11 Artificial Intelligence in Cardiovascular Medicine

ZACHI ATTIA, SURAJ KAPA, PETER NOSEWORTHY, AND PAUL FRIEDMAN

Artificial intelligence (AI) is ubiquitous. It autocompletes the sentences we type, populates web searches before we complete our thoughts, enables our phones to understand verbal commands, permits cars to drive themselves to destinations we speak, and increasingly supports medical diagnostic tests. In medicine it has identified retinal pathology with a skill that exceeds that of a trained ophthalmologist, can tirelessly detect mammographic lesions, and identify abnormalities on a pathologic slide. Some revile it as a technology that will lead to massive unemployment, economic disruption, and serve as an existential threat to humanity; others embrace it as the tool that will liberate humanity from drudgery and elevate the most noble of human tasks.[1]

Three broad capabilities of AI apply to the field of medicine. The first is the automation of fatiguing processes that involve analysis of massive amounts of data, such as continuous ECG tracings acquired over months. In this context, AI performs human-like tasks at massive scale. AI also permits embedding technology in novel forms such as clothing and other wearables to extract physiologic information to enable continuous monitoring of health. The application of AI also, by extension, enables remote monitoring in rural locations, space exploration, and extreme conditions. The second is the ability to extract signals beyond that which a human is capable of recognizing, for example determining the presence of ventricular function from a standard 12-lead electrocardiogram or single-lead ECG acquired from a watch- or smartphone-enabled electrodes. In this context, AI brings new value to well-established medical diagnostic tests that exist in current clinical workflows and practice. Thirdly, and more broadly, the ability to specifically, richly, and uniquely characterize an individual's physiologic data allows for a new level of personalized predictive models, potentially creating a whole new category of individual "previvors" who know a disease is impending before any signs of symptoms develop, opening the doors for potential interventions, and with associated social, legal, and economic implications. This deep phenotyping may inform additional fields, such as genetics. AI in medicine is in its early stages; the promise is large, but its application requires rigorous testing, vetting, and validation, as do all tests that impact human health. Here we focus on AI and its role in cardiovascular medicine.

DEFINITIONS AND KEY TERMS

> If intelligence is a cake, the bulk of the cake is unsupervised learning, the icing on the cake is supervised learning, and the cherry on the cake is reinforcement learning (RL).
>
> *Yann LeCun, 2016*

AI is a lay term, referring to machine learning (ML). In his cake analogy, Dr. Yann LeCun* divides ML into its three main branches and presents one of the technology's main challenges—the amount of data required for implementation. In all three types of learning (supervised, unsupervised, and reinforcement), instead of using an explicit set of human-devised rules to interpret a signal, large volumes of data are fed to a model, which uses statistical processes to identify relationships within the data. In short, the data train the model, free from human hypothesis (**eFig. 11.1**).

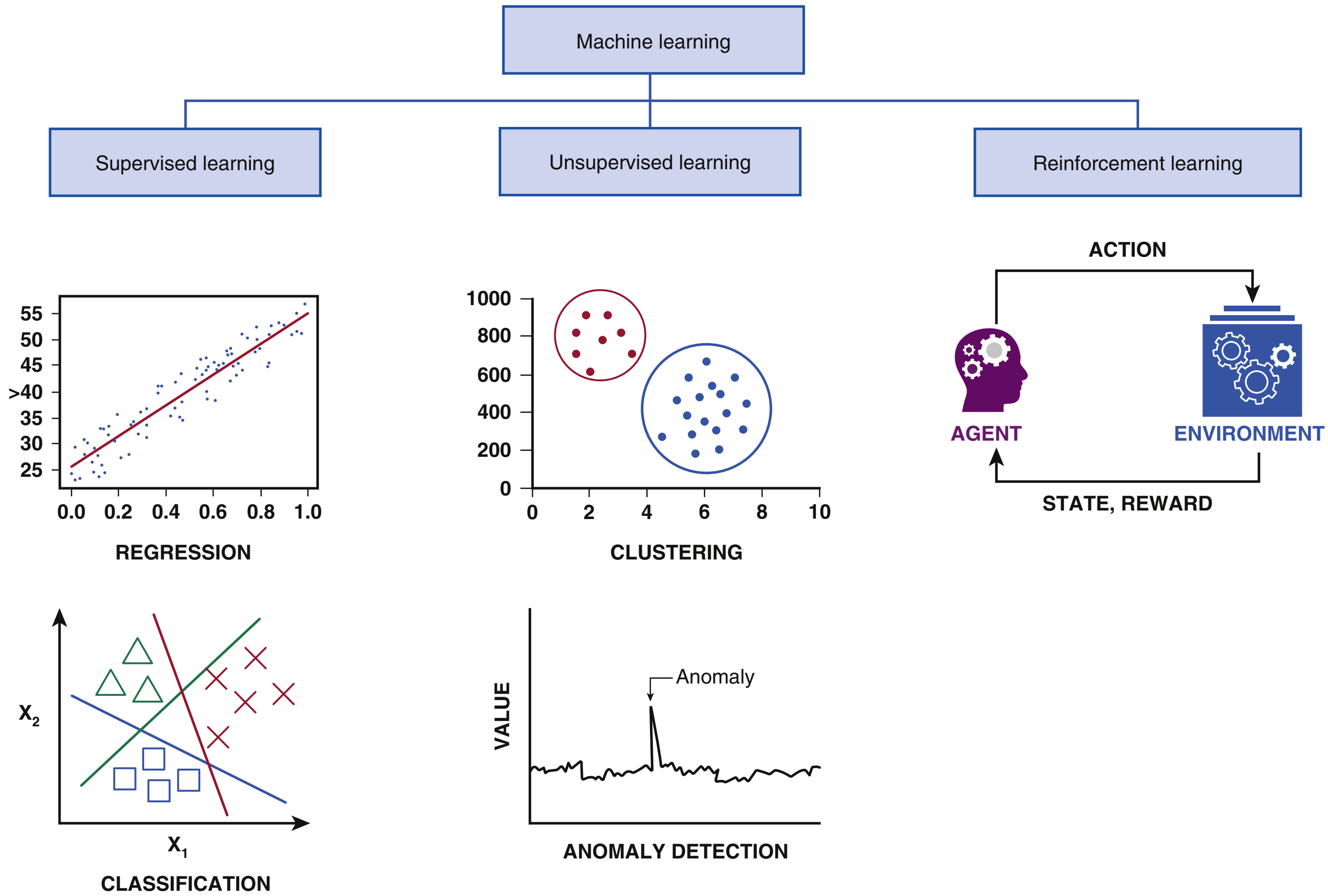
Learning

Learning is the process of improving the ability to complete a task based on experience. As the task is repeated, ML improves by getting feedback (via an error or loss function) and changing the way it performs the task (by changing with weights and biases of the mathematical functions that comprise the "neurons" in a neural network, for example), until the feedback is that the task is done correctly, or at least above a certain standard. In all three types of ML, the feedback is the loss function—the difference between a wanted outcome (how we think the task should have been performed) to the actual outcome (how the task was performed). Learning, or training, is often computationally intensive. Once trained, many networks can then operate with limited computational resources, for example on a smartphone. This makes many AI tools massively scalable.

Supervised Learning

Supervised learning is the most commonly used form of ML. Supervised learning requires labeled data (images and captions, ECGs and their rhythm description), with labels often provided by human experts. The discovery of the rules that explain the relationship between the input (a signal) to the output (a label) is called *training.* For example, if ECG samples labeled normal rhythm or atrial fibrillation (AF) are fed to a model, it will learn to differentiate between the two rhythms. The specific features of the signal used to generate model output are determined by the computer during training and are not discernible to humans (Fig. 11.1). Thus, AI is at times referred to as a "black box." In most cases, the model will be a parametric function (F) of the inputs, and it will be initialized using a set of random parameters (weights). During training, in an iterative manner, F is applied on a set of inputs with known outputs (the labels). The results of applying the function

*Yann LeCun, Geoffrey Hinton and Yoshua Bengio—often referred to as the "Godfathers of AI" or the founding fathers of modern AI research, have were awarded together the prestigious Turing award in 2018 for their contribution to the AI revolution.



