

# UC San Diego

## UC San Diego Previously Published Works

### Title

Machine Learning and Deep Neural Networks in Thoracic and Cardiovascular Imaging

### Permalink

<https://escholarship.org/uc/item/6kb012qf>

### Journal

Journal of Thoracic Imaging, 34(3)

### ISSN

0883-5993

### Authors

Retson, Tara A  
Besser, Alexandra H  
Sall, Sean  
et al.

### Publication Date

2019-05-01

### DOI

10.1097/rti.0000000000000385

Peer reviewed

# Machine Learning and Deep Neural Networks in Thoracic and Cardiovascular Imaging

Tara A. Retson, MD, PhD,\* Alexandra H. Besser, MD, PhD,\* Sean Sall, BA,†  
Daniel Golden, PhD,† and Albert Hsiao, MD, PhD\*

**Abstract:** Advances in technology have always had the potential and opportunity to shape the practice of medicine, and in no medical specialty has technology been more rapidly embraced and adopted than radiology. Machine learning and deep neural networks promise to transform the practice of medicine, and, in particular, the practice of diagnostic radiology. These technologies are evolving at a rapid pace due to innovations in computational hardware and novel neural network architectures. Several cutting-edge post-processing analysis applications are actively being developed in the fields of thoracic and cardiovascular imaging, including applications for lesion detection and characterization, lung parenchymal characterization, coronary artery assessment, cardiac volumetry and function, and anatomic localization. Cardiothoracic and cardiovascular imaging lies at the technological forefront of radiology due to a confluence of technical advances. Enhanced equipment has enabled computed tomography and magnetic resonance imaging scanners that can safely capture images that freeze the motion of the heart to exquisitely delineate fine anatomic structures. Computing hardware developments have enabled an explosion in computational capabilities and in data storage. Progress in software and fluid mechanical models is enabling complex 3D and 4D reconstructions to not only visualize and assess the dynamic motion of the heart, but also quantify its blood flow and hemodynamics. And now, innovations in machine learning, particularly in the form of deep neural networks, are enabling us to leverage the increasingly massive data repositories that are prevalent in the field. Here, we discuss developments in machine learning techniques and deep neural networks to highlight their likely role in future radiologic practice, both in and outside of image interpretation and analysis. We discuss the concepts of validation, generalizability, and clinical utility, as they pertain to this and other new technologies, and we reflect upon the opportunities and challenges of bringing these into daily use.

**Key Words:** machine learning, artificial intelligence, thoracic, cardiovascular

(*J Thorac Imaging* 2019;00:000–000)

## CARDIOTHORACIC IMAGING AT THE FOREFRONT OF TECHNOLOGICAL ADVANCES

From its inception, radiology has been a discipline filled with innovators and pioneers, who have ambitiously sought to solve challenging clinical problems with new technology. From the earliest application of x-rays for diagnosis and intervention,

to the advent of computed tomography (CT) and magnetic resonance imaging (MRI), to 3D postprocessing and computer-aided detection (CAD), the specialty has often been the first to embrace innovations that allow efficient diagnosis of disease to improve the care of patients. Increasingly, advanced CT and MRI scanners are capable of capturing images that freeze heart motion and exquisitely delineate fine anatomical structures at a level of detail never before seen outside of the operating room. In addition, advances in software and computational fluid mechanical models are enabling complex 3D and 4D reconstructions to visualize and assess the dynamic motion of the heart and quantify blood flow and hemodynamics (Fig. 1). Because of its intrinsic dynamic and anatomic complexities, cardiothoracic imaging is poised at the forefront of transformative technologies, as machine learning and deep neural networks begin to revolutionize medicine and image analysis.

## MACHINE LEARNING AND THE ADVANCEMENTS OF DEEP NEURAL NETWORKS

Since the development of the earliest computers, there has always been an interest in utilizing these systems to automate tasks. Starting with simple numerical calculation, computer programs quickly progressed in complexity. Modern high-level languages such as Python abstract away computational intricacies and have made it simpler to develop algorithms to ingest data and subsequently use them to make predictions. Commonly used supervised (using input and output pairs) and unsupervised (using only inputs) machine learning methods were developed and quickly expanded to include linear and logistic regression, tree-based methods such as random forests and gradient boosted trees, kernel-based methods such as support vector machines, and clustering methods like k-means and E-M clustering<sup>1</sup> (Fig. 2). Traditional machine learning methods share a common limitation—they require features to be specified in advance. The need for computational features meant that highly specialized computer scientists were required to conceptualize relevant features and manually translate them into a code to be used as input for the machine learning algorithms, a process sometimes referred to as “feature engineering.” These earlier methods were, therefore, somewhat inaccessible, as they required a considerable amount of time and expertise to train and fine tune. In contrast, neural networks can automatically learn input features from the data, alleviating much of the time and expertise necessary for training and fine tuning.

The field of artificial neural networks developed in parallel to traditional machine learning methods, albeit with a rockier history. The first precursor to current neural networks was the perceptron, a “brain model” for supervised learning of binary classifiers, developed in 1958.<sup>2</sup> Despite the initial excitement this generated, the perceptron was not capable of

From the \*Department of Radiology, University of California San Diego; and †Arterys Inc., San Francisco, CA.

T.A.R. and A.H.B. contributed equally to this work.

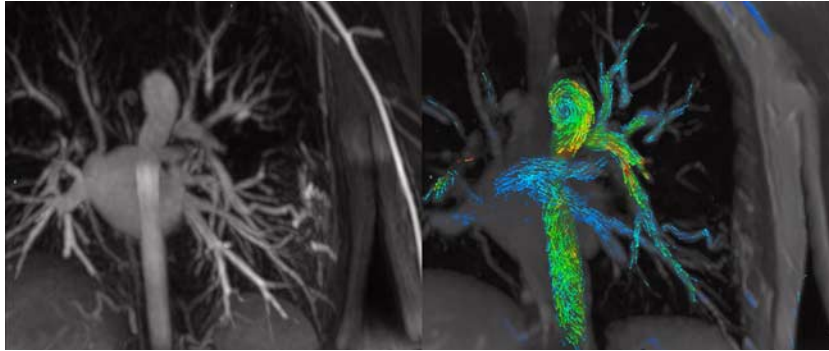
T.A.R.: receives support under NIH T32 EB005970 and RSNA RR1879.

A.H.B.: receives support under NIH T32 EB005970. S.S. and D.G.: are employed by Arterys Inc. A.H.: is a founder and shareholder of Arterys Inc., is a speaker for Bayer AG and receives research funding from GE Healthcare and Bayer AG.

Correspondence to: Albert Hsiao, MD, PhD, 9300 Campus Point Drive, MC 0841, La Jolla, CA 92037 (e-mail: hsiao@ucsd.edu).

