BJR

https://doi.org/10.1259/bjr.20180416

Received: 05 May 2018 | Revised: 05 October 2018 | Accepted: 10 October 2018

Cite this article as:
Chan S, Siegel EL. Will machine learning end the viability of radiology as a thriving medical specialty?. *Br J Radiol* 2019; **91**: 20180416.

REVIEW ARTICLE

Will machine learning end the viability of radiology as a thriving medical specialty?

[1]STEPHEN CHAN, MD, MBA, MPH and [2]ELIOT L SIEGEL, MD

[1]Department of Radiology, Harlem Hospital Center and Columbia University, New York, NY, USA
[2]Department of Diagnostic Radiology and Nuclear Medicine, VA Maryland Health Care System, Baltimore, MD, USA

Address correspondence to: Dr Stephen Chan
E-mail: *sc56md@gmail.com*

ABSTRACT

There have been tremendous advances in artificial intelligence (AI) and machine learning (ML) within the past decade, especially in the application of deep learning to various challenges. These include advanced competitive games (such as Chess and Go), self-driving cars, speech recognition, and intelligent personal assistants. Rapid advances in computer vision for recognition of objects in pictures have led some individuals, including computer science experts and health care system experts in machine learning, to make predictions that ML algorithms will soon lead to the replacement of the radiologist. However, there are complex technological, regulatory, and medicolegal obstacles facing the implementation of machine learning in radiology that will definitely preclude replacement of the radiologist by these algorithms within the next two decades and beyond. While not a comprehensive review of machine learning, this article is intended to highlight specific features of machine learning which face significant technological and health care systems challenges. Rather than replacing radiologists, machine learning will provide quantitative tools that will increase the value of diagnostic imaging as a biomarker, increase image quality with decreased acquisition times, and improve workflow, communication, and patient safety. In the foreseeable future, we predict that today's generation of radiologists will be replaced not by ML algorithms, but by a new breed of data science-savvy radiologists who have embraced and harnessed the incredible potential that machine learning has to advance our ability to care for our patients. In this way, radiology will remain a viable medical specialty for years to come.

INTRODUCTION

Recent articles in the medical and lay press have underscored the tremendous progress made in "artificial intelligence", and raised the prospect that computers, using machine learning (ML) algorithms, will soon replace radiologists.[1,2] As recently as 2016, Geoffrey Hinton - founder of the branch of machine learning known as "deep learning"– was quite emphatic in stating this perspective, recently stating, "I think that if you work as a radiologist you are like Wile E. Coyote in the cartoon. You're already over the edge of the cliff, but you haven't yet looked down. There's no ground underneath. It's just completely obvious that in five years deep learning is going to do better than radiologists. It might be ten years".[2] At that time, Hinton clearly indicated that machine learning would be a disruptive technology for radiologists. As described in Christiansen's seminal work,[3] there are three essential characteristics of a disruptive technology, each of which is satisfied or could potentially be satisfied by machine learning [Table 1]. Since machine learning appears to fulfill these three essential characteristics, one could conclude that machine learning represents a disruptive technology. However, more recent work by Christiansen et al suggests that there are two other criteria defining a disruptive technology.[4] These include: (1) the presence of only a low-end foothold or a new market foothold in the industry; and (2) the unknowing or deliberate ignorance of the new technology by the incumbent leaders in the industry. As indicated later in this article, we believe that neither of these latter two criteria is met by machine learning in radiology. With respect to machine learning, it now appears that leading radiology organizations have begun to adopt strategies for handling this potentially disruptive technology.[5]

In 2016, Chockley (a medical student) and Emanuel (an internal medicine physician and "Obamacare" architect) identified three threats to the future practice of diagnostic imaging, with machine learning singled out as the "ultimate threat".[6] They made the following two assertions: (1) "machine learning will become a powerful force in radiology in the next 5 to 10 years, not in multiple decades"; and (2) "indeed, in a few years there may [be] no specialty called radiology".[6] If they meant that the computer will largely replace the radiologist in 5 to 10 years (as implied in

Table 1. Three fundamental characteristics of a disruptive technology (as related to machine learning)

Key characteristics of disruptive technology	Is this true of machine learning and why?
The overall performance level offered by early versions of the disruptive technology is far inferior to the current technology.	Partially true. As of 2018, there is no version of an ML algorithm whose performance can match the accuracy and breadth of a human radiologist.
The customers currently served by the incumbent industry leaders often provide little (or even negative) feedback about the value of the new technology.	True. No one in the current generation of clinicians is requesting that radiology interpretations be provided solely by ML systems.
The customers who benefit most from the emergence of a new technology with inferior performance characteristics are often different from the ones currently served by the market leaders.	Probably true. Initial customers for ML systems have not yet been identified, although they may include clinicians (or hospitals) from developing nations, research subjects from population health studies, or large corporations with preventive health imaging needs.