EFIGURE 11.1 Graphic chart depicting the three major subtypes of machine learning: supervised learning, unsupervised learning, and reinforcement learning. In medicine, supervised learning has been most widely used, as it powerfully identifies subtle relationships in data, but at the cost of requiring large datasets for training. Further details in text.

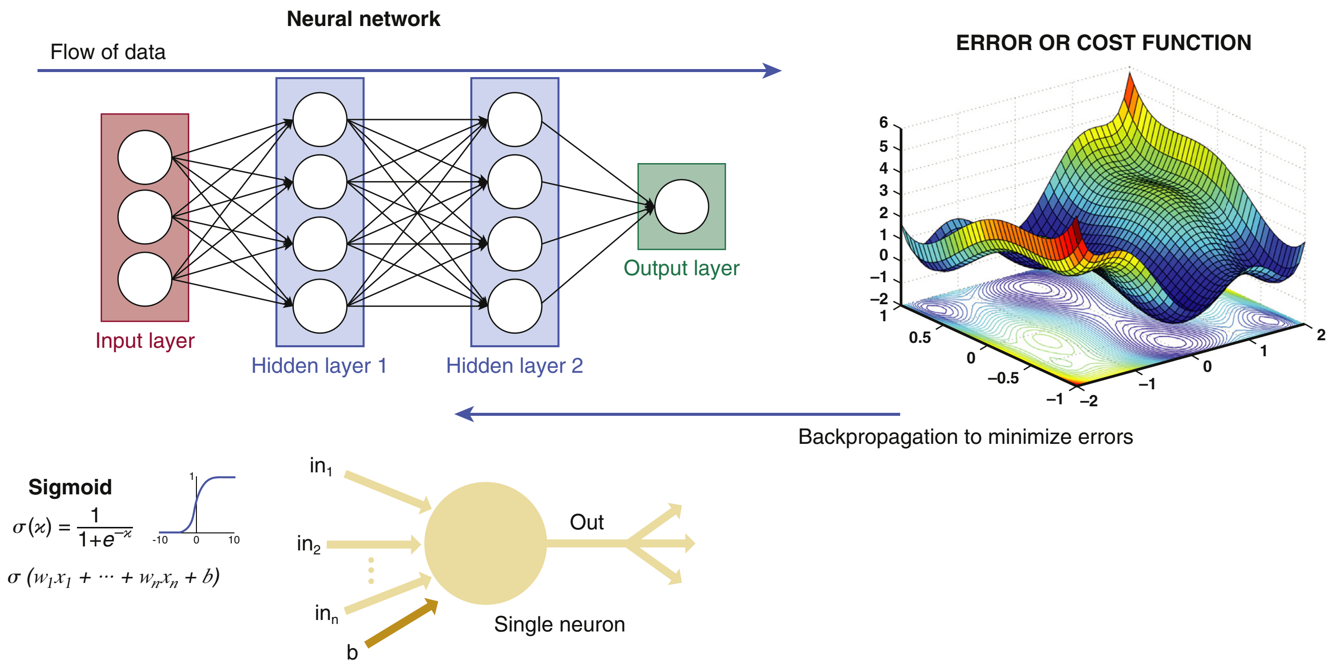


FIGURE 11.1 Graphical depiction of a neural network. **Top left:** The neural network shown contains four layers. Each layer is composed of "neurons" **(bottom panel)**. Each neuron receives multiple inputs, each multiplied by a weight (in1…inn), and a bias (offset) "b" and applies a nonlinear function to generate its output. During network training, weights and biases of each neuron are adjusted via backpropagation to minimize an error function (example error function shown **top right**).

F on the inputs yields estimated outputs (in the example, the probability of AF), and with each iteration, using the error between the estimated outputs and the real labels, model performance is assessed and the function weights are adjusted in a direction to minimize the error, improving model performance. The methods used to adjust to weights will be described in the "optimization and hyperparameters" subsection. The task can be either a classification—determination of the appropriate class for a data sample from a limited set of options (dogs versus cats, male versus female)—or a regression—a continuous value for each sample (age from an image). Because supervised learning in a neural network is an iterative process, with each step inching toward an improved solution, the biggest challenge is that large datasets are required. Because each sample in the dataset requires a label, attaching an accurate label to each element may be a limiting factor.

Unsupervised Learning

In unsupervised learning, the task revolves around the structure of the data itself. One most common form of unsupervised learning is *clustering*, in which the model clusters data based on its characteristics, instead of based on labels during the training stage. The model is fed only unlabeled data, and clusters samples based on similarity, using each sample's distance (Euclidian or other) from other samples. If the label of just a few samples in each cluster is known, the label of other samples in the cluster can be inferred because all the samples in that cluster would have similar features, but the model itself created the clusters without specific labels on the data elements. An example would be the acquisition of multiple ECG segments from a patient during a dialysis session at various potassium blood levels. The ECG segments could be clustered, and the potassium value of each cluster should be similar. Because unsupervised learning requires only the raw samples and basic assumptions regarding the data structure (such as the number of clusters), the barrier imposed by labeled data is lowered.

Reinforcement Learning

RL develops the optimal strategy for an agent in an environment with known rules and rewards. An example would be a chess player—learning chess by playing against itself, without labels or recorded human games. It uses only the rules and the game score. As RL requires known rules and rewards, its use in health care is still limited[2] and it is outside of the scope of this chapter.

Fully Connected and Convolutional Neural Networks

Inspired by the human brain, a fully connected neural network is a multi-layer parametric function that implements a nonlinear function between the inputs to the outputs (see Fig. 11.1). Each node (neuron) in each layer receives a weighted sum of all the nodes in the preceding layers and is activated using a nonlinear function. The values of the weights are defined during training, as the network learns the relationship between the input and output.

In convolutional neural networks, convolutional filters extract feature information from images in the convolutional layers, with the weights of the filters determined during training, so that the features selected are those that best define the desired network output. Both types of networks can be either used for classification tasks or for regression tasks.

Optimization and Hyperparameters

During training, the weights of the inputs to each neural in a neural network are adjusted so that the output of the function when fed samples with known labels will be the "closest" to their real labels. The difference between network output and the actual label is measured using the loss function, and a small loss value indicates network outputs close to the real labels. While weights could be randomly set to initiate a network and then randomly varied until the loss function is sufficiently small, this is clearly an inefficient and impractical way of training a network. A more efficient approach is to test a network on a set of samples, assess the effect of each of the network weights on the error function, and change the neuron weights accordingly. Mathematically, this is done by taking the derivative (actually, the gradient) of the function the network is implementing and changing the weights in the opposite direction (as the gradient points toward a higher loss). The gradient is approximated using a batch of samples, and the number of samples in the batch ("*batch size*") affects how accurate and smooth the gradient will be. While the gradient indicates the direction in which the weights should be adjusted, the magnitude of change (step size) is unknown. If a step size is too big, during training the function may step over a minimum of the error function (overcorrect), whereas if it is too small, the impact of the updated weights may be too diminutive to improve network accuracy, and the network may not converge or take a long time to train. Step size and the batch size are network "hyperparameters"—variables that affect how network parameters are changed (i.e., how the network learns during training), but which are not part of the final network function itself. Finding optimal hyperparameters requires empiric assessments of various combinations and is part of the art of network training. Assessing promising combinations of hyperparameters is typically performed on a second set of samples not used during network training called the internal validation set. Once hyperparameters and the network are finalized, the network is tested on a third set of samples not previously seen by the network referred to as the holdout testing set.

Transfer Learning

Transfer learning is a method used to apply supervised learning to problems for which the datasets available to train a network are relatively small. In this method, a network is developed to solve a problem that has enough labeled samples (the "primer"), and then it is retrained to solve a similar task with a much smaller dataset. The underlying hypothesis is that some of the patterns learned by the model are common to both tasks but can only be learned with a sufficient number of samples. This is similar to a human learning one musical instrument proficiently over years, and then requiring much less time and effort to learn a related instrument (e.g., guitar and banjo). Using transfer learning, datasets that may initially appear irrelevant can be used to solve specific tasks, and the transfer can be applied to all model parameters (basically seeding the model with weights from a trained model, instead of random weights), or to only a subset of parameters by "freezing" some model layers during training to keep the primer model values but allowing the rest to change.