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

DOI: 10.1097/RTI.0000000000000385



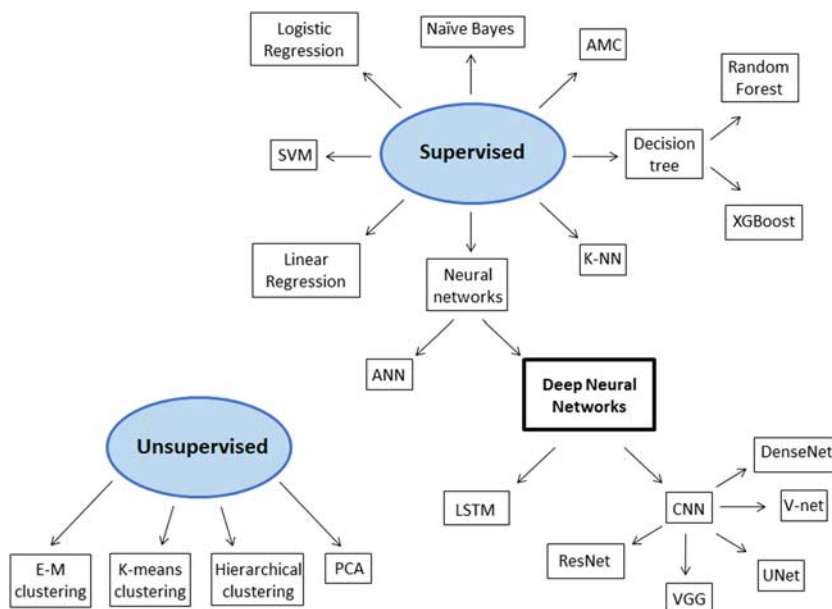
**FIGURE 1.** Static images from a cardiac-gated 4D MRA (left) and 4D flow (right), showing congenital branch pulmonary artery stenosis and enlarged intercostal collateral arteries supplying the left lower lobe. New technologies such as this provide an increasingly detailed understanding of pathophysiology, for greater patient-specific insights. full color online

solving some basic mathematical problems such as the exclusive-or (XOR), thus leading users to realize that it was not a universal solution. This led to several decades of an “AI winter” in which neural network development effectively came to a halt.<sup>3</sup> Neural network research remained relatively dormant until the work of Rumelhart et al<sup>4</sup> in 1986, which demonstrated that networks of neuron-like units could effectively be trained using gradient descent and back-propagation techniques. This discovery, along with a demonstrated ability for these networks to learn *any* function, revived the field of neural network research.<sup>5</sup>

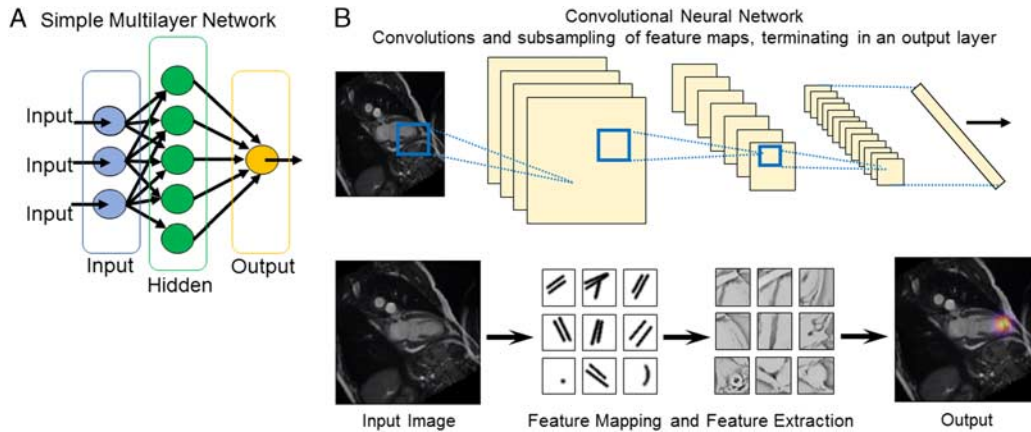
Through the 1990s and early 2000s, several breakthroughs were made. Neural networks were used to recognize handwritten digits, and pretraining with unspecified features enabled deep neural networks with multiple layers.<sup>6,7</sup> Progress was, however, intermittent and hindered by the large amount of data and computing resources needed to train these networks. A considerable breakthrough came in 2012 when Krizhevsky and colleagues trained a large, deep convolutional neural

network to classify objects in the ImageNet LSVRC contest and won by a significant margin.<sup>8,9</sup> The victory was attributed to a number of network architectural components, but also to the use of GPUs to accelerate network training with access to the large ImageNet data set (1.2 million images). Since then, there has been an explosion of neural network research (examples of simple architectures are shown in Fig. 3). Much attention has been directed to building larger network architectures up to hundreds of layers deep and novel ways to train them successfully.<sup>10–12</sup> In parallel, convolutional network architectures have also been created for tasks including object detection, localization, and segmentation.<sup>13,14</sup>

As neural network research has grown, high-level software frameworks like Keras were developed, allowing increased accessibility of image analysis and enabling its use in multiple other fields, including medicine.<sup>15</sup> Medical imaging has presented its own challenges, and network architectures originally built to operate on natural scene images needed to be optimized for medical image analysis.<sup>16</sup>



**FIGURE 2.** Traditional machine learning algorithms are categorized into “unsupervised” learning, requiring input data only, and “supervised” learning, requiring input and output data. Deep neural networks, a form of supervised learning, comprise the most recent advances in machine learning. AMC indicates associate memory classifier; ANN, artificial neural network; K-NN, K-nearest neighbor; SVM, support vector machines. full color online



**FIGURE 3.** Layout of an example (A) multilayer neural network and (B) deep convolutional neural network, for which many architectures are possible. Convolutional neural networks are capable of learning and extracting combinations of image features to render a result, such as the localization of the left ventricular apex (bottom right) without being explicitly programmed to look for specific features. [full color online](#)

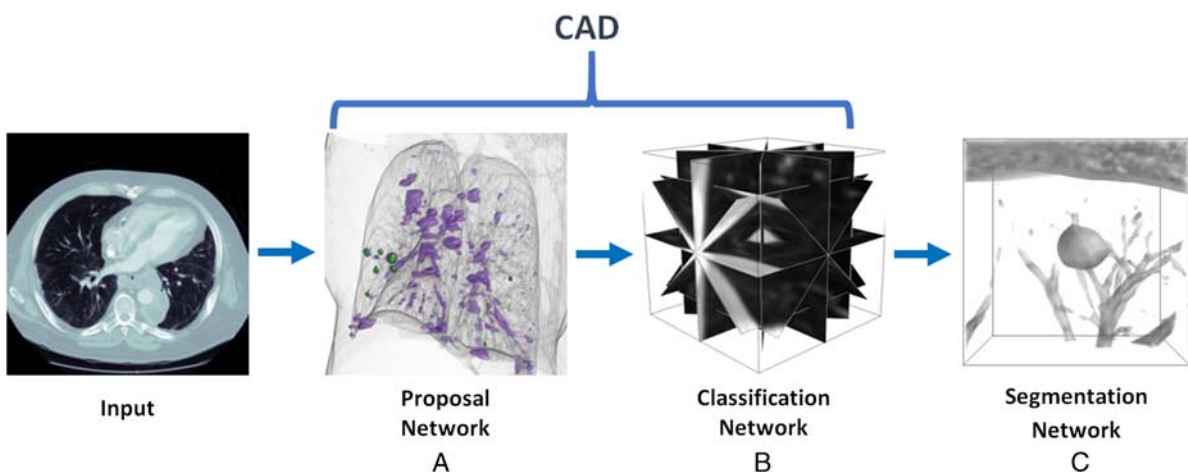
These obstacles have sparked inventive solutions, including novel architectures such as UNet, which was designed for semantic segmentation in the context of challenges that are commonly encountered in biomedical images (eg, small sample sizes and high image resolution).<sup>17</sup> Numerous networks have since been developed to solve a diverse range of problems in medicine, such as organ volumetry and lesion identification (Figs. 4, 5).