ML, machine learning;

their work), then we completely disagree and believe that this is an ill-informed prediction borne out of a lack of domain knowledge of radiology. Their view reflects a fundamental misunderstanding of the nature of the work performed by radiologists, as well as a lack of appreciation of exactly how difficult it will be for machine learning to replace the wide variety of imaging interpretation and patient care tasks inherent in the practice of radiology. We note that the specter of a future in which radiologists are no longer needed to provide image interpretation services has seriously alarmed forward-thinking medical students, radiology residents, and fellows, impelling some to ask if they should quit or avoid radiology residency because of the risk of not getting a job after residency.[7,8] Indeed, that fear could potentially damage the radiology profession by discouraging talented medical students from choosing radiology as their future career. We seek to allay such fear by careful examination of the recent developments in machine learning, and by detailed evaluation of the kind of technological development necessary to render the broad range of radiological diagnosis. Specifically, there are two fundamental sources of misunderstanding that lead many individuals to conclude that radiologists can be easily replaced by machine learning.

Misunderstanding #1: Machine learning can easily absorb and process the wide variation of information and ambiguity inherent in interpretation of medical images.

Remarkable achievements have been made in machine learning such as the impressive computer vision performance on identification of objects in everyday pictures from the Stanford ImageNet challenge[9] and the victory of Google's AlphaGo over the 2016 human champion of Go.[10] Computer scientists cite these accomplishments to assert that unsupervised machine learning will soon be rendering medical imaging findings and diagnoses. However, board games such as "Go" focus on a very "narrow" artificial intelligence task where a winning *vs* losing status can be assessed, whereas medical imaging is associated with far greater amount of ambiguity, and a larger variety of features, classifications, and outputs. It is also likely that thousands of "narrow" algorithms based on separate large, well-annotated databases will be required for a computer to begin to compete with a radiologist for comprehensive diagnostic assessment of even a single modality covering a single anatomical region of the body.

Advances in self-reinforcement learning have led to substantial further improvements in "AlphaGo" resulting in "AlphaGoZero" which utilizes an approach in which the computer is provided with the basic rules of the game and learns by playing itself large numbers of games rather than learning by analyzing the play of human experts.[11] Although possible in games with simple defined rules such as Go or chess, analogous self-reinforcement learning is not so easily attainable in radiology, given the lack of a simple set of rules of the "radiology game" to allow this sort of self-play. Barring an unforeseen major technological breakthrough, it is likely that human annotation and guidance will likely be necessary at multiple stages in the development of machine learning in medical imaging, augmented by increases in computing power and conceptual advances in artificial intelligence. This pattern is exemplified by the technological development of the Google Translate app in which significant conceptual advances in ML-based language translation needed to be made by computer scientists, who then were able to render the sequential and contextual information inherent in languages far more amenable to deep neural networks.[1]

Misunderstanding #2: Computer-aided detection and computer-aided diagnosis is an immediate technological precursor to ML algorithms.

Chockley and Emanuel cite the current performance of computer-aided detection (CADe) and computer-aided diagnosis (CADx) in various areas of radiology - including the field of mammography—as evidence of success stories, with machine learning "working as well as or better than the experienced radiologist". Indeed, many papers and presentations describing CAD systems in mammography have claimed a performance level in lesion detection similar to that of an experienced radiologist.[12–14] Based on that research, CAD was approved by the Food and Drug Administration (FDA) for use with mammography and has been widely introduced into radiology practices across the U.S. as an adjunctive technology for mammography.[15] However, in spite of its widespread use for the past decade, it has not been shown to improve detection rates in academic settings, and it is unclear whether or not CAD improves the detection rate of invasive breast carcinoma in community practice.[16] In addition, the use of CAD can be detrimental if its limitations are not understood.[17] While review of mammographic images with adjunctive CAD would likely be considered the de facto standard of care in community mammography practice,[18] we note that CAD systems have not replaced the practicing radiologist. In practice, survey data suggests that more than half (~62%) of radiologists have never or rarely changed their report as a result of CAD findings in mammography, and about a third of radiologists never or rarely use the findings generated by CAD.[19] There has been an initial demonstration of a machine learning tool to help separate

high-risk breast lesions that are more likely to become cancerous from those that are at lower risk.[20] However, we know of no CAD program in clinical use that continually receives feedback about its diagnostic performance – a task that is essential to learning from experience. Finally, we are not aware of any mammography CAD/machine learning software program that formally compares a prior mammogram to a current mammogram, just as a human reader would do. Yet, comparison with prior imaging studies remains a fundamental diagnostic task in mammography and radiology, especially in assessment of interval change.

