Machine Learning Approaches in Cardiovascular Imaging

Mir Henglin, BA; Gillian Stein, BA; Pavel V. Hushcha, MD; Jasper Snoek, PhD; Alexander B. Wiltschko, PhD; Susan Cheng, MD, MMSc, MPH

Abstract—Cardiovascular imaging technologies continue to increase in their capacity to capture and store large quantities of data. Modern computational methods, developed in the field of machine learning, offer new approaches to leveraging the growing volume of imaging data available for analyses. Machine learning methods can now address data-related problems ranging from simple analytic queries of existing measurement data to the more complex challenges involved in analyzing raw images. To date, machine learning has been used in 2 broad and highly interconnected areas: automation of tasks that might otherwise be performed by a human and generation of clinically important new knowledge. Most cardiovascular imaging studies have focused on task-oriented problems, but more studies involving algorithms aimed at generating new clinical insights are emerging. Continued expansion in the size and dimensionality of cardiovascular imaging databases is driving strong interest in applying powerful deep learning methods, in particular, to analyze these data. Overall, the most effective approaches will require an investment in the resources needed to appropriately prepare such large data sets for analyses. Notwithstanding current technical and logistical challenges, machine learning and especially deep learning methods have much to offer and will substantially impact the future practice and science of cardiovascular imaging. (***Circ Cardiovasc Imaging***. 2017;10:e005614. DOI: 10.1161/CIRCIMAGING.117.005614.)

Key Words: algorithms ■ artificial intelligence ■ automation ■ workflow

The technical capabilities of cardiovascular imaging modalities are rapidly growing and producing vast amounts of data. Clinicians and researchers alike have more opportunities than ever before to engage in the development and evaluation of novel image analysis algorithms with the ultimate goal of creating new tools to optimize patient care. Herein, we provide a framework for understanding current and future approaches to using machine learning in the increasingly data-rich arena of cardiovascular imaging.

Learning From Medical Images

Derived from both statistics and artificial intelligence, machine learning is a rapidly expanding field focused on building systems that make accurate predictions from data (Figure 1). These systems are built not by explicitly programming large sets of rules into a computer but by writing programs that can automatically learn those rules from the available data by example. Given the availability of large human-labeled data sets, many industry domains are now entirely reliant on machine learning. Examples include email spam filtering, online advertising, speech recognition, text translation, and image recognition. In medicine, early applications of machine learning can be traced back to algorithms such as the Patient Outcomes Research Team Score, which became a widely used tool for assessing the severity of pneumonia.[1,2] Image analysis algorithms include those commonly used to aid in the interpretation of ECGs. More recent advances in research include algorithms that can identify retinopathy from retinal scans[3] and grade biopsy-positive skin malignancies from photographs of skin lesions.[4]

With respect to cardiovascular imaging, machine learning can augment clinical and research activities in many ways, and specific approaches can be described using the terminology in Table 1. Machine learning methods have to date been used in 2 broad and interconnected areas: (1) automating tasks that might otherwise be performed by a human and (2) generating clinically important new knowledge, distilled from large amounts of imaging data. In Table 2, we summarize recently published studies that have used machine learning methods to analyze cardiovascular images collected from ultrasound, computed tomography (CT), magnetic resonance imaging, and nuclear imaging platforms. Most studies have focused on task-oriented problems, but studies of algorithms aimed at uncovering novel clinical insights—including new ways to predict mortality based on existing clinical data—are increasing in frequency.[16]

The types of problems that machine learning is able to solve can be illustrated by referring to an outline of image analysis workflow (Figure 2) and by using examples from ECG algorithm development. For the standard 12-lead ECG, humans were originally needed to manually perform the tasks of data acquisition, cleaning, and interpretation. Over time, algorithms were created to minimize motion artifacts and produce cleaner ECG tracings (ie, registration) and then to identify and provide reproducible measurements of PR

© 2017 American Heart Association, Inc.

Circ Cardiovasc Imaging is available at http://circimaging.ahajournals.org　　　　DOI: 10.1161/CIRCIMAGING.117.005614
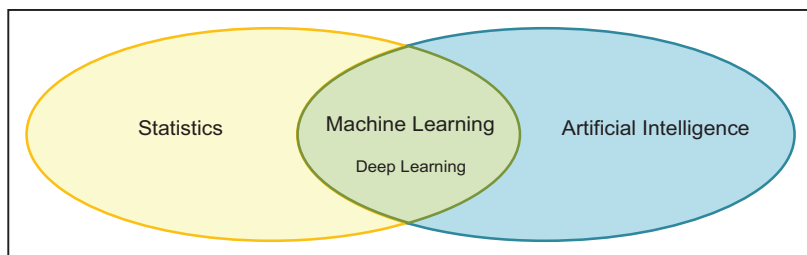
Figure 1. Machine learning in context.

and QRS, as well as other interval durations (ie, segmentation). Advanced algorithms can now assign clinical traits to a patient's ECG (ie, labeling): more straightforward labeling might include defining the presence of chamber hypertrophy, while more sophisticated labeling might involve distinguishing between ST-segment–elevation myocardial infarction and pericarditis. When an algorithm analyzes ECG features to assign the correct label and, in effect, to predict whether the tracing is from a patient with ST-segment–elevation myocardial infarction or pericarditis, the task may be accomplished by using a combination of supervised and unsupervised learning methods (Figure 2). Supervised methods require the use of tracings labeled by a cardiologist to teach an algorithm how to recognize malignant versus benign patterns. Meanwhile, unsupervised methods allow an algorithm to identify individual ECG features that tend to cluster (ie, co-occur) in the setting of ST-segment–elevation myocardial infarction or pericarditis. By operationalizing an extensive part of the total workflow (including registration, segmentation, and labeling), an algorithm-driven approach to analyzing an ECG tracing can come close to replicating the work of a human expert, thus, saving on time and human resources. As an example, automated external defibrillator algorithms have averted the absolute need for trained personnel to decide when to shock in the setting of a cardiac arrest. It is important to note, however, that no new knowledge is generated from the process just described (above and beyond what a human expert would produce). Therefore, further research is now underway to explore the extent to which machine learning techniques can identify new inter-relations between previously characterized ECG features and even use original tracing data to identify novel ECG features of potential clinical importance.[17–19]

Learning From Cardiovascular Images

A typical cardiovascular imaging data set is much larger and higher in dimension than a typical ECG data set. However, the same general framework applies with respect to how machine learning techniques can be used to augment various parts of the image analysis workflow (Figure 2).