Realizing clinical usage of this technology in medicine presents a range of unique complexities that may be new to the machine learning community. Challenges that must be considered include integration into the health care system while maintaining patient security and privacy, wide-ranging normal and abnormal human morphology, and liability and accountability concerns. Artificial intelligence heralds a sea change in medical imaging, offering an opportunity for radiologists to be leaders in, and shape the field of computer-aided diagnostics. Early adopters of technologies have the opportunity to set precedents and

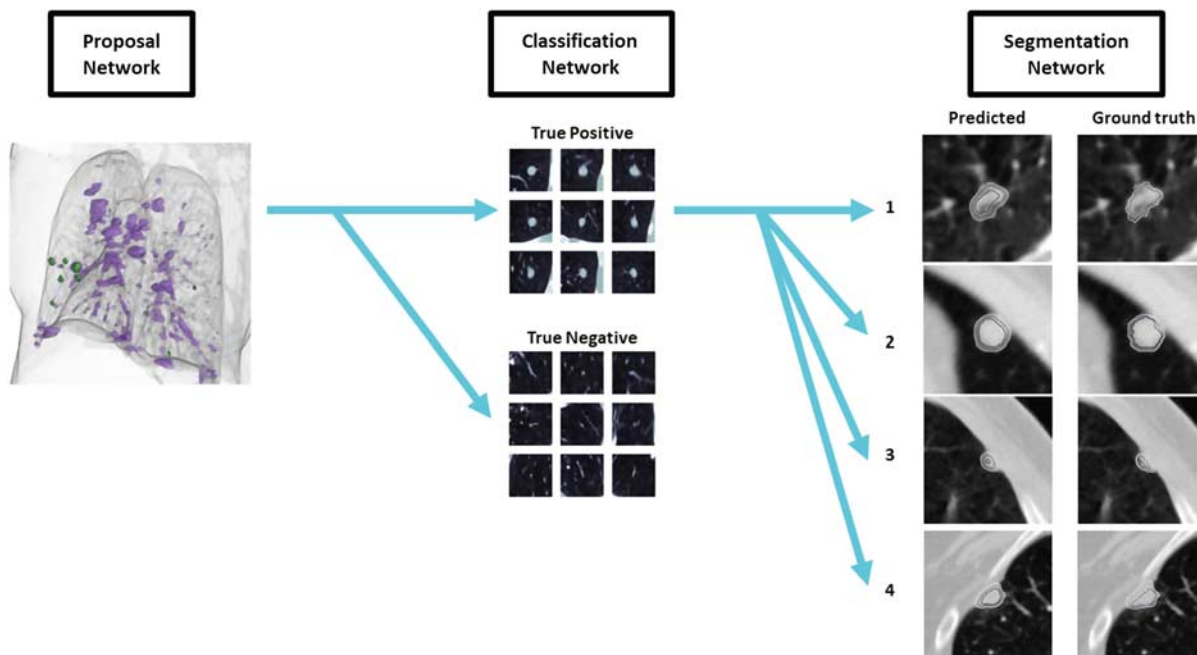
standards for their use and validation. As such, radiologists may play a critical role to shape the role of machine learning-based diagnostics in clinical practice.

### APPLICATIONS IN CARDIOVASCULAR AND THORACIC IMAGING

One of the earliest introductions of machine learning into cardiovascular medicine involved reading and interpreting ECG tracings, a technology that has now been incorporated into everyday clinical practice<sup>18</sup> and has benefited from the development of convolutional neural networks that allow detection and classification of arrhythmias.<sup>19–21</sup> Machine learning has since been applied to various tasks in cardiovascular and thoracic imaging including segmentation, characterization, quantification, lung nodule detection and measurement, and lung cancer prognosis and treatment. Many of these applications have been consolidated in prior reviews,<sup>22–24</sup> and we provide a survey of those relevant to



**FIGURE 4.** Architecture of a cascaded system of multiple neural networks, each building upon the outputs of a preceding network. In this example, the (a) proposal network identifies candidate pulmonary nodules, the (b) classification network identifies “true” pulmonary nodules from false positives, and the (c) segmentation network delineates the boundaries of each nodule. Together, the proposal and classification networks may be used as a method for CAD. [full color online](#)



**FIGURE 5.** Results of a cascaded convolutional neural network designed for pulmonary nodule detection and segmentation. The neural networks appear to overcome limitations of earlier algorithms in identifying boundaries of nodules near pleural surfaces. full color online

cardiovascular imaging in Table 1 and those relevant to thoracic imaging in Table 2.

**Applications to Cardiovascular Disease**

One major emphasis of machine learning applications in cardiac MRI is ventricular segmentation for quantification of volumetry and function. This is a particularly attractive problem for machine learning, as it is typically a time-consuming aspect of these examinations, often requiring manual outlining by a skilled operator. Deep learning algorithms have been generated that approximate the cardiac measurements of expert readers.<sup>25-27,48,49</sup> One such algorithm proposed by Avendi et al<sup>25</sup> was well correlated

with ground-truth measurements (0.99 for end systolic and 0.98 for end diastolic). Most recently, several commercial vendors have begun to take an interest in this technology and to integrate these algorithms into their software.<sup>50-52</sup> Outside of cardiac segmentation on MRI, algorithms have also been developed to detect subacute or chronic myocardial scar<sup>28</sup> and to predict patient survival and mechanisms of right heart failure in pulmonary hypertension.<sup>29</sup>

Machine learning techniques have also been applied to characterizing cardiac disease on echocardiography (ECHO). Ortiz et al<sup>30</sup> used neural networks to analyze cardiac contractility to predict 1-year mortality in patients with heart failure. Since this early work, supervised machine learning