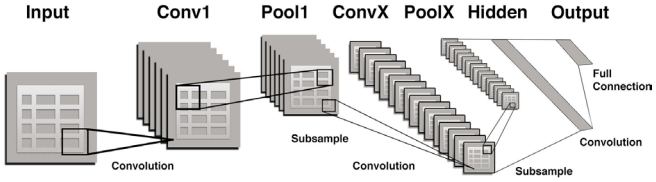
What is machine learning?

The term "machine learning" encompasses a variety of advanced iterative statistical methods used to discover patterns in data and, although inherently non-linear, is based heavily on linear algebra data structures. It can be utilized to help to improve prediction performance, dynamically adjusting the output model when the data change over time. Historically, there have been two very broad groupings of artificial intelligence applied to cognitive problems in everyday human work issues. The first is expert systems, in which software programs are constructed to mimic human performance, based upon rules that were derived from "experts" by the programmer. An example of this was the medical diagnostic program "Internist I", which was designed to capture the expertise of the chairman of internal medicine at the University of Pittsburgh School of Medicine, Dr Jack Myers.[21] The second is machine learning, in which the most recent advances in computer vision and speech recognition have come from a form of machine learning known as "deep learning", which uses a technique known as convolutional neural networks, and is based on a set of algorithms that attempt to model high-level abstractions in data. Neural networks use a variety of approaches loosely based on what are referred to as interconnected cells, analogous to the interneurons of the human nervous system.

Convolutional neural networks (CNNs) are a special type of neural network that is optimized for image pattern recognition.[22] Unlike other types of artificial neural networks, the majority of nodes (neurons) in a CNN are only connected to a subset of other nodes, particularly those in closer proximity in an image which enhances their ability to recognize local features of an image. In brief, a CNN consists of multiple layers between the input and output layers (Figure 1) The main building blocks are the convolution layer which can be thought of as a series of adjustable image filters that can emphasize or de-emphasize certain aspects of an image such as borders, colors, noise, and texture. Each of the multiple convolution layers within a CNN is followed by a pooling layer which serves to reduce the number of parameters. For example, the commonly used technique referred to as "max pooling" simply chooses the maximum pixel value within each small portion of an image and assigns all pixels to that value. The final layer of a CNN is a fully connected one similar to other types of artificial neural networks.

The term "artificial intelligence (AI)" is currently commonly utilized in medical imaging in both the lay and scientific literature to refer to machine learning in general and CNNs specifically. Although the architecture of both CAD programs and

Figure 1. A portion of the input (image) is passed to each successive pair of convolutional/pooling layers (filter/parameter reducers) with several convolution and pooling layers added before an output (prediction) is made. Initial layers tend to represent general features such as edges and colors and later layers represent features of increasing complexity, such as corners, and then textures, followed by even complex features such as a snout or whiskers, and finally entire objects such as a dog or cat. Finally, there is a "fully connected layer" that flattens input from other layers transforming them into a decision on whether the output belongs to a certain class (*e.g.*dog *vs* cat). Errors in classification in a training set are then "back-propagated" to modify and/or update the filters so that overall errors are minimized.



ML-based algorithms are designed by humans, the essential discriminatory functions of the AI algorithms emerge directly from the data, and, unlike CAD, do not require humans to identify and compute the critical features.[23] This emergence of algorithms from the data is what prompted Wired magazine to suggest that machine learning may represent "the end of code".[24] Specifically, the algorithms to predict such things as the presence of an intracranial bleed, or malignancy in a prostate MRI study will emerge directly from the "learned", iteratively adjusted values of the nodes in a CNN. Those values themselves represent the trained model and the "training" continues with the introduction of each annotated dataset. Although inputs to CNNs are not always raw images (and may be segmented and co-registered prior to classification), the many steps such as feature extraction, segmentation, registration, and statistical analysis utilized by the previous generation of so-called CAD (computer-aided detection or computer-aided diagnosis) software are not required. Both AI and CAD techniques can be utilized to develop medical imaging software, but AI algorithms typically require much more annotated data but subsequently take much less time to develop using fewer steps. Most developers previously utilizing these more human, understandable basic steps have made the transition to the use of CNNs. Creation of a machine learning model can be performed much faster (*e.g.* 5 to 6 days rather than 5 to 6 months or years) than traditional computer-aided detection and diagnosis (CAD_e/CAD_x).