CLINICAL USES IN CARDIOVASCULAR MEDICINE

ECG-Based Screening, Detection, and Prevention

Achieving human-like automated ECG interpretation has been a goal since the advent of digital ECG more than 60 years ago.[3] Early iterations of the technology were designed to identify fiducial points, make discrete measurements, and define common quantifiable abnormalities,[4–6] whereas contemporary approaches have moved beyond these rule-based approaches to recognize patterns in massive quantities of labeled ECG data.[7–9] Some early success has been achieved training deep neural networks (DNNs) on large datasets of single-lead ECGs and applying the algorithms to the 12-lead ECG,[8] sometimes outperforming expert over-readers.[7] In general, however, most algorithms lack the accuracy needed for widespread application without human oversight,[10] and it is likely that these technologies remain a tool to improve rather than replace human expertise for the foreseeable future.

For some discrete applications, these algorithms may enable rapid diagnosis on novel, patient- or consumer-facing devices. For example, algorithms have been demonstrated to be effective for AF diagnosis on a variety of single-lead ECG devices,[11–14] and there is great potential for making other important diagnoses including QT prolongation,[15] acute myocardial infarction,[16] or other arrhythmias (**Video 11.1**).[17] This "democratization" of ECG technology will exponentially increase the volume of signals that demand interpretation, and this will quickly outstrip the capacity of human ECG readers. We anticipate that these models will be essential in facilitating telehealth technologies—automatic patient- and consumer-facing technologies.

By leveraging massive labeled datasets, various neural networks can be used to move beyond human-like tasks to uncover more subtle patterns in the ECG that have gone unrecognized by even expert ECG readers. In doing so, these networks can bring new diagnostic power and value to the ECG. For example, the ECG can identify low ejection fraction (EF),[18] propensity toward AF (observable during normal sinus rhythm),[19] hypertrophic cardiomyopathy,[20] left ventricular hypertrophy,[21] hyperkalemia,[22] age and sex,[23] medical comorbidity/frailty, and studies are ongoing to identity markers of valvular heart disease, amyloidosis, and many other characteristics.

The AI ECG is an example of adding AI to an existing clinical test (the 12-lead ECG), which is already embedded in clinical workflows and widely available. In that context it can readily screen for underdetected disease, for which therapies exist (Fig. 11.2A). The use of the 12-lead ECG to identify left ventricular dysfunction (present asymptomatically in 3% to 9% of people) is undergoing prospective evaluation in a large cluster-design pragmatic trial (EAGLE, discussed later); the AI ECG has also been embedded into a stethoscope form factor (i.e., a stethoscope with embedded electrodes to record the ECG during a normal examination, permitting the application of AI). Due to the potential sensitivity of AI tests to detect disease early and provide deep phenotyping, it may appear to "predict" future disease, creating a class of "previvors" who have not yet experienced a disease (Fig. 11.2B). The AI ECG may also have a role to permit inexpensive at-home follow-up of patients at risk for ventricular dysfunction, such as those receiving chemotherapy or cardiac transplants (Fig. 11.2C). Prospective studies assessing such use cases are underway (TACTIC, NCT03879629). Whether AI tests achieve a sufficient level of predictive power to warrant intervention before a disease is manifest by currently used tests requires validation (Fig. 11.3), which at present is under development. To date, AI-based tools have received FDA approval for rhythm determination (i.e., to allow scalability of human capabilities). AI to extract information beyond what humans can determine (such as ventricular dysfunction) is currently under regulatory review.

Image Interpretation and Procedural Guidance

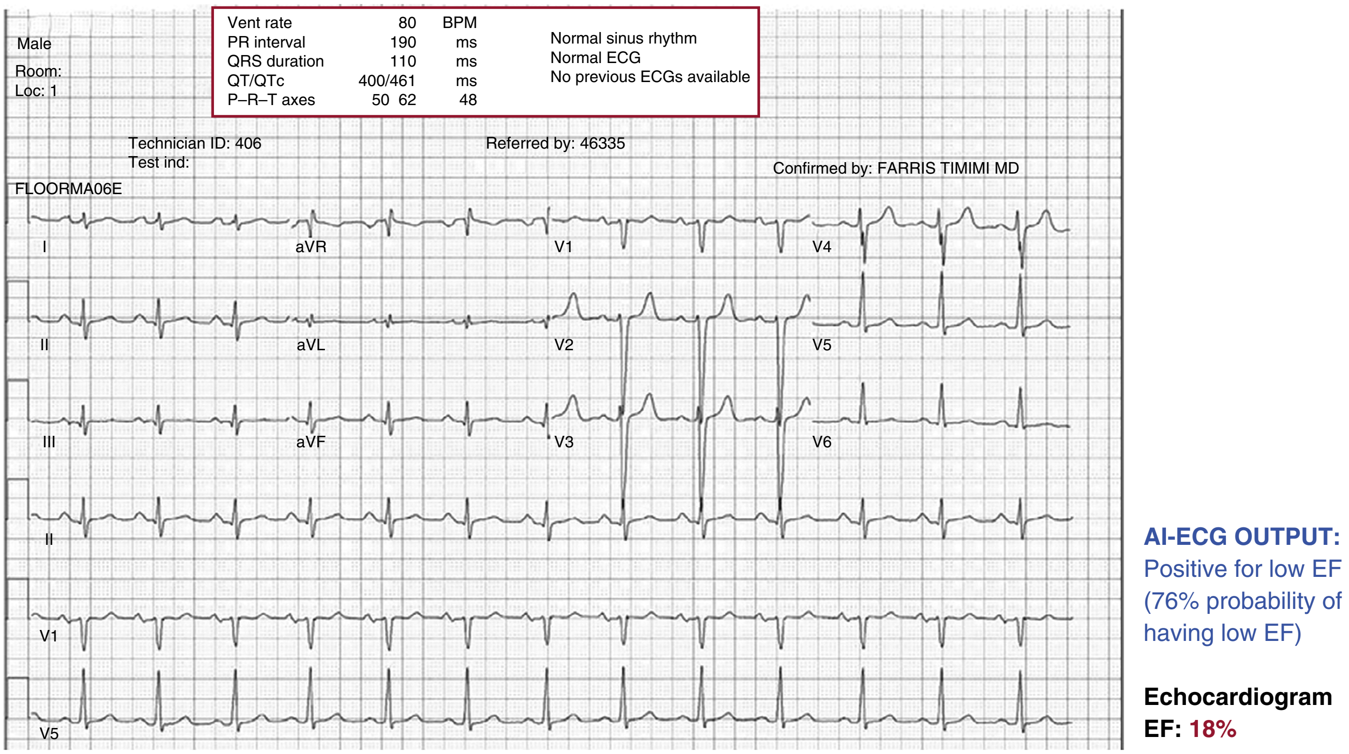
Cardiac imaging has been a particular focus of modern AI and ML work. ML has been used to improve image acquisition, image quality, accuracy of interpretation, and enhancing insights into cardiac physiology.

Nuclear Cardiology and Stress Testing

Stress testing by ECG or by nuclear imaging can yield false-positive and false-negative results. The prognostic value of either modality depends in part effected by patient-specific pre-test likelihood of disease among other electrocardiographic, clinical, perfusion, and functional variations. The number of variables that can impact accuracy of interpretation of functional evaluations for coronary disease may lead to a wide variation in predictive accuracy of humans. One study of over 2000 patients suggested that an ML algorithm that integrated all available patient data (including clinical and electrocardiographic) along with imaging data performed better in predictive major adverse cardiac events than ML focused on imaging alone or physician interpretation. Furthermore, deep learning approaches can offer statistically significant improvements in identification of per-vessel and per-patient sensitivity for detection of obstructive coronary disease.[1,24]