**TABLE 1.** Applications of Machine Learning in Cardiovascular Imaging

Modality	References	Description	ML Technique
MRI	Avendi et al <sup>25</sup>	RV segmentation	CNN
	Winther et al <sup>26</sup>	RV, LV endocardium and epicardium	CNN
	Tan et al <sup>27</sup>	LV segmentation	ANN
	Baessler et al <sup>28</sup>	Myocardial scar detection	Random forests
	Dawes et al <sup>29</sup>	Pulmonary hypertension prognosis	PCA
ECHO	Ortiz et al <sup>30</sup>	HF prognosis	ANN
	Narula et al <sup>31</sup>	HCM vs athlete's heart	SVM, Random forests, ANN
	Sengupta et al <sup>32</sup>	Constrictive pericarditis vs restrictive cardiomyopathy	AMC, random forest, k-NN, SVM
	Sengur <sup>33</sup>	Valvular disease	SVM
	Moghaddasi and Nourian <sup>34</sup>	MR severity	SVM
CT	Vidya et al <sup>35</sup>	MI detection	SVM
	Wolterink et al <sup>36</sup>	CAC scoring	CNN
	Isgum et al <sup>37</sup>	CAC scoring	k-NN, SVM
	Itu et al <sup>38</sup>	FFR estimation	deep neural network
	Motwani et al <sup>39</sup>	Prognosis	Logistic regression
	Mannil et al <sup>40</sup>	MI detection	Decision tree, k-NN, random forest, ANN

AMC indicates associate memory classifier, ANN, artificial neural network; CAC, coronary artery calcium; CNN, convolutional neural network; FFR, fractional flow reserve; HCM, hypertrophic cardiomyopathy; HF, heart failure; k-NN, k-nearest neighbor; LV, left ventricle; ML, machine learning; MI, myocardial infarction; PCA, principle component analysis; RV, right ventricle; SVM, support vector machines.

**TABLE 2.** Applications of Machine Learning in Thoracic Imaging

Topic	References	Description	ML Technique
Lung Nodules	Lo et al <sup>41</sup>	Pulmonary nodule detection	CADE
Radiomics	Li et al <sup>42</sup> Song et al <sup>43</sup>	NSCLC prognosis NSCLC prognosis	Unsupervised 2-way clustering SVM
Radiogenomics	Yamamoto et al <sup>44</sup>	Genetic classification	Random forest
COPD	Ying et al <sup>45</sup> González et al <sup>46</sup>	COPD classification COPD staging	Deep neural network CNN
Abdominal aortic thrombus	López-Linares et al <sup>47</sup>	Thrombus detection	CNN

CADE indicates computer aided detection; CNN, convolutional neural network; COPD, chronic obstructive pulmonary disease; ML, machine learning; NSCLC, non–small cell lung cancer; SVM, support vector machines.

techniques have used ECHO to differentiate hypertrophic cardiomyopathy from athlete’s heart,<sup>31</sup> classify and differentiate constrictive pericarditis from restrictive cardiomyopathy,<sup>32</sup> diagnose valvular heart disease,<sup>33</sup> grade severity of mitral valve regurgitation,<sup>34</sup> automate ejection fraction measurement,<sup>53</sup> and detect the presence of myocardial infarction.<sup>35,54</sup>

Several machine learning applications have also been developed to assist in the interpretation of CT. For example, algorithms have been developed for the automation of coronary artery calcium scoring<sup>36,37,55,56</sup> and assessment of the functional significance of coronary lesions. More recently, ML techniques have been leveraged to better assess the hemodynamic significance of coronary stenosis. It has recently become possible to simulate the results obtained from an intracoronary pressure wire at the time of coronary catheterization with computational fluid simulations based on coronary CTA, the so-called fractional flow reserve.<sup>57–59</sup> However, the time and expertise needed for coronary segmentation and flow simulation may delay the availability of results and limit its use. With this concept in mind, several groups have attempted to use machine learning approaches to accomplish this similar task.<sup>38,60,61</sup> Other applications of machine learning on CT include prediction of 5-year all-cause mortality,<sup>39</sup> and detecting the presence of myocardial infarct using texture analysis methods.<sup>40</sup>

### Applications to Thoracic Imaging

Outside of the heart, another area of recent rapid development has been detection and measurement of lung nodules. Lung cancer has long been the leading cause of cancer-related mortality in the United States, and significant attention has been focused on establishing screening guidelines for the purpose of earlier detection. A landmark study published in 2011 compared low-dose CT with CXR in screening high-risk patients for pulmonary nodules.<sup>62</sup> Despite the low incidence of detected cancer in the patient group and high false-positive rates, the study was terminated early, as screening with low-dose CT showed reduced mortality from lung cancer. In response, the US Preventative Services Task Force recommended routine screening for this high-risk group, which is predicted to increase the number of pulmonary CTs performed annually.<sup>63</sup> The volume of studies combined with the time-consuming nature of detection, measurement, and comparison of pulmonary nodules, has prompted a boom in thoracic machine learning research. To address the need to improve accuracy and efficiency in utilizing machine learning approaches, multiple research

groups and challenges have focused on developing CAD algorithms for pulmonary nodules.<sup>64,65</sup>

In the late 2000s, there was considerable activity and interest in developing CAD algorithms to address this, utilizing traditional machine learning techniques.<sup>66</sup> These programs were shown to slightly improve radiologist detection of pulmonary nodules when used concurrently in interpretation. However, the sensitivity of these algorithms by themselves was relatively low compared with experienced radiologists. They were also often associated with a high false-positive rate, leading to increased time spent examining false nodules detected by the software. In addition, detection of nodules adjacent to structures such as vessels or pleura tended to be missed by the software.<sup>67</sup> This is an issue for both CAD programs and radiologists, as difficulty with confidently identifying the borders of a nodule can decrease the ability to assess growth and changes over time, a characteristic imperative for lung cancer treatment planning and follow-up. More recently, a CAD nodule detection system evaluated by Lo et al<sup>41</sup> incorporated a pulmonary vessel image suppression function. This improved detection of nodules that were initially missed due to their close relationship to vessels, and decreased interpretation time. Lung nodule detection increased from 64% to 80%, although at the cost of a slight reduction in specificity.<sup>41</sup> Although traditional machine learning approaches like this still yield modest nodule sensitivity in this range of performance,<sup>66</sup> early data from public challenges seem to show that deep learning approaches may improve upon this considerably.<sup>68</sup>

The explosion of available data in conjunction with advances in machine learning has opened the field of image analysis to an exciting direction with the potential to predict patient prognosis and even response to treatment. Components of a software package developed at UPenn were used to stratify patients with similar treatments for early non–small cell lung cancer into 2 distinct survival groups. This was achieved using an unsupervised clustering analysis method based on distinctive radiographic imaging features.<sup>42</sup> This field of image analysis, termed “radiomics,” is geared at characterizing image features and correlating them with a tumor phenotype, with the intent of classifying and staging tumors noninvasively. Extracted features convert radiographic images into mineable data and can be utilized to build predictive and prognostic models.<sup>24</sup> Subsequently, several groups have used machine learning to associate phenotypic descriptors on CT with overall survival<sup>43,69</sup> or disease-free survival<sup>70</sup> in non–small cell lung cancer. A review by Kolossváry et al<sup>71</sup> also describes how radiomic techniques may be implemented to assist with coronary artery calcium scoring.