The nature of learning in machine learning can be confused with that associated with humans. Machine learning has been defined as "algorithm-driven learning by computers such that the machine's performance improves with greater experience" and indicated that it "involves the construction of algorithms that learn from data and make predictions on the basis of those data."[6] Although this definition may imply a type of self-reinforcement learning "with greater experience", in actuality, these algorithms in diagnostic imaging have improved largely by the addition of annotated data based on human review or patient

outcome information, rather than repetitive application of work-in-progress. This is a fundamental difference from the use of machine learning in strategy games such as Chess and Go in which there are well-defined parameters of success. The current regulatory constraint on progressive learning of these algorithms is that clearance of these products has been based on a well-defined training set and a test set (in order to establish performance against ground truth). Therefore, from a regulatory perspective, it is unclear whether or not the FDA would allow the continued modification or improvement of an ML system by incorporation of additional local patient data from a given clinical radiology practice. As is true of other related statistical techniques such as linear regression, additional "processed" data are required to enhance the model rather than simply "learning from experience" *per se*. Therefore, computers remain far less efficient than humans at learning and generalizing concepts from a relatively small dataset. This has been mitigated somewhat by techniques such as transfer learning which takes patterns learned from a related task as a starting point effectively kickstarting the training (also known as "one-shot learning") that takes advantage of existing knowledge to train using just a single or very few examples.[25]

Will an ML system soon be capable of replicating the work of a radiologist?

Major technological developments in machine learning have been made over the last few years, including advances in deep learning algorithms, further advances in graphics processing units speed and memory, and the exponential growth of corporate investment. However, there are several independent factors which suggest that successful replacement of the radiologist's work is likely to be substantially more difficult than is currently envisioned by some non-radiologist health care experts and computer science futurists. These potential challenges are based upon unique aspects of the radiological image, the visual processing capability of the radiologist, and the role of the radiologist in maximizing and maintaining clinical relevance in image interpretation. In particular, the job of the radiologist is not simply to detect findings related to various imaging studies, but to determine "what is wrong with this picture" and help determine the future course of action in the diagnostic evaluation and therapeutic decision-making. Determining "what is wrong with this picture" is a much harder task that extends far beyond the capabilities of the current generation of computer vision systems. Contextualization of the imaging information in diagnostic evaluation and therapeutic decision-making may be an even more difficult task to replicate.

A major challenge for ML algorithms is the greater technical complexity of the radiological image as compared to those images typically used in object-recognition tests for computer "visualization". In addition to the differing imaging modalities, this complexity includes a wide variety of manifestations of normal and pathological findings, multiple sequential images in a cross sectional/volumetric dataset, with much higher complexity of data and raw number of pixels/voxels in medical images. It also includes a high level of ambiguity and difficulty in annotation that is not inherent in the ImageNet challenges that have used common objects such as dogs, cats, bikes, cars, etc. Another major technical challenge is the development of a "reasonable" detection rate of abnormalities without an excessive rate of false-positive findings as compared with human performance. For more than 20 years, CADe and CADx programs, such as those used to detect lung nodules or breast masses, have been fraught with the issue of frequent false-positive findings (*i.e.* low specificity) and we suggest that this problem may also be an intrinsic problem for deep learning algorithms.[22,26] This problem is further complicated by: (1) the multiple classes of imaging abnormalities detected on diagnostic imaging studies; (2) the time and expense associated with the collection of large annotated datasets (such as ChestX-ray14 and the Cancer Imaging Archive) required for deep learning, of which a fair number are available in the public domain,[27,28] but many more are needed[29]; (3) the difficulties associated with ensuring sufficient, detailed image annotation; and (4) the rapid changes in imaging technology (*e.g.* 2D to 3D mammography tomosynthesis) that makes a multiyear annotation effect obsolete due to major technological improvements in imaging modalities. All of these challenges must be addressed before machine learning can replicate the work of a radiologist.

What are the technical details underlying challenges in object recognition and identification of abnormalities on diagnostic imaging studies?

First, it is true that computers with deep learning algorithms have approached human levels of performance in object recognition – as demonstrated in the Stanford ImageNet Large Scale Visual Recognition Competition (ILSVRC).[9] However, object recognition is a necessary but not sufficient prerequisite to performing this task on medical imaging studies. The set of validation images used in the ILSVRC are characterized by lower resolution, fewer classes and instances of objects per image, and larger objects, as compared to those features on the typical medical image used for diagnostic purposes. Stated another way, the task of object recognition on medical images is far more difficult because the objects (*i.e.* imaging findings) are more numerous, more varied, and far more complex than those on the standard test images for the ILSVRC. The issues of greater resolution, increased frequency of objects per unit space, and wider variety of object shapes and characteristics on medical images together pose a far greater challenge for computer-based object recognition than those posed by simple recognition of discrete objects. Medical evaluation of imaging findings typically requires analysis of multiple features, requiring several levels of analysis beyond object detection and classification (extending beyond the classic visual task of discriminating "dog *vs* cat"). Unless this learning algorithm can be trained with hundreds or thousands of additional algorithms to distinguish varying features of a recognized object, it will not yield any useful information about such questions. In the medical imaging realm, many kinds of imaging pathology require detailed analysis of a combination of features, likely requiring a greater degree of testing and validation, as well as an ensemble of multiple narrow algorithms. However, we recognize that focused applications of deep learning to specific medical imaging problems have already been devised and evaluated, especially in the fields of cardiothoracic imaging and breast imaging.[30,31]