Information Extraction

The first important task is to acquire high-quality imaging data. No matter how flexible, all algorithms will require some interface customization with respect to locale, institution, and modality-specific configurations and infrastructures. As cardiovascular imaging modality techniques continue to evolve, algorithms also need to accommodate measures (ie, variables) and layers of imaging data added over time; thus, highly customized algorithms may need substantial edits to handle new data types. With respect to optimizing image quality, filtering

and other algorithms can be applied in real time even though interindividual variation in clinical characteristics will always lead to variations in image quality.

Most data sets will inevitably require some amount of data cleaning. An important part of preprocessing image-based data is registration, which involves aligning all images into a standardized format, so that specific imaging features or anatomic regions of interest are displayed in time and in space as consistently as possible. Such procedures are common for modalities, such as nuclear imaging, wherein patient scans are frequently subject to attenuation correction. Registration also refers to the assimilation of multiple intrinsic images (eg, T1, T2, and diffusion tensor imaging for cardiovascular magnetic resonance imaging) and the aggregation of images from different views to create a single coherent image (eg, 3D CT reconstruction of coronary artery anatomy). The ultimate goal of registration is to standardize image quality and views across multiple patients to facilitate cohort-wide and even population-scale analyses.

After registration, there is the need to identify anatomic regions of interest for analysis (ie, segmentation). Segmentation traditionally requires an expert to identify which parts of the image are deemed important (eg, cardiac chamber structures, myocardial wall regions, coronary artery branches, valves, etc). Specialized segmentation may be required to identify specific regions of interest (eg, the middle scallop of the posterior mitral valve leaflet) or define the parameters of a novel measure (eg, skewness of pixel density distribution within a prespecified region of interest). Some early successes in automating segmentation include the application of a random forest classifier to CT angiography data to efficiently and accurately segment the pericardium and calculate volume of epicardial fat.[20] More complex segmentation tasks require defining not only a specific anatomic structure but also during what parts of the cardiac cycle that structure should be measured. A great deal of effort has been devoted to improving automated segmentation given the potential to avoid what is usually a laborious, time-consuming, and error-prone process when performed by humans (Table 2).

The information extraction portion of the image analysis workflow, once completed, should produce a data set that includes features ready for measurement, as well as an organized structure of data elements that can be used to potentially create novel features for analysis.

Image and Data Analysis

Traditionally, data analysis involves trained technicians selecting anatomic structures and performing measurements that are over-read by a cardiologist or radiologist who often adds diagnostic information to the record. The data are then combined

Table 1.   Glossary of Terms

Term: Definition
Registration is used to align multiple images into a single integrated image. The original images may be from different slices, views, times, and modalities. Registration is intended to overcome distortions, such as those from artifact, attenuation, rotation, scale, and skew, that will vary from image to image. Registration is required to combine different images together into a more complete source of information.
Segmentation involves dividing an image into multiple meaningful parts, with regions of interest clearly identified. Segmentation is typically used to identify objects, boundaries, or other relevant information contained in an image.
Artificial intelligence refers to the broad set of academic disciplines within computer science that strives to use computer hardware and software to build systems capable of goal-directed behavior. The term may also refer to the constructed system itself.
Machine learning is a computer science discipline and a subfield of both artificial intelligence and statistics (Figure 1). Machine learning is focused on teaching computers to perform predictive tasks without explicitly programming in the rules to perform this task. It involves getting computers to learn from experience, which is typically provided in the form of data, through fitting complex statistical models. As such, machine learning and statistics are closely related fields, whereby many machine learning concepts are connected to or have a history in statistics.
Unlabeled data are any data not associated with any clinical trait or outcome of interest. Typically, unlabeled data consist of samples when they are first generated or measured. Some examples of unlabeled imaging data might include raw ultrasound, CT, MR, or nuclear images. There are no descriptors or categories ascribed to unlabeled data.
Labeled data result from associating unlabeled data with one or more meaningful descriptions. A label may be the definition of a measurement, the definition of a clinical trait, or the definition of a clinical outcome. For instance, a linear measure may be labeled as LVEF, a binary variable may be labeled as denoting the presence or absence of LV hypertrophy, and another binary variable may be labeled as denoting the presence or absence of heart failure. Labels for data are often obtained by asking humans to carefully analyze or make judgments about unlabeled data (eg, asking a technical expert to trace the LV endocardium in multiple views to derive a biplane Simpson LVEF or asking an expert over-reader to adjudicate the presence or absence of rheumatic mitral valve disease). Thus, the process of labeling data often incurs substantial time and resource costs. The most successful machine learning algorithms, namely supervised learning algorithms, all require labeled data.
Weak labels are labels that convey limited information but are easier to create than nonweak labels. A data set may be said to have weak labels if the labels are inaccurate, sparse (eg, the data set is missing many labels), or incomplete (eg, the labels indicate the presence, not the location, of a tumor in an image). Training most modern methods for segmentation or abnormality detection requires manually labeled data, but acquiring large sets of reliable labels in many clinical machine learning tasks requires large amounts of time from technical experts. This process may be prohibitively difficult because of not only the monetary cost of expert attention but also because of the fact that, for some labeling tasks, even domain experts may disagree on how to label a particular image (ie, interobserver reproducibility). It is, therefore, of interest to find ways to develop algorithms that could learn from data with weaker labels that are less expensive, possibly less accurate, but that are much more easily obtainable.
Supervised learning algorithms are machine learning algorithms that fit models using pairs of input features and labeled data. The goal of these models is to correctly predict the label associated with some input features. An input feature might be an image of the LV, and the label might be a number indicating that the LV is present versus absent (or that the structure is an LV and not an atrium or a valve). Examples of supervised learning algorithms include linear regression, logistic regression, support vector machines, and decision trees. Supervised learning is currently the most successful type of machine learning—it is both the subject of the majority of machine learning research and it is by far the most commercially successful application of machine learning.
Unsupervised learning algorithms are models constructed using unlabeled data. These models attempt to capture relationships inherent to the structure of the features themselves. Examples include clustering and principal components analyses.
Neural networks are a family of supervised and unsupervised learning algorithms characterized by stacked layers of processing, often alternating linear and nonlinear transforms. Neural networks, historically known as multilayer perceptrons, were originally inspired by how the brain processes information but are now widely regarded to be only weakly related to real biological systems.
Deep learning is a subfield of machine learning, concerned with the research and development of deep neural networks. Deep neural networks are neural networks with many layers stacked on top of each other (often 5–25, but sometimes hundreds of layers deep). Until recently, training these algorithms on very large data sets (eg, millions of images) was completely impractical. The advent of faster machines, in particular the GPU, enabled the resurgence of this technique. Deep neural networks have led to major breakthroughs recently in speech recognition, machine translation, natural language, and image processing.
CNN are deep neural networks that are characterized by having efficient convolutional operations as the base layers of the network. Originally introduced for handwritten digit recognition,[5] convolutional networks use sliding window operations on images rather than per-pixel parameters. This helps to save memory and computation while encoding useful translational invariance into the model (eg, the model is not reliant on identifying a specific value for a specific pixel). Recent methodological advancements in computing and their application to large data sets have led to breakthroughs across many imaging tasks using CNNs.[6]
GPU is specialized computer hardware built specifically to expedite the processing of operations typically occurring in computer graphics. Many of these operations involve extensive use of linear algebra, which is also the bulk of processing in deep learning. The fitting of deep learning models can be accelerated many times over through the use of this specialized hardware.
Transfer learning is related to the concept that humans can apply relevant knowledge from previous learning experiences to new tasks. Most machine learning algorithms, by contrast, address isolated tasks. A model trained with one set of features and labels cannot be adapted to similar tasks without full retraining, which is often costly and time-consuming. Transfer learning refers to taking knowledge gained on an original task and efficiently learning to perform well on a separate but related task, often with much less data required than if the related task was attempted in isolation.