The related concept of “radiogenomics” encompasses the effort to combine imaging phenotypes and tumor genetic data to perform better prognostication and target therapeutic decisions. Vardhanabhuti and Kuo<sup>72</sup> comprehensively review the development of this concept in lung cancer. Of note, a recent publication by Zhou et al<sup>73</sup> found that certain nodule imaging characteristics correlated with specific metagene groups, speculating that noninvasive imaging may help direct targeted therapeutic treatment through inference of the genetic or cell surface markers. Significant interest has been placed on attempting to define a radiomic signature for individual gene mutations (eg, EGFR, ALK, K-ras), and to correlate this with treatment response to targeted inhibitory agents.<sup>44,74,75</sup> Although machine learning was not utilized in most of these studies to connect gene expression data with individual imaging characteristics of a patient’s tumor, it is possible that deep learning may influence and expedite the next generation of radiomic studies. Specifically, the use of deep learning and its ability to automate the process of “feature engineering” across all scales of imaging phenotypes may provide an opportunity to bridge the gap between genetic, histologic, and imaging data.

The opportunities for machine learning to assist in diagnosis, prognostication, and treatment certainly are not limited to cardiac disease and oncologic applications. Other promising areas of thoracic research include the use of machine learning for aortic segmentation, thrombus detection, and fibrotic and COPD disease classification.<sup>45–47,76</sup>

## DEVELOPING COMMUNITIES, COMPETITIONS, AND THE CHALLENGES OF PUBLIC DATA SETS

The rapid growth of machine learning technologies and their potential applications has prompted the formation and growth of multiple research conferences. This includes the Conference on Machine Intelligence in Medical Imaging (C-MIMI) and has encouraged the formation of machine learning interest groups within existing radiologic societies such as the International Society for Magnetic Resonance in Medicine Machine Learning (ISMRM ML). A number of these conferences and groups have created open challenges aimed at promoting the development of algorithms for specific tasks or to solve clinical problems. The International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), is an annual conference that hosts multiple challenges across many fields, and is notable for highlighting cardiac segmentation, coronary artery reconstruction, and left ventricle shape modeling.<sup>77</sup> Several challenges have also been devoted to thoracic imaging, specifically lung nodule/lesion detection. The International Symposium on Biomedical Imaging (ISBI) offered a challenge called LUNA (LUng Nodule Analysis) in 2016, which called for algorithms that could automatically detect nodules on chest CT. Similar challenges were announced for the SPIE MI 2015 conference, the Data Science Bowl in 2017, and this year’s RSNA conference.

With these challenges, the organizers have offered data sets for the algorithms to be trained and tested on. The availability of large imaging data sets is currently a bottleneck for many areas of research and can be difficult to obtain outside of a major hospital-affiliated research group. Multiple factors contribute to this, including IRB approvals for data gathering, the need to assure patient privacy and anonymity, the complexities of working with DICOM files

and their size, and the cost of assembling and maintaining a database. Several public repositories of thoracic CT images are available, however, and have fueled machine learning growth and algorithm development (reviewed in Morris et al<sup>78</sup>). The National Lung Screening Trial (NLST) enabled the collection of a large volume of patient data of former smokers in an effort to determine the utility of low-dose CT screening, a recommendation later made by the US Preventative Service Task Force.<sup>63,79</sup> This imaging databank, in addition to blood, sputum, and urine data, has been made available to researchers. Similarly, the UK Biobank study collected genetic data and imaging, and has made this available for research,<sup>80</sup> and the publicly available Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) contains over 7000 annotated nodules for the purposes of CAD development.<sup>81</sup>

Although machine learning scientists continue to innovate with the available data and the community works to address the issue of data availability, researchers have been able to use these data sets to achieve excellent performance.<sup>25,49,82</sup> For example, the highly utilized UNet architecture was developed for, and won, a segmentation challenge for electron microscopy at ISBI 2012.<sup>17</sup> Although primarily developed to segment neuronal structures in electron microscopy stacks, it has since been successfully applied to a variety of biomedical image segmentation tasks.<sup>48,83</sup> Although these early studies show great promise, the most common training and testing sets from challenges are still limited in scope compared with the range of diseases that would be seen in clinical practice. Thus, these algorithms must yet prove themselves in the real world, as larger and more routine clinical data sets become available.

With the rapidly expanding availability of imaging data and the desire to develop clinically useful technology, there has been increasing effort to establish guidelines and offer guidance by larger radiologic societies. The American College of Radiology (ACR) recently established a Data Science Institute (DSI) to steer the introduction of AI into the practice of radiology. The ACR has begun defining standards for training, testing, validating, integrating, and monitoring AI algorithms in clinical practice. To do this, the Data Science Institute has defined “use cases” pinpointing precise scenarios within radiology workflows wherein potential automation could enhance patient care. These “use cases” are organized by organ system and include 25 cardiac and 1 thoracic application offered in an open-source directory labeled TOUCH-AI.<sup>84</sup> The ACR will also evaluate algorithm performance by providing an ongoing postmarket assessment.<sup>85</sup> A goal of developing standardized pathways for algorithm validation will be to expedite the Food and Drug Administration (FDA) regulatory review process to allow these modalities to have a clinical impact as soon as possible while maintaining a high level of quality.

## FDA APPROVAL PROCESS FOR NEW TECHNOLOGIES AND SOFTWARE

FDA currently has 2 approval processes for new medical devices relevant to imaging analysis software: the 510k clearance and the premarket approval (PMA) clearance. PMA clearance is a more stringent application path, and it typically involves clinical trials to demonstrate safety and efficacy. 510k clearance does not require the same rigorous clinical trials necessary for a new class of drugs or devices, but it demands that equivalence to existing products

or technologies be shown.<sup>86</sup> In December of 2016, the FDA expedited the regulatory process for certain medical devices by adding its Breakthrough Devices provision to the 21st Century Cures Act. The goal was to “help patients have more timely access to devices and breakthrough technologies,” with submissions ranging from dental implants to deep learning–based computer-aided diagnostic programs.<sup>87</sup> Through this, artificial intelligence in some forms can be classified as a medical device, and gain clearance through a separate “de novo” pathway without a predicate device for comparison.<sup>87</sup>

Many applications may fall under the umbrella of CAD. The FDA notes 2 major CAD categories: CADe, which includes tools aimed at automating detection and focusing the attention of a clinician onto an area of an image, and CADx, which includes tools aimed at assessment or likelihood of a disease and automation of diagnosis.<sup>88</sup> While CADe products fall under the 510k application path, CADx products have typically fallen into the more stringent PMA application path. As CADx products have the potential for more significant patient impact, the FDA has historically been more cautious. Recently, however, the requirements for CADx tools have been reduced.<sup>89</sup> Certain cancer detection tools will now fall under a new category of device and can be approved under the less stringent 510K pathway. The growing number of applications integrating neural networks has likely contributed to this change. Even with more lenient approval processes, CAD products will continue to require regulatory attention going forward.