In order for an ML system to replicate fully the multifactorial nature of the radiologist's assessment of an image (for example, a chest radiograph), it will likely need to be trained not by a single large dataset (containing many disparate types of radiographic abnormalities), but by the presentation of multiple datasets that specifically reinforce the learning associated with each class of imaging abnormalities (such as cardiac, mediastinal, pulmonary, and osseous) as well as additional datasets with various important subclasses of imaging abnormalities (for example, congenital heart disease). The final aggregate of the multiple datasets for chest radiographic images will need to be extremely large and extensively annotated, in order to ensure that the computer's experience matches both the depth and breadth of the radiologist's knowledge. Of course, a less ambitious training approach could be devised to ascertain whether a radiograph is normal or abnormal for triage purposes, but this approach would not replicate the bandwidth and detailed accuracy of expert performance.

Another major problem is the establishment of a gold standard. For example, within a large dataset of chest radiographs in patients suspected to have tuberculosis, there may be variability among several clinical radiologists in image interpretation. In clinical practice, one individual radiologist may want to not miss a case of tuberculosis due to its high clinical impact and thus would annotate cases as positive with subtle/non-specific findings of TB, while another radiologist may not want to overcall tuberculosis and may instead look for the more classical signs specific to the disease. Thus, when creating a predictive machine learning model, does one attempt to create different radiologist "personas" (*e.g.* high sensitivity *vs* high specificity profiles), or predict what a specific radiologist will report, or somehow create a middle-of-the-road report or "consensus" report? Alternatively, does the annotation of the final outcome of an imaging study get labeled as the actual sputum lab result or the actual clinical outcome? If so, then cases that are obviously normal or obviously strongly suggestive of TB will be labeled differently due to the clinical outcome. (In a recent academic study on this topic, the combination of sputum results, original radiologist interpretations, and confirmation by a single overreading, expert radiologist was required for inclusion into the pulmonary TB database.)[30] Finally, is the task to predict how a specific radiologist performs or how an "average" radiologist performs in interpretation of a radiograph or in prediction of the clinical outcome? If the goal is to predict clinical outcome, then issues such as prevalence of disease in a particular population may weight too heavily on the performance of the system. All of these questions raise important, clinically relevant issues that have not yet been resolved.

In machine learning, the computer's greatest strength - its abilities to process data endlessly and to repeat the same steps without tiring – could also represent a type of Achilles' heel. This problem is due to the issue of overfitting- defined as the functioning of a learning model (or prediction model) that fits so well with its training dataset to the extent that it models the statistical noise, fluctuations, biases, and errors inherent in the dataset, negatively impacting the performance on new data (*i.e.* diagnostic imaging studies not previously presented). This is more likely to occur in medical imaging than in other computer vision applications due to the relatively large number of categories of normal and abnormal findings and limited numbers of annotated training sets. More succinctly, Domingos indicates that overfitting has occurred "when your learner outputs a classifier that is 100% accurate on the training data but only 50% accurate on test data, when in fact it could have output one that is 75% accurate on both".[22] While the notion of accuracy in machine learning was relatively simple in the reported studies of object recognition, we note that radiology has a rich scientific history of measurement of diagnostic accuracy, including the development of receiver-operating characteristic (ROC) analysis.[32–34]

Classifier performance is central to making informed decisions about machine learning, and yet the typical use of a single measure of diagnostic accuracy, while simple, is inadequate for technical evaluation. Publications of medical machine learning studies are much more informative and rigorous when they utilize ROC analysis because its measures of sensitivity and specificity are not dependent on prevalence of disease (as is true of accuracy). In addition, the measure of diagnostic accuracy is typically derived from use of a single arbitrary threshold, whereas ROC analysis demonstrates the performance using all known threshold values. However, since the prevalence of a disease does affect the performance of any diagnostic classifier, it would also be helpful to know the prevalence of the disease in the test population, so that the false-positive and false-negative rates could be determined. Precision, which roughly translates as the likelihood that a positive test means that the disease or finding is truly present (otherwise known as the positive predictive value), can demonstrate the relative strength or weakness in a classifier for findings or diseases that are low prevalence.[26,35]