CNN indicates convolutional neural network; CT, computed tomography; GPU, graphics processing unit; LV, left ventricle; LVEF, left ventricular ejection fraction; and MR, magnetic resonance.

Table 2. Overview of Machine Learning Algorithms Applied in Cardiovascular Imaging Studies

Author	N	Study Design	Methods	Measures	Main Findings
Berchialla et al[7]	228	Cross-sectional	• Bayesian network • Logistic regression • Random forest • Artificial neural network • SVM	• Use data from stress echo and CTA to predict future cardiovascular events (myocardial infarction or death)	• Bayesian network outperformed other methods • Measures of LV dysfunction and CAD extent had greater impact in predicting target event
Isgum et al[8]	584	Longitudinal	• Linear and quadratic discriminant • k-NN • SVM	• Automatically score coronary calcium in low-dose, noncontrast-enhanced chest CT scans	• Cardiovascular risk was best determined by merging results of 3 best-performing classifiers (2-stage classification with k-NN, 2-stage classification with k-NN and SVM, 1-stage classification with k-NN with selected features) • Detected on average 157/198 mm³ (sensitivity 79.2%) of coronary calcium volume with average 4 mm³ false-positive volume
Lee et al[9]	205	Cross-sectional	• Decision tree • Naive Bayes • k-NN • SVM	• Analyze AAA geometry on contrast CT images • Determine whether AAA wall surface curvatures predict rupture risk	• k-NN demonstrated the highest accuracy (85.5% compared with 68.9% using maximum diameter alone) • Accuracy of SVM, decision tree, and naive Bayes was 83.4%, 83.3%, and 80.1%, respectively
Mohammadpour et al[10]	115	Cross-sectional	Fuzzy rule-based classifying system	Use myocardial perfusion scan and clinical variables to predict CAD	Classifier determined most important risk factors for CAD and correctly detected patients who did not need invasive coronary angiography with 92.8% accuracy
Xiong et al[11]	140	Cross-sectional	• Naive Bayes • Random forest • AdaBoost	• Determine physiological manifestation of coronary stenoses by assessing myocardial perfusion on CTA images	• Method may improve diagnosis of obstructive coronary artery stenoses • AdaBoost performed better than other algorithms with accuracy 0.70, sensitivity 0.79, and specificity 0.64
Knackstedt et al[12]	255	Cross-sectional	• Vendor-independent software AutoLV	• Obtain measures of LV volumes, EF, and average biplane longitudinal strain using ultrasound images • Compare values with visual estimation and manual tracking	• Algorithm was time efficient (8±1 s/patient), reproducible, and technically feasible for LVEF and longitudinal strain assessment
Arsanjani et al[13]	713	Longitudinal	• Machine learning algorithm LogitBoost	• Use SPECT perfusion data to predict early revascularization in patients with suspected CAD	• LogitBoost sensitivity (73.6±4.3%) for predicting revascularization was similar to one expert reader (73.9±4.6%) and perfusion measures only (75.5±4.5%) • LogitBoost specificity (74.7±4.2%) was better than both expert readers (67.2±4.9% and 66.0±5.0%) and similar to total ischemic perfusion deficit (68.3±4.9%) • LogitBoost AUC (0.81±0.02) was identical to one reader but superior to another reader (0.72±0.02) and perfusion measures only (0.77±0.02)