## VALIDATION, UTILITY, AND GENERALIZABILITY

As the population ages and becomes more medically complex, the need for imaging will continue to increase. Deep learning–based diagnostic tools offer many potential benefits, and may increase physician efficiency, improve the accuracy of diagnosis, and enhance consistency between different sites and institutions. As early adopters, radiologists have the opportunity to set the standards for this technology and direct evaluation of programs to ensure their accuracy and safety for patients. Bringing a technology to the clinic will involve several key elements: peer review, analytical validity, clinical validity, clinical utility, and generalizability. Within these, the overall performance compared with a ground truth, performance on individual components of the task, and mechanisms of failure should be examined.

Although the medical literature has historically undergone a process of peer review, it is important to recognize that preprint websites are increasingly used as a form of communication, especially in the machine learning community. In this rapidly changing and evolving field, a repository for the prerelease of manuscripts such as arXiv is valuable for sharing information and distributing cutting-edge findings. However, it is important to recognize the limitations and pitfalls of this method of communication. Manuscripts on prerelease repositories are moderated upon submission, but they are not required to undergo the peer-review process, and new versions may be uploaded or updated at any time. Therefore, authors must use caution when making clinical decisions on the basis of manuscripts from preprint websites until peer-review scrutiny is applied.

Analytical validity is typically assessed during the testing and development of new technologies, and it tests the performance relative to preexisting technologies. For example, an algorithm might measure the caliber of the



**FIGURE 6.** Consideration of the components of the overall task (e.g., dice per slice when segmenting the ventricles of the heart) can help identify sources of error and understanding of areas wherein oversight may be necessary. Without considering the individual components (eg, considering only overall dice per volume), systematic errors may be overlooked. full color online

aorta, and a comparison of the measurements between the algorithm and a human observer may be sufficient. It is harder to assess analytical validity when the outcome metric is subjective. In such situations, a consensus among experts may be necessary. For example, when contouring cardiac chambers, the apical and basal heart boundaries are a source of disagreement between readers, and protocols vary between institutions. However, a consensus average of segmentations from a panel of experts can be helpful for training algorithms.<sup>90</sup> Analytical validity can also be established by analyzing algorithm performance on individual components of a task. Understanding where an algorithm fails can provide insight into areas that require human supervision. If an algorithm has particular difficulty segmenting the cardiac apex, this might not be readily apparent if only the overall similarity scores are considered (Fig. 6). Further, these insights may direct research to improve the performance of the algorithms.

Clinical validity is a concept distinct from analytical validity, and it touches upon whether a technology works within the range of performance needed for clinical decision-making. In the example of an algorithm that measures the caliber of the ascending aorta, it is important to assess whether any differences in measurement between the algorithm and the current standard would impact the clinical management. In other words, it may be possible to prove differences between an algorithm and human observer, but these may not be clinically relevant. In the example of ventricular segmentation, the apex of the left ventricle may have considerably different contours between an algorithm and human observers, but it may contribute very little to the total volume, and therefore have little impact on the overall volumes or ejection fraction.

Clinical utility is a related concept for determining the potential value of a new tool. This is where radiologists may play a key role in the development of technology. We may routinely only provide linear or bilinear measurements of pulmonary nodules and lymph nodes in daily clinical practice due to efficiency and time constraints, but it may also be clinically useful to provide volumetric measurements. We may choose not to provide such measurements if the measurement tools are too cumbersome, inaccurate, or work too inconsistently to be useful. Deep convolutional neural networks have the potential to address this need if well implemented into the radiologists' workflow, and clinical domain knowledge is critical for directing the



development of new applications. With such collaboration, new technologies will be more likely adopted into practice.

One major consideration for algorithms that are to be used in clinical practice is the extent to which they will generalize to clinical populations. The recent tragic accidents from Uber and Tesla have shown that overestimation of the robustness and generalizability of this technology can yield fatal results.<sup>91,92</sup> Many algorithms are trained on a limited patient population, or from public data sets of predominantly normal patients, and may not necessarily work on the typical population seen in clinical practice. Another pitfall is that algorithms may be developed on a single scanner type or imaging technique that is institution specific. Such algorithms may not generalize to other scanners or protocols. For example, an algorithm designed to detect pulmonary nodules on 1.25-mm hard kernel reconstructions may not work as well on thick slices. As such, the developers of these tools must work to ensure that the patient populations and imaging protocols on which models are developed are representative of clinical patient populations, and users must appreciate limitations of models as they deploy them in clinical practice. One of the major advantages of modern deep learning models is their ability to learn and improve by being “fine tuned” on new training data. This allows refinement of models and retraining over time on broader populations.

## CONCLUSIONS

The applications of deep neural networks and other machine learning technologies to long-standing problems in radiology are rapidly advancing and promise to shape the future of the specialty. Both supervised and unsupervised techniques have been applied in current technologies (Tables 1, 2). They will likely become fundamental to practice, and soon, as pervasive and unnoticeable as other technologies we have integrated into daily use. In many ways, these technologies are already a large part of our lives. Phones recognize us, speak to us, manage our schedules, and reroute our wrong turns. Entertainment systems and online shopping trackers recommend things we would enjoy on the basis of previous selections and others with similar behaviors. These machine learning processes are so well used that they have become transparent in our daily activities. Outside of digital transcription, machine learning algorithms are only beginning to emerge in the daily practice of radiology. The breadth of problems that machine learning can help address is immense, and will likely mature rapidly in areas of detection, characterization and prognostication of disease, and individualized treatment decisions. As early adopters of this new technology, radiologists should be cautious consumers, and think critically about the new advancements to ensure that they are safe and effective tools in clinical practice. Integration of machine learning into the daily workflow has the potential to augment our capabilities and make radiologists more efficient, more focused on diagnosis and higher order tasks, and better able to address the needs of referring physicians and patients.