The problem of overfitting in medical imaging is also magnified by the wide variety of "odd" shapes of normal structures, and the myriads of anatomical variants related to extra or missing anatomical structures (such as accessory ossicles or congenitally absent or hypoplastic structures). This problem is made most evident by considering the problems faced by a radiology researcher who is collecting and classifying the many types of anatomical structures and abnormalities that are found on chest radiography. That researcher would have to obtain images and related data for the computer to demonstrate abnormalities of the heart, mediastinum, lungs, bones, pleura, and various other structures. Distinguishing anatomical variants from pathological entities has been an important function of the practicing radiologist, with a whole atlas devoted to helping them avoid making a false-positive diagnosis.[36] In other scientific fields, such as the field of genomics, there has been recognition of the unacceptably high "false-positive" rate associated with various kinds of "wild-type" variations that mimic findings associated with genetic mutations associated with cancer.[37] In one study of ML algorithms devoted to this problem, they characterized the types of false-positive errors into six different groups and suggested that "feature-based analysis of 'negative' or wild-type positions can be helpful to guide future developments in software".[37] This is akin to the problem with anatomical variants in diagnostic radiology.

Because the deep learning approach is highly complex, and because no method has been developed that allows a given algorithm to "explain" its reasoning, technology experts are generally

not able to understand fully the reasons for the algorithm's conclusions, and not able to predict the occurrence and frequency of failure or error in performance of the algorithm.[38] Therefore, validation and regulatory approval could take more time due to the "black box" nature of machine learning approaches. Fortunately, major advances have been made in recent years in illuminating the contents of the CNN black box.[39] One such advance, saliency maps, was originally proposed in 1998, and is based on the "feature-integration theory" of human visual attention.[40] In 2013, two image visualization techniques for visualization inside deep convolutional networks were demonstrated, one of which involved saliency maps.[41] For a given output category value (*e.g.* a type of interstitial lung disease), saliency maps display the pixels of the image (*e.g.* CT of the thorax) that were most important for image classification. More recently, other more sophisticated techniques have been developed that organize non-human interpretable convolution layers into an explanatory and potentially interactive graph or image that can be used to speed up the learning process and identify inaccuracies or important areas of an image ignored by a CNN allowing refinement of the model and improving performance.[39,42]

In contrast, CAD algorithms have been developed over several decades, many of which are focused on specific clinical imaging problems, and therefore have relatively narrow imaging applications. Examples of these applications include: (1) fracture detection, bone age determination, and bone mineral density quantitation in orthopedic radiology; (2) brain hemorrhage detection, multiple sclerosis detection and quantitation, and regional brain segmentation and volumetry in neuroradiology; and (3) coronary and/or carotid artery stenosis evaluation, and cardiac function assessment in cardiovascular radiology. In order for an ML system to replicate the performance of a radiologist, it would have to incorporate large portfolios of narrow ML algorithms, each of which has been devised to answer a specific clinical question. The use of combinations of algorithms to solve a single narrow machine learning problem or problems has been referred to as ensemble methods in machine learning and has been successful in winning machine learning competitions on classification of complex datasets. Yet the integration and orchestration of such a wide and varied array of learning algorithms - possibly from several different developers–into a single clinical system would likely require substantial amounts of time and effort in validation and testing (according to the "no free lunch" theorem of ensemble learning),[43] not to mention the potential regulatory challenges. In the field of artificial intelligence, the "holy grail" is to devise a form of "general artificial intelligence", which could replicate average human intelligence. General artificial intelligence, as opposed to a collection of narrow artificial intelligences, could help overcome this technological hurdle. Unfortunately, the majority of computer scientists do not believe that generalized artificial intelligence will emerge in the next 20 years, if ever. However, there are other ways that narrow artificial intelligence can help to improve the radiology work process, aside from diagnostic interpretation. There is a wide range of opportunities to increase operational efficiency, improve the radiology workflow, and provide decision support to clinicians and radiologists.

Is it likely that the job of the practicing radiologist is going to be completely displaced by artificial intelligence in the near future?