Table 2.   Continued

Author	N	Study Design	Methods	Measures	Main Findings
Berikol et al[14]	228	Longitudinal	• SVM • Artificial neural network • Naive Bayes • Logistic regression	• Diagnose acute coronary syndrome and decide whether to discharge or admit patients considering their symptoms, electro- and echocardiographic findings, levels of cardiac enzymes	• SVM had the highest predicting accuracy 99.13%, sensitivity 98.22%, and specificity 100% • Accuracy of artificial neural network, naive Bayes, and logistic regression was 91.26%, 88.75%, and 90.1%, respectively
Celutkiene et al[15]	256	Longitudinal	Custom multiparametric mathematical model	Analyze dobutamine stress echocardiography with speckle tracking (compared with conventional wall motion analysis) to detect myocardial ischemia	Algorithm detected myocardial ischemia in patients with coronary stenoses ≥50% with sensitivity 91.6% and specificity 86.3%, compared with 76.8% and 89%, respectively, for visual assessment
Motwani et al[16]	10 030	Longitudinal	Custom-built predictive classifier	Predict 5-year all-cause mortality in patients with suspected CAD undergoing CCTA	• Method showed performance superior to use of clinical and CCTA findings alone • AUC was 0.79 vs 0.61 for Framingham risk score, 0.64 for segment stenosis score, 0.64 for segment involvement score, and 0.62 for modified Duke index

AAA indicates abdominal aortic aneurysm; AUC, area under the curve; CAD, coronary artery disease; CCTA, coronary CTA; CT, computed tomography; CTA, CT angiography; EF, ejection fraction; LV, left ventricle; k-NN, k-nearest neighbor; SPECT, single positron emission computed tomography; and SVM, support vector machine.

with clinical data into a single data set, and conventional statistics are used to determine whether a given measurement is relevant to a clinical outcome. Machine learning approaches have now expanded to assist not only with registration and segmentation but also with performing the measurements normally made by a human. Just as ECG algorithms have continued to improve the ability to correctly identify PR, QRS, and QT intervals on a 12-lead tracing and, in turn, determine measurements of these interval durations, machine learning algorithms have been developed to automatically detect the left ventricle (LV) endocardium and provide measures of LV volumes and ejection fraction.[12] Many machine learning approaches are developed as part of online data science competitions, including one focused on determining ejection fraction from CMR scans collected from >1000 patients.[21] Although not yet able to perform completely reliable measures of LV ejection fraction (especially when image quality is limited), all such algorithms continue to improve in accuracy.[22]

There are 2 potential benefits of algorithms designed to automate measurements. First, automated standardized measurements can be applied to large data sets, permitting detection of subtle relations between anatomic variance and clinical outcomes. For instance, automated measures of mitral valve thickening and prolapse could facilitate large-scale analyses of early mitral valve disease while avoiding inter-reader bias and the need for case–control studies of manually curated measurements.[23–25] Second, advanced measures that are expensive or time-consuming to perform can also be automated and applied to large cohorts if the advanced measures are closely related to standard measurements. Examples of advanced measures

include coronary plaque volume by CT, extracellular volume fraction by cardiovascular magnetic resonance, coronary flow reserve by positron emission tomography, and myocardial deformational strain by echocardiography. Although a human may need specialized training to learn how to reproducibly perform speckle tracking strain analysis of the LV[26], for instance, an algorithm already optimized to trace the LV for conventional volumetric measurements can be adjusted to automate endocardial tracings for speckle tracking measures.[27]

Beyond automating measurements, a well-developed strength of machine learning is in the analysis of the measurement data in relation to outcomes (ie, prediction). When building a predictive model, both machine learning and conventional statistical techniques will attempt to characterize the relationships between predictors (eg, imaging-based measures) and outcomes (eg, cardiovascular events). However, traditional statistics will create models that permit describing these relationships in easily comprehensible terms, such as an odds ratio. Although the field of machine learning includes the application of conventional statistics (Figure 1), newer machine learning approaches will fit models to capture the relationships between predictors and outcomes with the highest degree of fidelity possible, even at the expense of easy interpretability. Given their flexibility, modern machine learning methods are especially valuable when (1) the most accurate predictions are desired, even at the expense of a clear understanding of how these predictions are made or (2) the relationships between predictors and outcomes are likely to be complex or nonlinear, such that interpretability is not
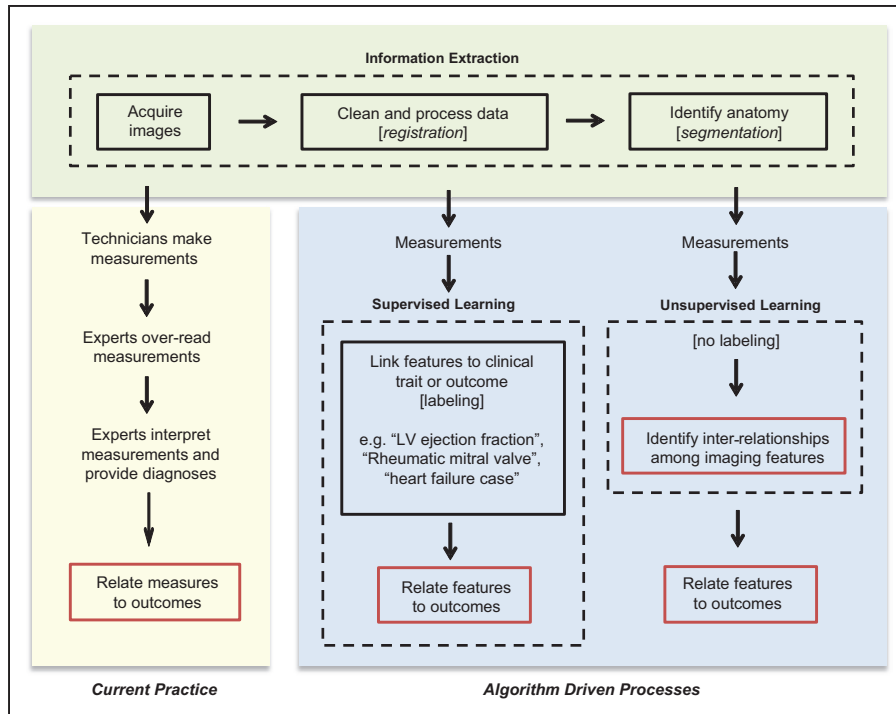
Figure 2. Image analysis workflow. LV indicates left ventricle.

expected from the outset. In these situations, traditional statistical methods may be extremely hampered or completely fail.[28] However, such situations are relevant to cardiovascular imaging databases because it is often unclear at the outset what input imaging features will matter for predicting a clinical outcome while the imaging features themselves are often too numerous or complex to decipher manually.