## REFERENCES

- Hastie T, Tibshirani R, Friedman J. *Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction*. New York, NY: Springer; 2009.
- Rosenblatt F, Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev*. 1958;65:386–408.
- Minsky M, Papert S. *Perceptrons; An Introduction to Computational Geometry*. Cambridge, Massachusetts: MIT Press; 1969.
- Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*. 1986;323:533–536.
- Cybenko G. Approximation by superpositions of a sigmoidal function. *Math Control Signals Syst*. 1989;2:303–314.
- Lecun Y, Boser B, Denker JS, et al. Handwritten digit recognition with a back-propagation network. 1990.
- Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. *Neural Comput*. 2006;18:1527–1554.
- Russakovsky O, Deng J, Su, H, et al. ImageNet large scale visual recognition challenge. *Int J Comp Vision*. 2015;115:211–252.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing System, Vol. 1. 2012; 1097–1105.
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014.
- Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. 2015.
- He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. 2015.
- Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. 2013.
- Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. 2014.
- Chollet F Keras. 2018.
- Ding J, Li A, Hu Z, et al. Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks. 2017.
- Ronneberger O, Fischer P, Brox T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Cham: Springer; 2015:234–241.
- Maglaveras N, Stamkopoulos T, Diamantaras K, et al. ECG pattern recognition and classification using non-linear transformations and neural networks: a review. *Int J Med Inform*. 1998;52:191–208.
- Kiranyaz S, Ince T, Gabbouj M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Trans Biomed Eng*. 2016;63:664–675.
- Afsar F, Arif M, Yang J. Physiological Measurement Detection of ST segment deviation episodes in ECG using KLT with an ensemble neural classifier recent citations. *Physiol Meas*. 2008;29:747–760.
- Rajpurkar P, Hannun AY, Haghpanahi M, et al. Cardiologist-level arrhythmia detection with convolutional neural networks. 2018.
- Al SJ, Anchouche K, Singh G, et al. Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. *Eur Heart J*. 2018. Doi:10.1093/eurheartj/ehy404.
- Singh G, Al'Aref SJ, Van Assen M, et al. Machine learning in cardiac CT: basic concepts and contemporary data. *J Cardiovasc Comput Tomogr*. 2018;12:192–201.
- Thawani R, McLane M, Beig N, et al. Radiomics and radiogenomics in lung cancer: a review for the clinician. *Lung Cancer*. 2018;115:34–41.
- Avendi MR, Kheradvar A, Jafarkhani H. Automatic segmentation of the right ventricle from cardiac MRI using a learning-based approach. *Magn Reson Med*. 2017;78:2439–2448.
- Winther HB, Schmidt B, Wacker FK, et al. *v-net: deep learning for generalized biventricular cardiac mass and function parameters*. 2017.
- Tan LK, McLaughlin RA, Lim E, et al. Fully automated segmentation of the left ventricle in cine cardiac MRI using neural network regression. *J Magn Reson Imaging*. 2018;48:140–152.
- Baessler B, Mannil M, Oebel S, et al. Subacute and chronic left ventricular myocardial scar: accuracy of texture analysis on nonenhanced cine MR images. *Radiology*. 2018;286:103–112.
- Dawes TJW, De Marvao A, Shi W, et al. Machine learning of three-dimensional right ventricular motion enables outcome

- prediction in pulmonary hypertension: a cardiac MR imaging study 1. *Radiology*. 2017;283:381–390.
30. Ortiz J, Ghefter CG, Silva CE, et al. One-year mortality prognosis in heart failure: a neural network approach based on echocardiographic data. *J Am Coll Cardiol*. 1995;26:1586–1593.
  31. Narula S, Shameer K, Mabrouk, A, et al. Machine-learning algorithms to automate morphological and functional assessments in 2D echocardiography. *J Am Coll Cardiol*. 2016;68:2287–2295.
  32. Sengupta PP, Huang YM, Bansal M, et al. A cognitive machine learning algorithm for cardiac imaging: a pilot study for differentiating constrictive pericarditis from restrictive cardiomyopathy HHS public access. *Circ Cardiovasc Imaging*. 2016;9:6. Doi:10.1161/CIRCIMAGING.115.004330.
  33. Sengur A. Support vector machine ensembles for intelligent diagnosis of valvular heart disease. *J Med Syst*. 2012;36:2649–2655.
  34. Moghaddasi H, Nourian S. Automatic assessment of mitral regurgitation severity based on extensive textural features on 2D echocardiography videos. *Comput Biol Med*. 2016;73:47–55.
  35. Vidya KS, Ng EYK, Acharya UR, et al. Computer-aided diagnosis of myocardial infarction using ultrasound images with DWT, GLCM and HOS methods: a comparative study. *Comput Biol Med*. 2015;62:86–93.
  36. Wolterink JM, Leiner T, De Vos BD, et al. Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks. *Med Image Anal*. 2016;34:123–136.
  37. Işgum I, Prokop M, Niemeijer M, et al. Automatic coronary calcium scoring in low-dose chest computed tomography. *IEEE Trans Med Imaging*. 2012;31:2322–2334.
  38. Itu L, Rapaka S, Passerini T, et al. A machine-learning approach for computation of fractional flow reserve from coronary computed tomography. *J Appl Physiol*. 2016;121:42–52.
  39. Motwani M, Dey D, Berman DS, et al. 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.
  40. Mannil M, von Spiczak J, Manka R, et al. Texture analysis and machine learning for detecting Myocardial infarction in non-contrast low-dose computed tomography. *Invest Radiol*. 2018;53:338–343.
  41. Lo SB, Freedman MT, Gillis LB, et al. Computer-aided detection of lung nodules on CT with a computerized pulmonary vessel suppressed function. *AJR Am J Roentgenol*. 2018;210:480–488.
  42. Li H, Galperin-Aizenberg M, Pryma D, et al. Unsupervised machine learning of radiomic features for predicting treatment response and overall survival of early stage non-small cell lung cancer patients treated with stereotactic body radiation therapy. *Radiother Oncol*. 2018;129:218–226.
  43. Song J, Liu Z, Zhong W, et al. Non-small cell lung cancer: quantitative phenotypic analysis of CT images as a potential marker of prognosis. *Nat Publ Gr*. 2016;6:38282.
  44. Yamamoto S, Korn RL, Oklu R, et al. ALK molecular phenotype in non-small cell lung cancer: CT radiogenomic characterization 1. *Radiol Radiol*. 2014;272:568–576.
  45. Ying J, Dutta J, Guo N, et al. Classification of exacerbation frequency in the COPDGene cohort using deep learning with deep belief networks. *IEEE J Biomed Heal Informatics*. 2016:1–1. Doi:10.1109/JBHI.2016.2642944. [Epub ahead of print].
  46. González G, Ash SY, Vegas Sanchez-Ferrero G, et al. Disease Staging and Prognosis in Smokers Using Deep Learning in Chest Computed Tomography. *Am J Respir Crit Care Med*. 2017;197:193–203.
  47. López-Linares K, Aranjuelo N, Kabongo L, et al. Fully automatic detection and segmentation of abdominal aortic thrombus in post-operative CTA images using Deep Convolutional Neural Networks. *Med Image Anal*. 2018;46:202–214.
  48. Lieman-Sifry J, Le M, Lau F, et al. FastVentricle: Cardiac Segmentation with ENet. 2017.
  49. Petitjean C, Zuluaga MA, Bai W, et al. Right ventricle segmentation from cardiac MRI: a collation study. *Med Image Anal*. 2015;19:187–202.
  50. Cardio—Arterys Available at: <https://www.arterys.com/cardio>.
  51. Cardiac MRI and CT Software—circle cardiovascular imaging. Available at: <https://www.circlecvi.com/>.
  52. HeartVista Cardiac Package, HeartVista. Available at: <https://www.heartvista.ai/>.
  53. Knackstedt C, Bekkers SCAM, Schummers G, et al. 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.
  54. Sudarshan VK, Rajendra Acharya U, Ng EYK, et al. Data mining framework for identification of myocardial infarction stages in ultrasound: a hybrid feature extraction paradigm (PART 2). *Comput Biol Med*. 2016;71:241–251.
  55. Takx RAP, De Jong PA, Leiner, T, et al. Automated coronary artery calcification scoring in non-gated chest CT: agreement and reliability. *PLOS One*. 2013;9:e91239.
  56. Işgum I, de Vos BD, Wolterink JM, et al. Automatic determination of cardiovascular risk by CT attenuation correction maps in Rb-82 PET/CT. *J Nucl Cardiol*. 2017;25:2133–2141.
  57. Min JK, Taylor CA, Achenbach S, et al. Noninvasive fractional flow reserve derived from coronary CT angiography. *JACC Cardiovasc Imaging*. 2015;8:1209–1222.
  58. Taylor CA, Gaur S, Leipsic J, et al. Effect of the ratio of coronary arterial lumen volume to left ventricle myocardial mass derived from coronary CT angiography on fractional flow reserve. *J Cardiovasc Comput Tomogr*. 2017;11:429–436.
  59. Coenen A, Lubbers MM, Kurata A, et al. Fractional flow reserve computed from noninvasive CT angiography data: diagnostic performance of an on-site clinician-operated computational fluid dynamics algorithm. *Radiology*. 2015;274:674–683.
  60. Coenen A, Kim Y-H, Kruk M, et al. Diagnostic accuracy of a machine-learning approach to coronary computed tomographic angiography-based fractional flow reserve. *Circ Cardiovasc Imaging*. 2018;11:e007217. Doi:10.1161/CIRCIMAGING.117.007217.
  61. Duguay TM, Tesche C, Vliegenthart R, et al. Coronary computed tomographic angiography-derived fractional flow reserve based on machine learning for risk stratification of non-culprit coronary narrowings in patients with acute coronary syndrome. *Am J Cardiol*. 2017;120:1260–1266.
  62. National Lung Screening Trial Research Team, Aberle DM, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365:395–409.
  63. Final update summary: lung cancer: screening—US Preventive Services Task Force. Available at: <https://www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/lung-cancer-screening>.
  64. Valente IRS, Cortez PC, Neto EC, et al. Automatic 3D pulmonary nodule detection in CT images: a survey. *Comput Methods Programs Biomed*. 2016;124:91–107.
  65. Luna 16 Grand Challenge. Available at: <https://luna16.grandchallenge.org/>.
  66. Goo JM. A computer-aided diagnosis for evaluating lung nodules on chest CT: the current status and perspective. *Korean J Radiol*. 2011;12:145–155.
  67. Das M, Ley-Zaporozhan J, Gietema HA, et al. Accuracy of automated volumetry of pulmonary nodules across different multislice CT scanners. *Eur Radiol*. 2007;17:1979–1984.
  68. Setio AAA, Traverso A, de Bel T, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *Med Image Anal*. 2017;42:1–13.
  69. Aerts HJWL, Rios Velazquez E, Leijenaar RTH, et al. ARTICLE decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006. Doi:10.1038/ncomms5006.
  70. Huang Y, Liu Z, He L, et al. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non—small cell lung cancer. *Radiology*. 2016;281:947–957.
  71. Kolossváry M, Keller Mayer M, Merkely B, et al. Cardiac computed tomography radiomics. *J Thorac Imaging*. 2018;33:26–34.