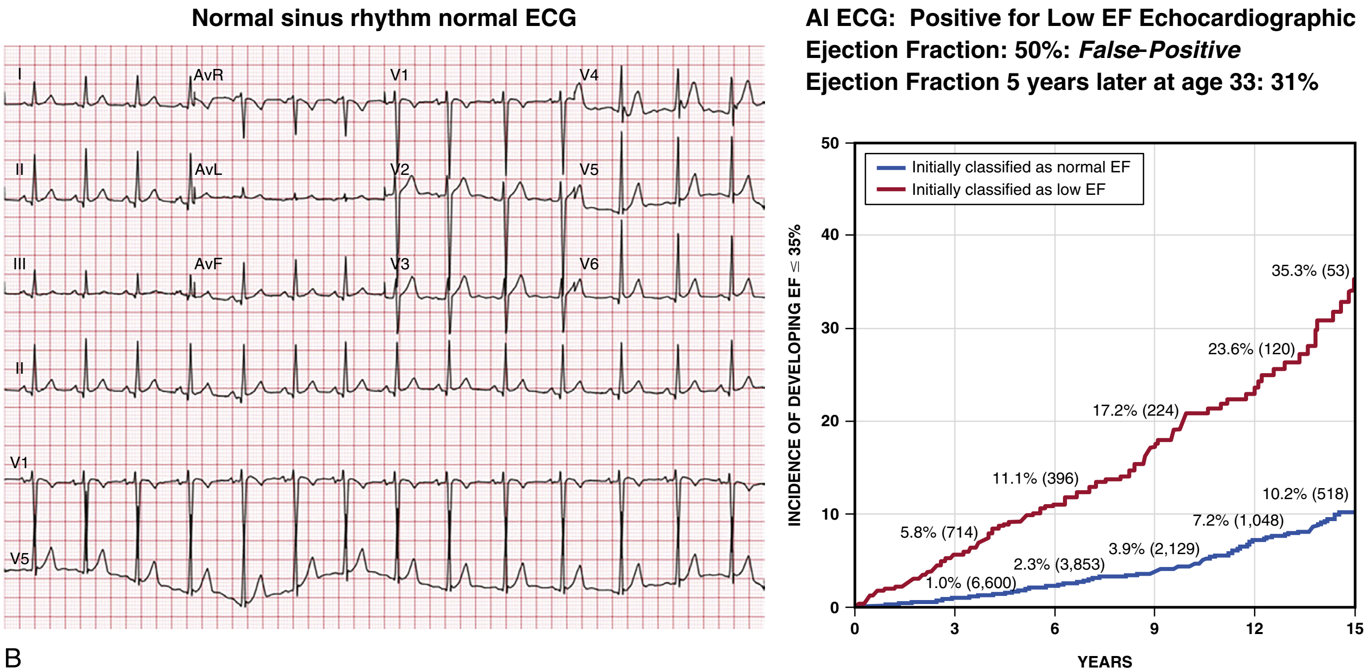
Echocardiography

The clinical utility of echocardiographic images depends on several factors: (1) skill related to image acquisition; (2) image quality; and (3) accuracy and consistency of image interpretation.[25] Newer handheld devices, some of which are smartphone enabled, are inexpensive and have made bedside echocardiographic imaging available to many individuals with limited training in image acquisition and/or interpretation. Recent work has focused on using AI approaches to facilitate remote training of unskilled sonographers as well as for robot-assisted echocardiography. The latter significantly improves diagnostic process time when done in combination with telemedicine-enabled cardiac consultation. Such approaches may help scale not just the availability of tools to acquire echocardiographic images but also the capacity to acquire high-quality, interpretable images with minimal prior experience. Because embedded AI in the imaging tool recognizes images of diagnostic quality, immediate feedback is given to the bedside imager (for example, indicating an image is acceptable, or if not, suggesting specific maneuvers to acquire the desired image).



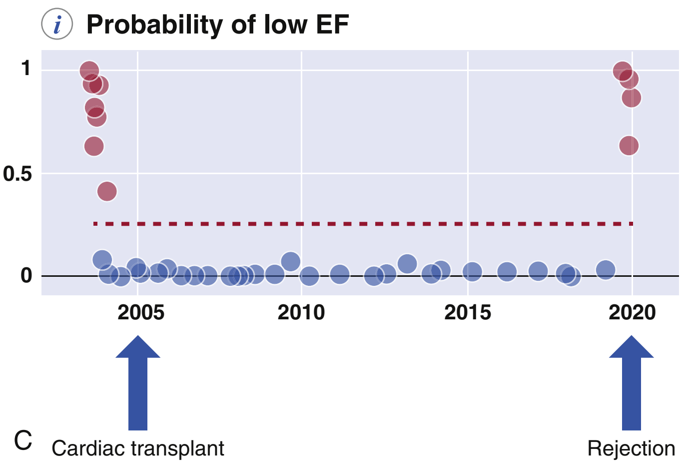
Male
Room:
Loc: 1
Vent rate 80 BPM
PR interval 190 ms
QRS duration 110 ms
QT/QTc 400/461 ms
P–R–T axes 50 62 48
Normal sinus rhythm
Normal ECG
No previous ECGs available
Technician ID: 406
Test ind:
Referred by: 46335
Confirmed by: FARRIS TIMIMI MD
FLOORMA06E
I
aVR
V1
V4
II
aVL
V2
V5
III
aVF
V3
V6
II
V1
V5
AI-ECG OUTPUT:
Positive for low EF (76% probability of having low EF)
Echocardiogram EF: 18%

A



Normal sinus rhythm normal ECG
I
AvR
V1
V4
II
AvL
V2
V5
III
AvF
V3
V6
II
V1
V5
AI ECG: Positive for Low EF Echocardiographic
Ejection Fraction: 50%: False-Positive
Ejection Fraction 5 years later at age 33: 31%
Initially classified as normal EF
Initially classified as low EF
INCIDENCE OF DEVELOPING EF ≤ 35%
YEARS
35.3% (53)
23.6% (120)
17.2% (224)
11.1% (396)
5.8% (714)
10.2% (518)
7.2% (1,048)
3.9% (2,129)
2.3% (3,853)
1.0% (6,600)

B



Probability of low EF
1
0.5
0
2005
2010
2015
2020
Cardiac transplant
Rejection

C

FIGURE 11.2 **A,** Electrocardiogram acquired from a 35-year-old asymptomatic man who presented after his sister died suddenly, read as normal. An AI-ECG algorithm reported a 76% probability of having a low ejection fraction. Subsequent echocardiography demonstrated an ejection fraction of 18%. He was ultimately diagnosed with familial dilated cardiomyopathy. **B,** *Left,* Electrocardiogram from a 28-year-old man, read as normal. The AI-ECG algorithm indicated a high probability of an ejection fraction less than 35% (positive test). Echocardiography at that time reported an ejection fraction (EF) of 50% suggesting a false-positive. However, the patient developed ventricular dysfunction, with an EF of 31% 5 years later. In some patients, the AI algorithm may identify subtle features that may predict future development of low EF. This situation illustrates the concept of disease "previvors," and in this case may result from pathophysiologic changes impacting ion channels and electrical impulse generation before mechanical function is affected, although the mechanism remains unproven. *Right,* The increased risk of developing left ventricular dysfunction with a positive AI-ECG screen for ventricular dysfunction. **C,** Plot of the AI-ECG outputs for all of the ECGs for an individual patient, taken from the Mayo Clinic Cardiology AI dashboard. Each point on the graph is generated by a single ECG, with the abscissa indicative of the date of the ECG, and the ordinate the probability of ventricular dysfunction. The patient had a dilated cardiomyopathy, and confirmed low ejection fraction (EF; *red points*, left of 2005 on the graph). He received a cardiac transplant, with normalization of his ejection fraction, and subsequent low probability of low EF by AI-ECG *(blue points)*. In 2020 he suffered rejection and ventricular dysfunction, identified by the AI-ECG (*red points* at 2020). (**B** from Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med*. 2019;25:70–74.)

In addition to facilitating high-quality image acquisition, the scalability of echocardiographic imaging is subject to the same limitations as other imaging modalities—namely the need for expert interpretation. Several recent studies have suggested that applying ML to echocardiographic images can accurately assess several key variables (ejection fraction [heart pump strength], heart chamber size, and valve function). Such approaches, when applied broadly, may allow for rapid and accurate identification of disease, and identification of those images that warrant expert over-read.