The ability of machine learning to handle large high-dimensional data sets is also attractive if only given the sheer size of cardiovascular imaging databases. It is well recognized that all imaging modalities routinely produce a vast amount of data per patient. A single CT scan results in tens to hundreds of images, each comprising millions of pixels. Cine imaging expands this volume of data many times over. Thus, there is enormous potential gain from comprehensively analyzing not only typical measurement data but also image-based raw data, from which new measures can be generated. However, if the task of predicting clinical outcomes is to begin at the raw pixel level for a given imaging study, a conventional data analysis approach is quickly overwhelmed by all the possible combinations of pixels, filters, or image processing techniques that could be used in attempts to reveal a relationship between the image and an outcome. Modern machine learning techniques are able to automatically discover such relationships at the pixel level, primarily via deep learning.

Deep Learning

Deep learning is a type of machine learning that includes a class of algorithms called neural networks, which are intended to model high-level abstractions of data from stacked layers of processing, often alternating linear and nonlinear transformations. Deep neural networks, consisting of up to dozens or even hundreds of layers stacked on top of one another, have led to major breakthroughs in image, speech, and text processing. These methods are currently considered state of the art for making predictions from imaging data. In particular, convolutional neural networks (CNNs),[5] a type of deep neural network, can automatically learn to discover and combine local image features (such as an edge or a color contrast) in increasing levels of abstraction to ultimately enable prediction of an outcome. Although CNNs are most effective only when applied to large data sets, medical imaging databases of adequate size are now commonplace. Special advantages of CNNs include their ability to manage complicated relationships between the inputs (ie, image data) and the outputs (ie, outcomes) that are not easily captured by manual measurements.

When applied to image-based data sets, ideally large in size,[29] deep learning algorithms enable unassisted approaches to discovering and testing new imaging features. Although still in development, some approaches to creating novel features may even be applied to completely unlabeled images.[30] Deep learning algorithms, in certain circumstances, can perform unsupervised learning, which is important because most existing imaging data sets are not linked to a clearly defined outcome at the outset. In some cases, training on a large body of unlabeled data can allow an algorithm to perform better when applied to only a small amount of labeled data.[31] Using deep learning to perform unsupervised feature generation is attractive for many reasons, including efficient and seamless potential application to almost any domain and problem. The main drawback, however, is the interpretability of any newly discovered imaging feature. Whether applied to imaging or nonimaging data sets, deep learning creates features that are often unintuitive and difficult to comprehend. Recognizing this limitation, ongoing research is focused on stimulating neural networks to visually manifest the most activating characteristics or regions of an image (Figure 3).[30,31]

It is possible that the most productive approaches to obtaining knowledge from large volumes of imaging data will involve combining deep learning features with adjunctive information. Technologies that integrate otherwise disparate information sources can possibly enable synthesis of all available information related to a particular medical image, including data pertaining to image acquisition, text from preliminary and finalized cardiovascular imaging and radiological reading reports, and data from clinical notes, complementary diagnostics, and any additional records that are related to the imaged patient (Figure 3). An analysis of images and associated source data can potentially lead to improved clustering, classification, and, in turn, discovery of precision phenotypes.[32,33] The discovery of any novel imaging biomarker, of course, will still require validation and rigorous further investigation before its integration into practice.

Notwithstanding all advances to date, it should be emphasized that a single deep learning network cannot easily synthesize images from different types and sources of data in the way that humans are currently able to do when performing a clinical or research evaluation. One day, such complex and intricate tasks may be handled by a collection of networks, each assigned to a different type of data set. Another option is transfer learning, which involves applying knowledge from one task to address another task (Table 1) and this approach may be helpful for related data sets. Ideally, CNNs trained on a large and laboriously hand-labeled data set can be retuned to perform relatively well on related smaller-sized data sets, so that a model that is well trained on one large original task can be leveraged to succeed quickly at another. For example, we may train a CNN to predict 1-year risk for cardiovascular death from a large original data set of echocardiograms. Later, for a separate data set, this same CNN can be repurposed to predict the likelihood of hospital readmission directly from ultrasound images even if the new data set may only be 1 of 10th or 1 of 20th the size of the original.

Challenges and Future Directions

The highly trained human act of creating a finely tuned and nuanced interpretation of a medical image involves assimilating objective data with intuition, prior clinical experience, and internalized knowledge that is not in the medical record. Therefore, while machine learning approaches rapidly advance in capabilities, their performance will always be limited by the availability and quality of the data and labels from which they can learn. Accordingly, machine learning algorithms may remain limited in areas of cardiovascular imaging that lack an abundance of data for training an algorithm (eg, rare diseases, historic and deeply archived images, and image-related data or text that may not be easy to access). Similarly, in the absence of large training data sets, algorithms may continue to be limited in identifying uncommon presentations of common diseases. Given these limitations along with challenges pertaining to the interpretability of results from deep learning algorithms, we anticipate that even the most advanced machine learning approaches are more likely to offer a powerful complement to, rather than a complete replacement of, expert over-reading.