72. Vardhanabhuti V, Kuo MD. Lung cancer radiogenomics. *J Thorac Imaging*. 2018;33:17–25.
73. Zhou M, Leung A, Echegaray S, et al. Non-small cell lung cancer radiogenomics map identifies relationships between molecular and imaging phenotypes with prognostic implications. *Radiology*. 2017;286:307–315.
74. Aerts HJWL, Grossmann P, Tan, Y, et al. Defining a radiomic response phenotype: a pilot study using targeted therapy in NSCLC. *Sci Reports*. 2016;6:33860.
75. Rizzo S, Petrella F, Buscarino, V, et al. CT radiogenomic characterization of EGFR, K-RAS, and ALK mutations in non-small cell lung cancer. *Eur Radiol*. 2016;26:32–42.
76. Kim SY, Diggans J, Pankratz D, et al. Classification of usual interstitial pneumonia in patients with interstitial lung disease: assessment of a machine learning approach using high-dimensional transcriptional data. *Lancet Respir Med*. 2015;3:473–482.
77. miccai2018. WORKSHOP & CHALLENGE & TUTORIAL. Available at: <https://www.miccai2018.org/en/WORKSHOP—CHALLENGE—TUTORIAL.html>.
78. Morris MA, Saboury B, Burkett B, et al. Reinventing radiology: Big Data and the future of medical imaging. *J Thorac Imaging*. 2018;33:4–16.
79. NLST. Datasets—NLST—The Cancer Data Access System. Available at: <https://biometry.nci.nih.gov/cdas/nlst/>.
80. UK Biobank. Resources, UK Biobank. Available at: <https://www.ukbiobank.ac.uk/resources/>.
81. Armato SG, McLennan G, Bidaut L, et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med Phys*. 2011;38:915–931.
82. Ringenberg J, Deo M, Devabhaktuni V, et al. Fast, accurate, and fully automatic segmentation of the right ventricle in short-axis cardiac MRI. *Comput Med Imaging Graph*. 2014;38:190–201.
83. Le BM, Lieman-sifry J, Lau F, et al. Computationally efficient cardiac views projection using 3D convolutional neural networks. 2017:109–116. Doi:10.1007/978-3-319-67558-9.
84. American College of Radiology. TOUCH-AI Directory. Available at: <https://www.acrdsi.org/DSI-Services/TOUCH-AI>.
85. American College of Radiology. ACR data science institute structures artificial intelligence development to optimize radiology care. Available at: <https://www.acr.org/Media-Center/ACR-News-Releases/2017/ACR-Data-Science-Institute-Structures-Artificial-Intelligence-Development-to-Optimize-Radiology-Care>.
86. Health C for D and R. Overview of device regulation. Available at: <https://www.fda.gov/medicaldevices/deviceregulationandguidance/overview/default.htm>.
87. Health C for D and R. How to study and market your device—expedited access pathway program. Available at: <https://www.fda.gov/medicaldevices/deviceregulationandguidance/howtomarketyourdevice/>.
88. Health C for D and R. Guidance documents (medical devices and radiation-emitting products)—clinical performance assessment: considerations for computer-assisted detection devices applied to radiology images and radiology device data - premarket approval (pma) and premarket N. Available at: <https://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/default.htm>.
89. Ridley E. C-MIMI: FDA decision paves the way for imaging AI. Available at: <https://www.auntminnie.com/index.aspx?sec=log&itemID=118370>.
90. Suinesiaputra A, Cowan BR, Al-Agamy AO, et al. A collaborative resource to build consensus for automated left ventricular segmentation of cardiac MR images. *Med Image Anal*. 2014;18:50–62.
91. Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian, Technology. *The Guardian*. 2018. Available at: [www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe](http://www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe). Accessed September 30, 2018.
92. Tesla autopilot is not having a great day. 2018. Available at: <https://slate.com/technology/2018/09/tesla-autopilot-problems-elon-musk.html>. Accessed September 30, 2018.