Finally, AI approaches have the potential to address other aspects of ultrasound image acquisition, including image noise and poor image quality. In various areas of ultrasound imaging, use of ML can improve image classification, detection, and segmentation. Integrated AI algorithms may streamline methodologic tasks ranging from optimizing imaging quality through the segmentation and registration of such images.[26]

Computed Tomography and Magnetic Resonance Imaging

Similar to echocardiography and nuclear stress imaging, a key limitation of computed tomography (CT) and magnetic resonance imaging (MRI) is expert clinician interpretation. AI approaches applied to cardiac imaging may improve the consistency and accuracy of interpretation of images. Beyond interpretation, AI approaches to image acquisition may reduce the time required to acquire high-quality MRI and CT images while limiting motion or other artifact. Finally, AI can scale and improve the speed of segmentation and reconstruction of data from MRI and CT. Recent reviews have summarized this evolving area.[27]

Coronary Arteriography

A final area of modern application of AI and ML to cardiac care is in the catheterization laboratory. The accuracy of pre-procedural evaluation (reducing false-positives and false-negatives from stress testing, accurate identification of lesion severity from noninvasive CT evaluation), consistency of interpretation of images acquired in real time, to enhancement of interpretation of complex lesions, and the appropriate choices in management may be facilitated by ML and deep learning algorithms. ML approaches may extract fractional flow reserve and lesion severity from CT coronary angiography.[28] This application may improve identification of those patients appropriate for coronary intervention. In addition, AI may offer value in interventional cardiology by predicting stent size and length, likelihood of future stenosis, and complex lesion characteristics (irregular lumen shape, invasive fractional flow reserve from cinegraphic images) which would traditionally require the use of additional tools (pressure wires, intravascular ultrasound).[29]

Natural Language Processing and Structured Data Analysis

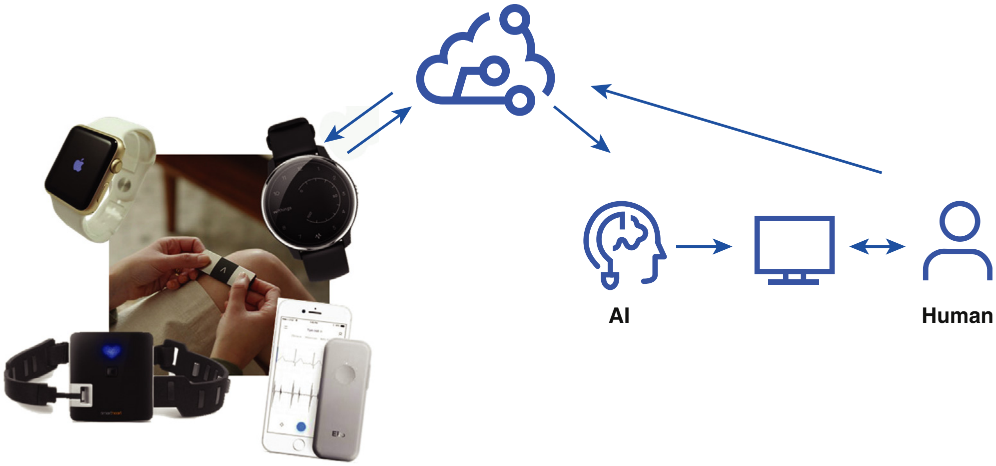
Structured data elements within the electronic health record are readily available for predictive analytics and can facilitate rapid, point-of-care decision support. However, much of the electronic health record (EHR) contains free text that requires additional processing for data abstraction. Traditional rule-based approaches to extract clinical data from free text are prone to misclassification due to the complexities of natural language structure, and more sophisticated models are emerging as more reliable alternatives. These approaches may include hybrid training of models using text vectorization and output tags that are fed into ML models or even completely unsupervised topic modeling. Because natural language processing (NLP) is not restricted by predefined diagnostic codes, an optimized NLP model has the ability to recognize even complex language patterns by comprehensively assessing all available documentation, thus improving the accuracy in capturing potentially ambiguous diagnoses.

Risk Scores (Deep Phenotyping)

The vast repositories of structured and unstructured patient data now available within EHRs offer an opportunity to generate risk scores and characterize patient phenotypes on a large scale. Such technologies promise to streamline patient care, identify individuals at risk of adverse outcome, and recognize and reinforce best practices. Often, these deep phenotyping approaches use a hybrid deep-learning model structure to distill the complicated relationships hidden in the data. This may include models that transform event structures into deep clinical-concept embedding and use a recurrent neural network (RNN) to predict outcomes over time. For example, one large-scale retrospective study using over 3 million patient records demonstrated that both traditional statistical approaches and novel ML models can predict risk of AF.[30] Similar approaches have been used to identify patients with heart failure,[31] to predict risk of hospitalization,[32] diagnose diabetes and peripheral artery disease, and generally have superior performance than relying on structured data alone.[26,33,34] Public repositories of these algorithms, such as the Phenotype KnowledgeBase (PheKB),[35] now contain algorithms for 50 to 60 medical conditions and many have demonstrated good performance when implemented across different health systems.[36–40]

IMPLEMENTING ARTIFICIAL INTELLIGENCE INTO CLINICAL PRACTICE

AI stands to increase the power of existing tests and transform many mundane accessories (e.g., stethoscopes, shirts, watches) into sources of medically diagnostic information (**eFig. 11.2**). Several key issues need to be addressed as computational algorithms are applied to clinical cardiology practice. Standards will likely need to be set, for example, on optimal approaches to testing and validation of these algorithms. Questions such as diversity of the training and testing sets and how to ensure that an algorithm will function similarly on data acquired at centers beyond the center(s) where the initial algorithm was developed are still evolving questions. Furthermore, studies are needed to understand how best to optimize real-time, clinical implementation of AI-enabled alerts. For example, AI algorithms that streamline data acquisition, interpretation, and reporting may be more easily integrated into practice. However, as such algorithms become more integrated into systems, the potential for human oversight and correction may decrease. In turn, AI-managed alerts that allow for advanced recognition of disease (e.g., EF from a 12-lead ECG) may not have significant impact if not implemented in such a way that physicians appropriately react to the alert. Large prospective studies to assess the impact on workflow and real-world impact are needed and are



EFIGURE 11.2 Diagram depicting the AI ECG as a platform from remote monitoring. *Far left,* Multiple form factors have been used to collect ECG and other physiologic signals. These are transmitted typically via cellular technology to a cloud for AI processing, designed to generate direct patient or technician alerts in the event of actionable abnormalities. *AI*, Artificial intelligence.