Acemoglu and Autor devised a 3 × 2 × 2 matrix model by which "work" can be classified, according to whether it is based upon (1) low, medium, or high skills; (2) cognitive or manual labor; and (3) routine or non-routine tasks.[44] Based upon their analysis, they found two interesting results relevant to a radiologist. First, the rapid diffusion of new technologies which substitute capital for labor – such as computerization - resulted in decreased demand for work based upon routine tasks. This effect was present whether the work is cognitive or manual, but was predominantly found among workers with medium-skill levels. Interestingly, the types of workers found to be more resistant to job displacement included financial analysts (a non-routine, cognitive job) and hairdressers (a non-routine, manual job). (We do note that the asset management industry is devoting substantial economic resources – even more than that devoted to radiology - to incorporate artificial intelligence into financial analysis.[45]) With respect to routine interpretation tasks performed by the practicing radiologist, it is likely that an ML system will soon perform some of the routine image interpretation tasks (for example, lung nodule screening or pre-operative chest radiography). However, many of radiologists' highly skilled work tasks, especially in complex image pattern recognition, will be more difficult to replicate over at least the next two decades, and therefore will require more time for adequate dataset generation and training, validation, and performance testing. This suggests that those radiologists who have acquired higher levels of skills (such as higher degrees of subspecialization, or greater experience in narrow, focused areas of clinical imaging) would be even more resistant to job displacement. Second, "technical change that makes highly skilled workers uniformly more productive" results in a lowering of the threshold for task difficulty that separates the medium-skill worker and the high-skill worker.[44] Therefore, in the face of potential displacement of radiologists from some image interpretation tasks, many radiologists will increasingly spend a higher percentage of time on other valuable radiology-based tasks. These radiology-based tasks include those listed in the ACR 3.0 Initiative, such as: consultation with referring physicians; timely oversight of ongoing complex imaging studies; direct patient contact including discussion about test results; verification of adherence to national imaging guidelines for proper test ordering; participation and data collection for radiology quality initiatives; and timely review of radiology-based patient outcomes.[46] The potential shift in the proportion of imaging interpretation activities in the daily work of the radiologist is also in keeping with the findings of the 2017 McKinsey report on the effects of automation on employment and productivity. While over half of all occupations have at least 30% work activities that could be automated, no more than 5% of all occupations could be entirely automated; this indicates that far more jobs will change than will be eliminated by automation.[47] In particular, the report states "high-skill workers who work closely with technology will likely be in strong demand, and may be able to take advantage of new opportunities for independent work." For radiologists, this could potentially include renewed focus on the entire spectrum of patient care in imaging. It is likely that new kinds of jobs for

radiologists will arise as a result of machine learning, similar to the way online retail activities led to both a decreased need for marketers and sales staff, and a tremendously increased demand for data scientists able to perform the data-mining activities needed to assess consumer wants and satisfaction.

For some expert radiologists, particularly those situated along the frontiers of their radiological subspecialties, there is also the possibility of being involved with a higher proportion of non-routine clinical work, including the interpretation of more complex imaging technologies that are found to be much more difficult to encode into an ML system. It is far less likely that sufficiently large datasets could be generated to provide neural networks the "experience" to answer questions about less common clinico-pathological entities, or to deal with non-routine clinical issues that often arise in medical practice. Therefore, there will remain an important role for the expert radiologist who can deal with the non-routine clinical work. This viewpoint is expressed by two experts in information systems and economics: "While computer reasoning from predefined rules and inferences from existing examples can address a large share of cases, human diagnosticians will still be valuable even after Dr Watson finishes its medical training because of the idiosyncrasies and special cases that inevitably arise. Just as it is much harder to create a 100 percent self-driving car than one that merely drives in normal conditions on a highway, creating a machine-based system for covering all possible medical cases is radically more difficult than building one for the most common situations".[48]

Finally, there is significant uncertainty as to whether or not certification by governmental regulatory agencies would initially allow these systems to operate autonomously, as opposed to requiring oversight by human radiologists. Similar to the steps established for CAD in mammography almost two decades ago, we believe that ML systems will, for the foreseeable future, be approved only for adjunctive use with radiologist oversight, over which time it could become the norm for machines and humans to work together in imaging study interpretation. At first, this may manifest as "worklist triage" in which cases suspected to be more likely to be abnormal by an ML algorithm will be prioritized for human interpretation.

Obtaining regulatory (FDA) clearance will continue to be an arduous process during the initial introductory phase of ML systems into the clinical care environment, because of all the intricate details involved in validation and approval of a plethora of ML systems. The FDA will likely need greater time, resources, and expertise to evaluate a completely different kind of imaging-based technology, and to understand the ramifications of a system wherein the underlying work processes–the learning algorithms themselves – are relatively opaque (*i.e.* a "black box"). Even after FDA approval, user acceptance of ML systems could be adversely affected if systemic errors or deviations are detected that cannot be explained. This suggests that post-market surveillance could become a more important feature with these systems.

