation*. 2018;138:1623–1635.

25. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med*. 2019;25:65–69.

26. Clifford GD, Liu C, Moody B, Springer D, Silva I, Li Q, Mark RG. Classification of normal/abnormal heart sound recordings: the physionet/computing in cardiology challenge 2016. *2016 Comput Cardiol Conf (CinC)*. 2016;43:609–612. Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7868816.

27. Potes C, Parvaneh S, Rahman A, Conroy B. Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds. *2016 Comput Cardiol Conf (CinC)*. 2016;2016:621–624.

28. Mallya S, Overhage M, Srivastava N, Arai T, Erdman C. Effectiveness of lstms in predicting congestive heart failure onset. arXiv preprint arXiv:1902.02443. 2019. Available at: https://arxiv.org/abs/1902.02443. Accessed August 11, 2019.

29. Diller GP, Kempny A, Babu-Narayan SV, Henrichs M, Brida M, Uebing A, Lammers AE, Baumgartner H, Li W, Wort SJ, Dimopoulos K, Gatzoulis MA. Machine learning algorithms estimating prognosis and guiding therapy in adult congenital heart disease: data from a single tertiary centre including 10 019 patients. *Eur Heart J*. 2019;40:1069–1077.

30. Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. *Bioinformatics*. 2005;21:3301–3307.

31. Dastin J. Amazon scraps secret ai recruiting tool that showed bias against women. 2018. Available at https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G. Accessed May 24, 2019.

32. U.S. Food and Drug Administration. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)–based software as a medical device (SAMD)—discussion paper and request for feedback. 2019. Available at https://www.fda.gov/media/122535/download. Accessed May 24, 2019.

33. Group S-I, Cerf, French Radiology C. Artificial intelligence and medical imaging 2018: French radiology community white paper. *Diagn Interv Imaging*. 2018;99:727–742.

34. Waymel Q, Badr S, Demondion X, Cotten A, Jacques T. Impact of the rise of artificial intelligence in radiology: what do radiologists think? *Diagn Interv Imaging*. 2019;100:327–336.

35. Niederer SA, Lumens J, Trayanova NA. Computational models in cardiology. *Nat Rev Cardiol*. 2019;16:100–111.

36. Kim B, Azevedo VC, Thuerey N, Kim T, Gross M, Solenthaler B. Deep fluids: a generative network for parameterized fluid simulations. arXiv preprint arXiv:1806.02071. 2018. Available at: https://arxiv.org/abs/1806.02071. Accessed August 11, 2019.

37. Jamal S, Ali W, Nagpal P, Grover S, Grover A. Computational models for the prediction of adverse cardiovascular drug reactions. *J Transl Med*. 2019;17:171.

38. Walonoski J, Kramer M, Nichols J, Quina A, Moesel C, Hall D, Duffett C, Dube K, Gallagher T, McLachlan S. Synthea: an approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *J Am Med Inf Assoc*. 2017;25:230–238.

39. Choi E, Biswal S, Malin B, Duke J, Stewart WF, Sun J. Generating multi-label discrete patient records using generative adversarial networks. arXiv preprint arXiv:1703.06490. 2017. Available at: https://arxiv.org/abs/1703.06490. Accessed August 11, 2019.

40. Seetharam K, Kagiyama N, Sengupta PP. Application of mobile health, telemedicine and artificial intelligence to echocardiography. *Echo Res Pract*. 2019;6:R41–R52.

41. Schwamm LH, Chumbler N, Brown E, Fonarow GC, Berube D, Nystrom K, Suter R, Zavala M, Polsky D, Radhakrishnan K, Lacktman N, Horton K, Malcarney MB, Halamka J, Tiner AC. Recommendations for the implementation of telehealth in cardiovascular and stroke care: a policy statement from the American Heart Association. *Circulation*. 2017;135:e24–e44.