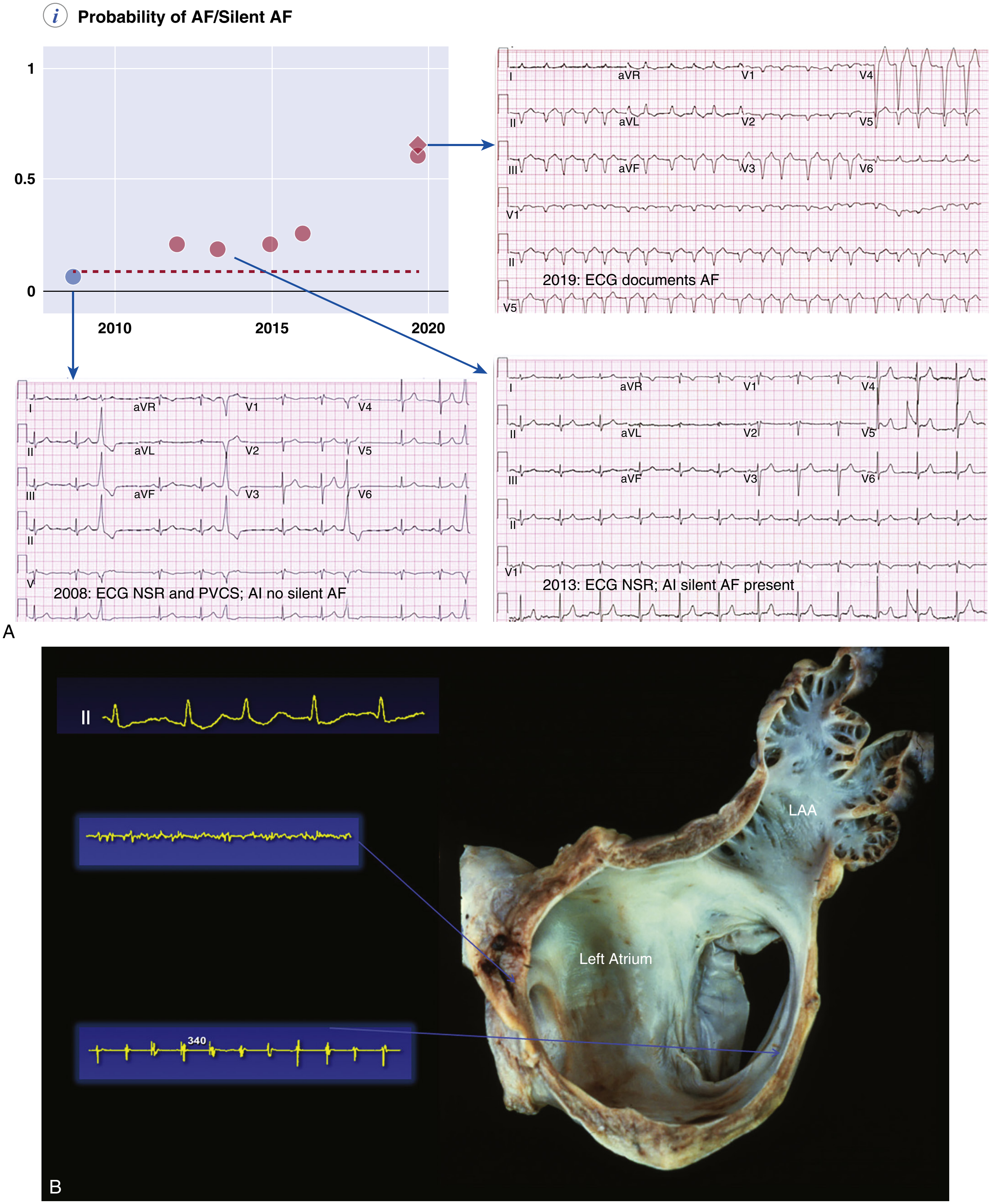


FIGURE 11.3 **A,** Shown is the AI-ECG dashboard for the probability of silent atrial fibrillation (AF; *top left*). In 2010, the probability of silent AF was low (representative ECG *bottom left*). From 2011 on, all ECG were recorded during sinus rhythm, but the AI ECG raised suspicion of episodic AF (*bottom right*, example from 2013), until 2019, when a tracing demonstrating AF was obtained. **B**, Surface ECG lead II *(top left)*, intracardiac electrogram (EGM) acquired from interatrial septum (*left*, middle recording), and EGM acquired from the coronary sinus (*left*, bottom tracing). Note the continuous fractionated signals recorded at the septum, suggestive of fibrosis, as opposed to the discrete signals, separated by isoelectric intervals recorded simultaneously at the coronary sinus during atrial fibrillation. The structural changes and/or transient repolarization changes before or after atrial fibrillation (AF) episodes may lead to subtle changes on the ECG that are used by the AI ECG to determine episodic atrial fibrillation is present, although the mechanisms by which AI determines the presence of AF from an ECG recorded during NSR is not known. *AI*, Artificial intelligence.

underway.[41] Mayo Clinic has developed an AI dashboard accessible via the EMR that automatically ingests all ECGs available in the EMR and displays multiple AI analyses in an interactive, graphical format (**Video 11.2**, eFig. 11.3).

Other key areas of consideration when implementing AI and ML into clinical practice include how to allow for continuous adaptation of the algorithms to new data, associated regulatory implications, and how algorithms and data should interact. Continuous exposure to new data, which improves diversity of cases to which a clinician gets exposed thus improves clinical expertise. Ideally, systems that use AI algorithms will continue to evolve in response to their correct and incorrect interpretations, strengthening the overall model over time. However, the minimal standards to support regulatory approval of a given algorithm still require review. Furthermore, regulatory standards and approvals may vary between countries.[42] Finally, the question of how algorithms are practically deployed across institutions or countries remains to be determined. Many algorithms require such computational power as to only be operational in cloud-based systems. However, many data frameworks impose restrictions on sending data to centralized, cloud-based servers. Whether algorithms should be individually supplied to individual institutions (limiting the opportunity for continuous learning and leading to potential scalability issues due to the availability of adequate computational power at every local site) or enabled through cloud-based systems that allow data to be sent to a centralized framework (creating concerns regarding data sharing and data privacy) is an area of active discussion and will likely require policy discussions at a national and international level.

Pitfalls and Limitations of Artificial Intelligence in Cardiovascular Medicine

Despite the immense promise of ML, a number of considerations have impeded its development and require addressing. Before neural networks can be trained, data must be accessed in a usable format, and for many forms of ML, clearly labeled. This requires subject matter experts as well as technical experts. Data ownership remains an unresolved issue, particularly with patient data. Use of an individual's data in ML exposes them to risk of loss of privacy, and at the same time a third party may yield financial benefit, raising potential conflicts of interest. The training of many networks requires large quantities of data, often necessitating accumulation of data from more than a single institution, again with concerns relating to privacy and data ownership. There is currently a lack of well-established quality standards or a centralized clearinghouse for vetted technologies.

Deep learning has the capability to make deep connections within data but can only learn from data to which it was exposed. Any preexisting biases that lead to exclusion from the training set may lead to unreliable results when fed into a clinically used network. Examples have included a higher rate of misidentification of black versus white populations in facial recognition software. In medicine, false associations could lead to a prediction of increased mortality due to place of residence, socioeconomic status, and other nonmedical correlates.

Neural networks have been subject to adversarial attacks in which pixels are modified of an image with no visible effects to a human observer, yet with complete change in classification, the network output (Fig. 11.4). Such attacks could lead to misclassification and misdiagnosis and raise questions about the lack of understanding of the mechanism of network classification. This leads to the "black box issue" in that the components of a signal used by a network to make its determination are not known to humans, raising concerns about their broad spread deployment. Careful clinical testing and vetting can mitigate this concern.

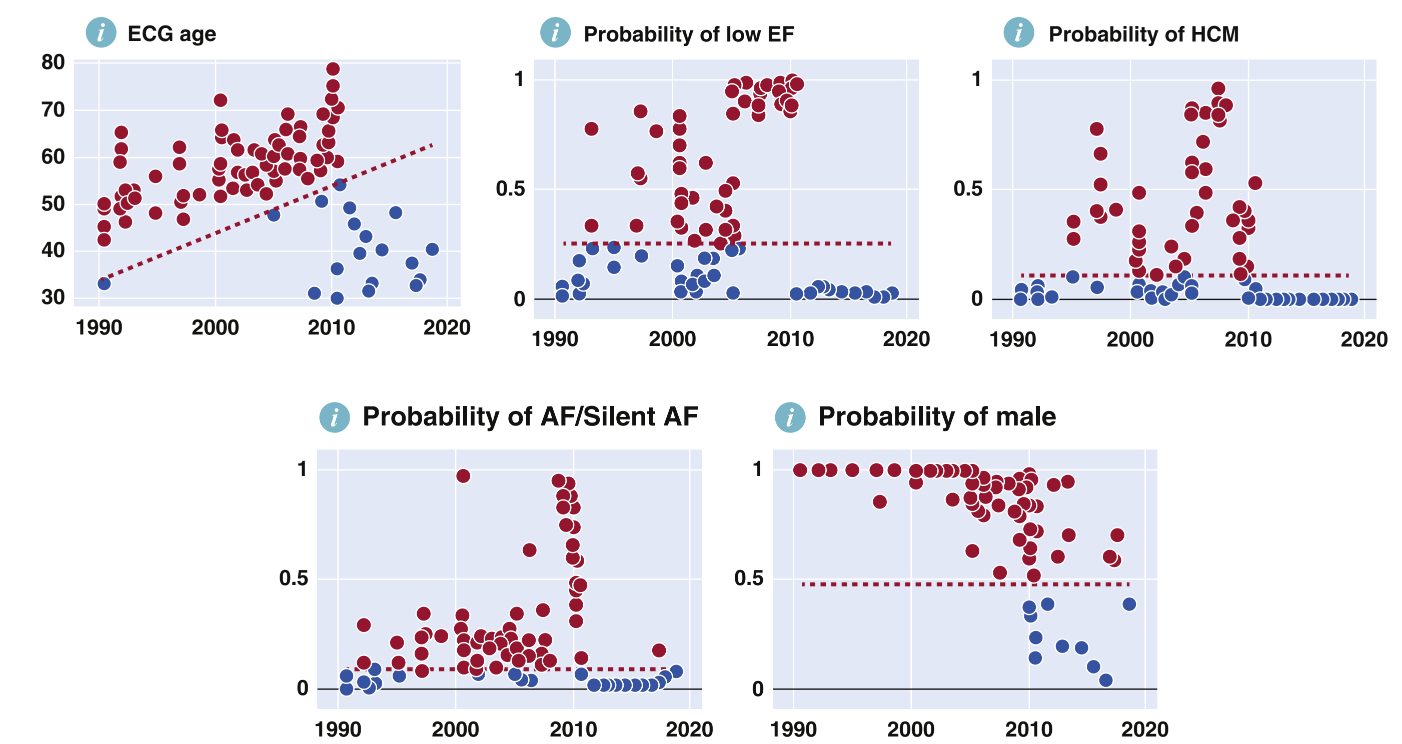
Lastly, physician engagement and thoughtful assessment of workflow and implementation are essential for the adoption of AI tools in clinical practice. Technology-driven solutions (such as many EHRs) have paradoxically led to physician burnout and patient dissatisfaction and have failed to fulfill their promise. Careful attention to user interfaces, patient and physician use requirements, meticulous validation, and outcomes-based observations will be essential to permit AI to improve clinical practice.