A perspective on the near-future role of machine learning as an aid in the clinical practice of cardiovascular imaging

may again be borrowed from electrocardiography. Algorithms that assisted with interpreting ECGs were initially promoted and provided by diagnostic equipment companies and came to be widely used clinically by the 1990s.[34] Although the diagnostic accuracy of ECG algorithms compared with cardiologists was initially questioned, their performance improved substantially over time. In some cases, algorithms may even be superior to cardiologists for characterizing some types of ECG traits (eg, diagnosing broad complex tachycardia).[35] However, algorithms may underperform for other traits (eg, diagnosing atrial fibrillation) and can even inadvertently lead over-reading cardiologists to misdiagnose more frequently when compared with cardiologists interpreting an ECG in clinical context without considering any input from an algorithm.[36] Thus, while ECG algorithms continue to improve alone or in combination,[37] certain limitations in performance persist. Nonetheless, the diagnostic accuracy of current ECG algorithms is high,[38] such that they now serve as an essential decision support tool routinely used in practice.

Despite the notion that machine learning is more likely to support rather than supplant clinical decision making, the potential to substantially transform the practice of medical imaging remains. There is a hope that, one day, whole imaging studies in their original raw digital imaging and communications in medicine format, which are collected from large numbers of patients in a clinical practice, can be usefully interpreted by advanced machine learning algorithms without the need for any preprocessing or labeling. Although an ideal goal, achieving such a scenario will continue to be challenging, while there remains a lack of standardization in how images are acquired and processed before being analyzed. Methods for automating registration and segmentation continue to be in development, but there is still much work to be done before their performance can match that of trained human curators for most imaging modalities. Importantly, labeling remains a critical step for algorithm development, and supervised learning methods, especially deep learning, require large and accurately labeled data sets. Although technical experts can be recruited to assist in performing all the steps of data processing required before machine learning analyses (ie, registration, segmentation, and labeling), the best algorithms are trained on the largest data sets, and the human effort required to prepare such data sets is both costly and subject to human variations in performance. Inadequate or poorly conducted data processing will lead to ambiguous data generation, and noninformative inputs lead to noninformative outputs. However, there are algorithms in development that can learn from weak labels (ie, labels that are incomplete or otherwise easier to obtain) at the cost of reliability or accuracy. Also, if nonexperts are, on average, able to correctly label an image with minimal training, researchers can crowdsource the generation of labeled data sets. For complex pathologies, however, crowdsourcing tasks will need to be compartmentalized, organized, and staged with varying degrees of expert review involved at multiple stages. Thus, the extent to which crowdsourcing alone can be used to effectively or efficiently develop algorithms for complete cardiovascular imaging interpretation, in a given setting, remains to be explored.

Beyond data set synthesis, a current practical limitation for running algorithms relates to the fact that handling enormous
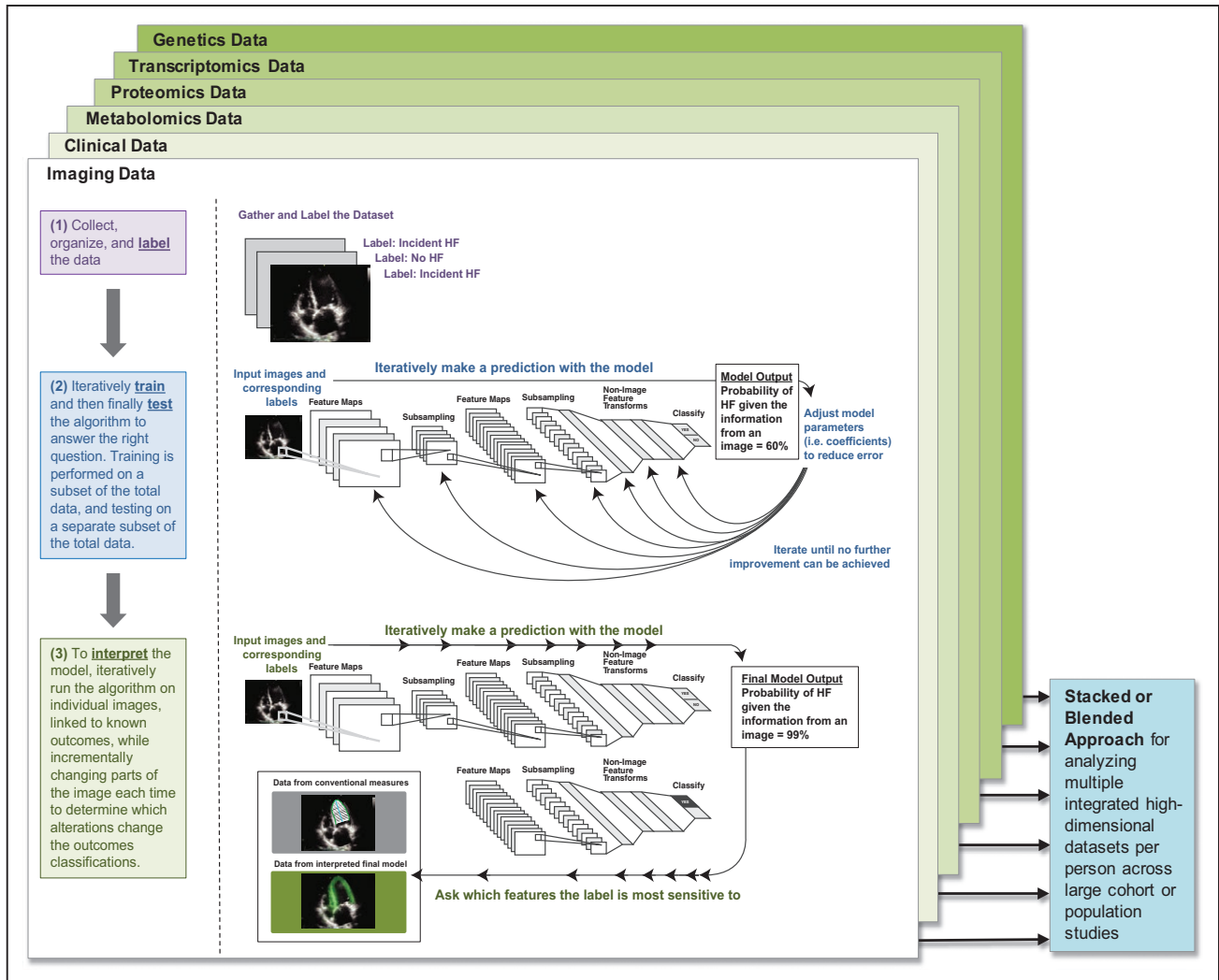
Figure 3. Image analysis in context. HF indicates heart failure.