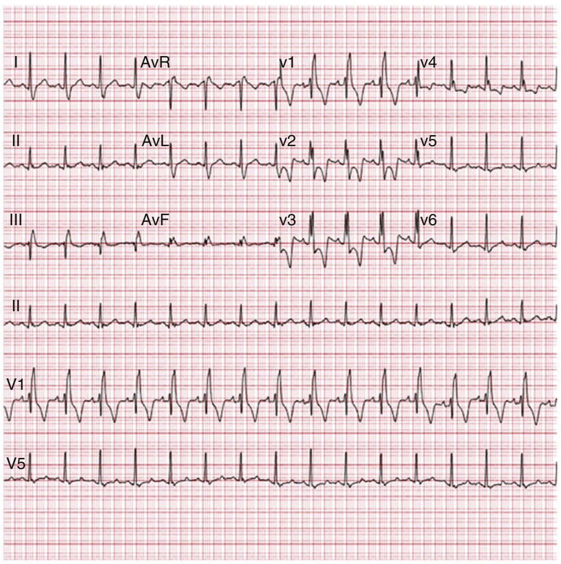
CONCLUSIONS

In summary, the application of ML to physiologic data stands poised to transform the practice of medicine. Many AI algorithms will be integrated into devices used by clinicians (including the electronic medical record); others may be stand-alone tools. While AI is unlikely to replace physicians, physicians who use ML tools will likely supplant those who do not. Much like the electrocardiogram at the turn of the century or the echocardiogram several decades ago, ML offers new ways to probe an individual's current state and to gauge more accurately its future state, and thus work to improve the human condition. But as with any medical tool, it requires proper testing, validation, and prospective assessment, as well as a compassionate and caring clinician to deploy, apply, and interpret its findings to help the human seeking care.

REFERENCES

1. Lopez-Jimenez F, Attia Z, Arruda-Olson AM, et al. Artificial intelligence in cardiology: present and future. *Mayo Clin Proc*. 2020;95:1015–1039.
2. Shen C, Gonzalez Y, Klages P, et al. Intelligent inverse treatment planning via deep reinforcement learning, a proof-of-principle study in high dose-rate brachytherapy for cervical cancer. *Phys Med Biol*. 2019;64:115013.
3. Pipberger HV, Freis ED, Taback L, Mason HL. Preparation of electrocardiographic data for analysis by digital electronic computer. *Circulation*. 1960;21:413–418.
4. Caceres CA, Rikli AE. The digital computer as an aid in the diagnosis of cardiovascular disease. *Trans N Y Acad Sci*. 1961;23:240–245.
5. Caceres CA, Steinberg CA, Abraham S, et al. Computer extraction of electrocardiographic parameters. *Circulation*. 1962;25:356–362.
6. Rikli AE, Tolles WE, Steinberg CA, et al. Computer analysis of electrocardiographic measurements. *Circulation*. 1961;24:643–649.
7. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med*. 2019;25:65–69.
8. Ribeiro AH, Ribeiro MH, Paixao GMM, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun*. 2020;11:1760.
9. Smith SW, Walsh B, Grauer K, et al. A deep neural network learning algorithm outperforms a conventional algorithm for emergency department electrocardiogram interpretation. *J Electrocardiol*. 2019;52:88–95.
10. Schlapfer J, Wellens HJ. Computer-interpreted electrocardiograms: benefits and limitations. *J Am Coll Cardiol*. 2017;70:1183–1192.
11. Halcox JPJ, Wareham K, Cardew A, et al. Assessment of remote heart rhythm sampling using the AliveCor heart monitor to screen for atrial fibrillation: the REHEARSE-AF study. *Circulation*. 2017;136:1784–1794.
12. Rajakariar K, Koshy AN, Sajeev JK, et al. Accuracy of a smartwatch based single-lead electrocardiogram device in detection of atrial fibrillation. *Heart*. 2020;106:665–670.
13. Wasserlauf J, You C, Patel R, et al. Smartwatch performance for the detection and quantification of atrial fibrillation. *Circ Arrhythm Electrophysiol*. 2019;12:e006834.
14. Perez MV, Mahaffey KW, Hedlin H, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. *N Engl J Med*. 2019;381:1909–1917.
15. Chung EH, Guise KD. QTC intervals can be assessed with the AliveCor heart monitor in patients on dofetilide for atrial fibrillation. *J Electrocardiol*. 2015;48:8–9.
16. Samol A, Bischof K, Luani B, et al. Single-Lead ECG recordings including Einthoven and Wilson Leads by a smartwatch: a new era of patient directed early ECG differential diagnosis of cardiac diseases? *Sensors*. 2019;19.
17. Ferdman DJ, Liberman L, Silver ES. A smartphone application to diagnose the mechanism of pediatric supraventricular tachycardia. *Pediatr Cardiol*. 2015;36:1452–1457.
18. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med*. 2019;25:70–74.
19. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet*. 2019;394:861–867.
20. Ko WY, Siontis KC, Attia ZI, et al. Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram. *J Am Coll Cardiol*. 2020;75:722–733.
21. Kwon JM, Jeon KH, Kim HM, et al. Comparing the performance of artificial intelligence and conventional diagnosis criteria for detecting left ventricular hypertrophy using electrocardiography. *Europace*. 2020;22:412–419.
22. Galloway CD, Valys AV, Shreibati JB, et al. Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram. *JAMA Cardiol*. 2019;4:428–436.
23. Attia ZI, Friedman PA, Noseworthy PA, et al. Age and sex estimation using artificial intelligence from standard 12-lead ECGs. *Circ Arrhythm Electrophysiol*. 2019;12:e007284.
24. Gomez J, Doukky R. Artificial intelligence in nuclear cardiology. *J Nucl Med*. 2019;60:1042–1043.
25. Seetharam K, Kagiyama N, Sengupta PP. Application of mobile health, telemedicine and artificial intelligence to echocardiography. *Echo Res Pract*. 2019;6:R41–R52.
26. Liu H, Bielinski SJ, Sohn S, et al. An information extraction framework for cohort identification using electronic health records. *AMIA Jt Summits Transl Sci Proc*. 2013;2013:149–153.
27. Dey D, Slomka PJ, Leeson P, et al. Artificial intelligence in cardiovascular Imaging: JACC State-of-the-Art Review. *J Am Coll Cardiol*. 2019;73:1317–1335.
28. von Knebel Doeberitz PL, De Cecco CN, Schoepf UJ, et al. Coronary CT angiography-derived plaque quantification with artificial intelligence CT fractional flow reserve for the identification of lesion-specific ischemia. *Eur Radiol*. 2019;29:2378–2387.
29. Sardar P, Abbott JD, Kundu A, et al. Impact of artificial intelligence on interventional cardiology: from decision-making aid to advanced interventional procedure assistance. *JACC Cardiovasc Interv*. 2019;12:1293–1303.
30. Hill NR, Ayoubkhani D, McEwan P, et al. Predicting atrial fibrillation in primary care using machine learning. *PloS One*. 2019;14:e0224582.
31. Choi DJ, Park JJ, Ali T, Lee S. Artificial intelligence for the diagnosis of heart failure. *NPJ Digit Med*. 2020;3:54.
32. Xiao C, Ma T, Dieng AB, et al. Readmission prediction via deep contextual embedding of clinical concepts. *PloS One*. 2018;13:e0195024.
33. Upadhyaya SG, Murphree Jr DH, Ngufor CG, et al. Automated diabetes case identification using electronic health record data at a tertiary care facility. *Mayo Clin Proc Innov Qual Outcomes*. 2017;1:100–110.





ECG date	Main rhythm	Heart rate	QT/QTc	Real age	ECG age	P of male (%)	P of low EF (%)	P of AF (%)	P of HCM (%)
2018-Aug-02 10:23:47		94	410/512	62.55	40.43	38.77	1.84	7.84	0.01
2017-jul-31 11:05:10		94	404/505	61.55	33.84	69.62	0.71	5.49	0.07
2017-Mar-15 06:30:52		88	408/493	61.17	32.77	58.78	0.91	16.96	0.01
2016-Dec-02 10:18:40		100	392/505	60.89	37.19	59.99	1.45	2.11	0.03
2016-Jul-11 11:14:22		94	368/460	60.49	52.14	4.37	2.59	0.68	0.01
2015-Jul-06 11:47:19		95	368/462	59.48	48.14	9.88	2.49	0.75	0.02
2014-jun-02 11:46:00		101	358/464	58.38	40.11	18.82	3.27	1.49	0.06
2013-Jun-		101	310/401	57.39	33.17	70.68	3.21	2.12	0.02

EFIGURE 11.3 Shown is an example of how the ECG dashboard may be implemented to electronic health records, thus facilitating easy access to enhanced reporting of novel insights beyond traditional ECG interpretation. This patient received a cardiac transplantation in 2010—note how the probability of low ejection fraction (EF) drops into the normal range (*red dots* replaced with *blue dots* after 2010). Marked abnormalities in one AI-ECG screen may lead to perturbations in other AI-ECG tests.

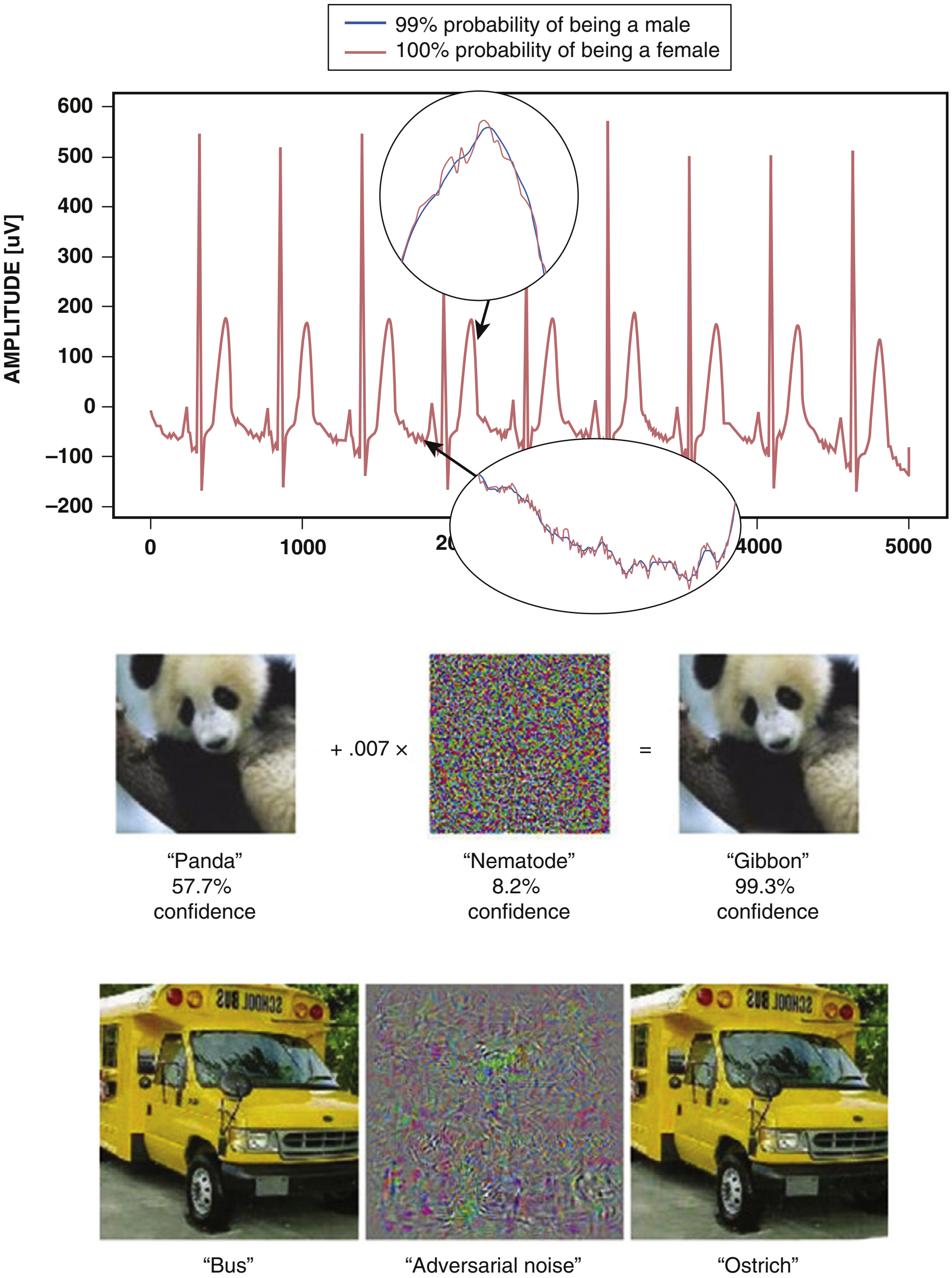


FIGURE 11.4 Examples of adversarial attacks on neural networks. An ECG *(blue tracing)* is correctly classified as being acquired from a man by a neural network. The addition of sub-clinical noise to the signal *(red tracing)* leads the network to misclassify the tracing as that belonging to a woman, despite the absence of significant change to a human observer. The images below the ECG depict similar adversarial network attacks against a network designed for image classification. The addition of apparent noise results in no visible change to a human observer, but misclassification of a panda as a gibbon by the network, as well as a similar disruption using a different image.

34. Afzal N, Sohn S, Abram S, et al. Mining peripheral arterial disease cases from narrative clinical notes using natural language processing. *J Vasc Surg*. 2017;65:1753–1761.
35. Kirby JC, Speltz P, Rasmussen LV, et al. PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability. *J Am Med Inform Assoc*. 2016;23:1046–1052.
36. Carroll RJ, Thompson WK, Eyler AE, et al. Portability of an algorithm to identify rheumatoid arthritis in electronic health records. *J Am Med Inform Assoc*. 2012;19:e162–e169.
37. Kho AN, Hayes MG, Rasmussen-Torvik L, et al. Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. *J Am Med Inform Assoc*. 2012;19:212–218.
38. Ritchie MD, Denny JC, Zuvich RL, et al. Genome- and phenome-wide analyses of cardiac conduction identifies markers of arrhythmia risk. *Circulation*. 2013;127:1377–1385.
39. Kawatkar A, Chu LH, Iyer R, et al. Development and validation of algorithms to identify acute diverticulitis. *Pharmacoepidemiol Drug Saf*. 2015;24:27–37.
40. Denny JC, Crawford DC, Ritchie MD, et al. Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: using electronic medical records for genome- and phenome-wide studies. *Am J Hum Genet*. 2011;89:529–542.
41. Yao X, McCoy RG, Friedman PA, et al. Clinical trial design data for electrocardiogram artificial intelligence-guided screening for low ejection fraction (EAGLE). *Data Brief*. 2020;28:104894.
42. He J, Baxter SL, Xu J, et al. The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. 2019;25:30–36.