amounts of data requires an extraordinary amount of computational power. Fortunately, large-scale computational resources are becoming more accessible. In particular, graphics processing units have been optimized for the computational tasks on which machine learning models depend. The price of these devices has been declining, and they are also readily available for ad hoc rental on cloud platforms. A flexible cloud computing environment, notwithstanding access and privacy issues that continue to be evaluated,[39] will be ideal for facilitating integration of imaging data that tend to exist in variable formats across multiple institutions and in siloed storage. On the research front, publicly funded initiatives are now underway to make biorepositories of imaging and imaging-derived data, collected from across research studies, available to the scientific community for advanced and large-scale analyses via cloud computing.

Conclusions

Machine learning approaches have formed the core of many cardiovascular image acquisition and processing algorithms that are already in routine use. Given the rapid evolution of machine learning capabilities, continued advancements are being made in developing tools for optimizing not only how cardiovascular

imaging measurements are performed but also how the results of these measurements can be interpreted. Currently available machine learning methods, particularly those based on deep learning, have generated growing interest in their potential to derive new insights from image-related data, as well as the images themselves, given the expanding size of existing databases. Increasingly large databases, however, will require increasing resources to create high-quality labels for enabling effective analyses. Therefore, continued progress will depend on a commitment to thoughtfully and strategically investing in such resources. Notwithstanding ongoing technical and logistical challenges facing the field, machine learning and particularly deep learning methods are very likely to substantially impact the future practice and science of cardiovascular imaging.

Sources of Funding

Disclosures

None.

References

1. Cooper GF, Aliferis CF, Ambrosino R, Aronis J, Buchanan BG, Caruana R, Fine MJ, Glymour C, Gordon G, Hanusa BH, Janosky JE, Meek C, Mitchell T, Richardson T, Spirtes P. An evaluation of machine-learning methods for predicting pneumonia mortality. *Artif Intell Med*. 1997;9:107–138.

2. Fine MJ, Auble TE, Yealy DM, Hanusa BH, Weissfeld LA, Singer DE, Coley CM, Marrie TJ, Kapoor WN. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med*. 1997;336:243–250. doi: 10.1056/NEJM199701233360402.

3. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316:2402–2410. doi: 10.1001/jama.2016.17216.

4. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115–118. doi: 10.1038/nature21056.

5. LeCun L, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE*. 1998;86:2278–2324.

6. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Proc Adv Neural Inf Process Sys*. 2012;25:1097–1105.

7. Berchialla P, Foltran F, Bigi R, Gregori D. Integrating stress-related ventricular functional and angiographic data in preventive cardiology: a unified approach implementing a Bayesian network. *J Eval Clin Pract*. 2012;18:637–643. doi: 10.1111/j.1365-2753.2011.01651.x.

8. Isgum I, Prokop M, Niemeijer M, Viergever MA, van Ginneken B. Automatic coronary calcium scoring in low-dose chest computed tomography. *IEEE Trans Med Imaging*. 2012;31:2322–2334. doi: 10.1109/TMI.2012.2216889.

9. Lee K, Zhu J, Shum J, Zhang Y, Muluk SC, Chandra A, Eskandari MK, Finol EA. Surface curvature as a classifier of abdominal aortic aneurysms: a comparative analysis. *Ann Biomed Eng*. 2013;41:562–576. doi: 10.1007/s10439-012-0691-4.

10. Mohammadpour RA, Abedi SM, Bagheri S, Ghaemian A. Fuzzy rule-based classification system for assessing coronary artery disease. *Comput Math Methods Med*. 2015;2015:564867. doi: 10.1155/2015/564867.

11. Xiong G, Kola D, Heo R, Elmore K, Cho I, Min JK. Myocardial perfusion analysis in cardiac computed tomography angiographic images at rest. *Med Image Anal*. 2015;24:77–89. doi: 10.1016/j.media.2015.05.010.

12. Knackstedt C, Bekkers SC, Schummers G, Schreckenberg M, Muraru D, Badano LP, Franke A, Bavishi C, Omar AM, Sengupta PP. Fully automated versus standard tracking of left ventricular ejection fraction and longitudinal strain: the FAST-EFs Multicenter Study. *J Am Coll Cardiol*. 2015;66:1456–1466. doi: 10.1016/j.jacc.2015.07.052.

13. Arsanjani R, Dey D, Khachatryan T, Shalev A, Hayes SW, Fish M, Nakanishi R, Germano G, Berman DS, Slomka P. Prediction of revascularization after myocardial perfusion SPECT by machine learning in a large population. *J Nucl Cardiol*. 2015;22:877–884. doi: 10.1007/s12350-014-0027-x.

14. Berikol GB, Yildiz O, Özcan IT. Diagnosis of acute coronary syndrome with a support vector machine. *J Med Syst*. 2016;40:84. doi: 10.1007/s10916-016-0432-6.

15. Čelutkienė J, Burneikaitė G, Petkevičius L, Balkevičienė L, Laucevičius A. Combination of single quantitative parameters into multiparametric model for ischemia detection is not superior to visual assessment during dobutamine stress echocardiography. *Cardiovasc Ultrasound*. 2016;14:13. doi: 10.1186/s12947-016-0055-6.

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. doi: 10.1093/eurheartj/ehw188.

17. Liu Y, Scirica BM, Stultz CM, Guttag JV. Beatquency domain and machine learning improve prediction of cardiovascular death after acute coronary syndrome. *Sci Rep*. 2016;6:34540. doi: 10.1038/srep34540.

18. Syed Z, Stultz CM, Scirica BM, Guttag JV. Computationally generated cardiac biomarkers for risk stratification after acute coronary syndrome. *Sci Transl Med*. 2011;3:102ra95. doi: 10.1126/scitranslmed.3002557.

19. Sharma LN, Tripathy RK, Dandapat S. Multiscale energy and eigenspace approach to detection and localization of myocardial infarction. *IEEE Trans Biomed Eng*. 2015;62:1827–1837. doi: 10.1109/TBME.2015.2405134.

20. Norlén A, Alvén J, Molnar D, Enqvist O, Norrlund RR, Brandberg J, Bergström G, Kahl F. Automatic pericardium segmentation and quantification of epicardial fat from computed tomography angiography. *J Med Imaging (Bellingham)*. 2016;3:034003. doi: 10.1117/1.JMI.3.3.034003.

21. Second Annual Data Science Bowl: Transforming How We Diagnose Heart Disease. https://www.kaggle.com/c/second-annual-data-science-bowl. Accessed January 23, 2017.

22. Avendi MR, Kheradvar A, Jafarkhani H. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Med Image Anal*. 2016;30:108–119. doi: 10.1016/j.media.2016.01.005.

23. Delling FN, Rong J, Larson MG, Lehman B, Fuller D, Osypiuk E, Stantchev P, Hackman B, Manning WJ, Benjamin EJ, Levine RA, Vasan RS. Evolution of mitral valve prolapse: insights from the Framingham Heart Study. *Circulation*. 2016;133:1688–1695. doi: 10.1161/CIRCULATIONAHA.115.020621.

24. Delling FN, Rong J, Larson MG, Lehman B, Osypiuk E, Stantchev P, Slaugenhaupt SA, Benjamin EJ, Levine RA, Vasan RS. Familial clustering of mitral valve prolapse in the community. *Circulation*. 2015;131:263–268. doi: 10.1161/CIRCULATIONAHA.114.012594.

25. Delling FN, Vasan RS. Epidemiology and pathophysiology of mitral valve prolapse: new insights into disease progression, genetics, and molecular basis. *Circulation*. 2014;129:2158–2170. doi: 10.1161/CIRCULATIONAHA.113.006702.

26. Cheng S, Larson MG, McCabe EL, Osypiuk E, Lehman BT, Stanchev P, Aragam J, Benjamin EJ, Solomon SD, Vasan RS. Reproducibility of speckle-tracking-based strain measures of left ventricular function in a community-based study. *J Am Soc Echocardiogr*. 2013;26:1258–1266.e2. doi: 10.1016/j.echo.2013.07.002.

27. Wong KC, Tee M, Chen M, Bluemke DA, Summers RM, Yao J. Regional infarction identification from cardiac CT images: a computer-aided biomechanical approach. *Int J Comput Assist Radiol Surg*. 2016;11:1573–1583. doi: 10.1007/s11548-016-1404-5.

28. Fan J, Han F, Liu H. Challenges of big data analysis. *Natl Sci Rev*. 2014;1:293–314. doi: 10.1093/nsr/nwt032.

29. de Bruijne M. Machine learning approaches in medical image analysis: from detection to diagnosis. *Med Image Anal*. 2016;33:94–97. doi: 10.1016/j.media.2016.06.032.

30. Le QV, Ranzato M, Monga R, Devin M, Chen K, Corrado GS, Dean J, Ng AY. Building high-level features using large scale unsupervised learning. *Proc ICML*. 2012:507-514.

31. Angermueller C, Pärnamaa T, Parts L, Stegle O. Deep learning for computational biology. *Mol Syst Biol*. 2016;12:878.

32. Mikolov T, Deoras A, Povey D, Burget L, Cernocky J. Strategies for training large scale neural network language models. *Proc Automatic Speech Recognition and Understanding*. 2011:196–201.

33. Oakden-Rayner L, Carneiro G, Bessen T, Nascimento JC, Bradley AP, Palmer LJ. Precision Radiology: predicting longevity using feature engineering and deep learning methods in a radiomics framework. *Sci Rep*. 2017;7:1648. doi: 10.1038/s41598-017-01931-w.

34. Willems JL, Abreu-Lima C, Arnaud P, van Bemmel JH, Brohet C, Degani R, Denis B, Gehring J, Graham I, van Herpen G. The diagnostic performance of computer programs for the interpretation of electrocardiograms. *N Engl J Med*. 1991;325:1767–1773. doi: 10.1056/NEJM199112193252503.

35. Lau EW, Pathamanathan RK, Ng GA, Cooper J, Skehan JD, Griffith MJ. The Bayesian approach improves the electrocardiographic diagnosis of broad complex tachycardia. *Pacing Clin Electrophysiol*. 2000;23(10 pt 1):1519–1526.

36. Anh D, Krishnan S, Bogun F. Accuracy of electrocardiogram interpretation by cardiologists in the setting of incorrect computer analysis. *J Electrocardiol*. 2006;39:343–345. doi: 10.1016/j.jelectrocard.2006.02.002.

37. Meyer C, Fernández Gavela J, Harris M. Combining algorithms in automatic detection of QRS complexes in ECG signals. *IEEE Trans Inf Technol Biomed*. 2006;10:468–475.

38. Shah AP, Rubin SA. Errors in the computerized electrocardiogram interpretation of cardiac rhythm. *J Electrocardiol*. 2007;40:385–390. doi: 10.1016/j.jelectrocard.2007.03.008.

39. Schweitzer EJ. Reconciliation of the cloud computing model with US federal electronic health record regulations. *J Am Med Inform Assoc*. 2012;19:161–165. doi: 10.1136/amiajnl-2011-000162.