The FDA's "Clinical and Patient Decision Support Software draft guidance" issued in December 2017 exempts software that provides decision support that merely makes it easier to perform simple calculations or retrieval of accessible data. However, deep learning applications are thought to be "black box" and thus must be FDA regulated.[49] In response to these challenges, the FDA has recently begun to make significant strides toward making the clearance process less onerous. One FDA guidance draft document, "Expansion of the Abbreviated 510(k) Program: Demonstrating Substantial Equivalence through Performance Criteria"[50] makes 510(k) clearance easier by allowing manufacturers to establish "substantial equivalence" functionally using performance metrics rather than requiring direct comparison testing and the same technology.[51] A few companies have managed to obtain FDA clearance for their deep learning-based algorithms related to diagnostic imaging and diagnostic testing. Arterys (San Francisco, CA) was the first company to receive clearance by the FDA for a deep learning application (for a suite of oncology software for automated segmentation of solid tumors on liver CT and MRI scans, and lung CT scans), thereby setting a precedent for other applications using CNNs.[52] Also, as of August 2018, the FDA has recently approved clinical decision support software for alerting providers of a potential stroke in patients,[53] an algorithm for detection of wrist fractures,[54] and an AI-based device to detect certain diabetes-related eye problems.[55]

What is the likely pathway of incorporating machine learning into radiology practice?

Even if the use of machine learning technology throughout society continues to increase exponentially, it is not at all clear that ML algorithms in a relatively well-defined field such as medical imaging will necessarily experience such astronomical growth. Advances in computational speed may only guarantee that the same answer–including the wrong answer–could be provided 1000 times faster, unless there are new techniques or new insights that emerge with approaches to deeper neural networks or future approaches such as Bayesian deep learning networks. Currently, machine learning for various image recognition algorithms requires presentation of many well-annotated imaging studies by human researchers, who then periodically test each algorithm for reliability and accuracy. Large imaging datasets will need to be developed and shared across institutions and radiology practices; this is an activity that requires work and trust to overcome technological, institutional, and regulatory barriers. The longstanding requirements of the medical field for high levels of diagnostic accuracy (as measured by sensitivity and specificity) and precision in differential diagnosis, will likely serve both as important benchmarks by which to judge the usefulness of these computer-aided diagnostic algorithms, and as essential "brakes" to the otherwise headlong rush to introduce labor-saving technology to reduce costs. The incorporation of these machine learning programs into the medical arena will likely be more gradual than in other sectors such as industrial, financial, chemistry, astronomy, etc., with a reasonable likelihood of a monotonic increase in the rate of progress over the years. Our healthcare system is a complex adaptive system and change in portions of that system–such as in the radiology industry–is typically characterized by "punctuated equilibrium"–*i.e.* relatively long periods of incremental change, interrupted by relatively short bursts of intense change.[56] Thus, the one caveat that

we make to our prediction of gradual incorporation of machine learning would be the advent of an earth-shattering technological innovation in generalized artificial intelligence–such as the invention of "the master algorithm," which is that universal learning algorithm that can be applied to disparate fields of knowledge, and yet still make robust, accurate predictions, when supplied with sufficient, appropriate data.[57] Only in that case would we suggest that machine learning has become a "10X force"–a change in the business force so large that it exceeds the usual competitive influences by an order of magnitude.[58] This "sea-change" would then motivate radiologists to prepare for an upcoming "strategic inflection point"–that point in time when the old ways of doing business and competing in the marketplace are no longer favored, and a new strategic paradigm takes over.[58] However, the history of science indicates that the timing of such an invention cannot be predicted in advance and likely will not occur any time soon.

In order to support their cognitive processes, current practicing radiologists have already learned to incorporate various kinds of technology, including quantitative analysis, three-dimensional imaging display tools, collaborative tools for consultation, and digital imaging resources. Future AI tools hold the promise of further expanding the work that radiologists can do, including in the realms of precision (personalized) medicine and population management. Rather than replacing radiologists, future AI tools could advance the kind of work that radiologists perform; this would be in line with the classic IBM Pollyanna Principle: "Machines should work; humans should think."[59,60] At the 2016 meeting of the Radiological Society of North America (RSNA), Keith Dreyer proposed that the future model of the radiologist is the "centaur diagnostician"; such a physician would team up with the ML system to optimize patient care.[61] This idea follows the observation that the performance of human-machine teams in playing chess could exceed that of a human or a machine system alone.[62] This partnership would yield greater precision and detail in their imaging-based report, including more quantitative information and evidence-based recommendations.[61] In addition, this could help facilitate advanced visualization techniques, refine clinical-radiological work procedures, and improve the timeliness and quality in communication between the radiologist and referring physician, as well as between the radiologist and patient. By viewing ML systems as a collaborator, not as a competitor, future radiologists could benefit from a partnership where the combined performance of the radiologist-computer team would likely be superior to either one alone, and feel enriched by the "luxury" of working with the advanced technological support offered by machine learning. In addition, the computer could allow the human to do more of what he or she does best – such as judicious use of the cognitive abilities associated with curiosity, experimentation, and insight. Just as in the example of Advanced Chess, it seems likely that the ability to work effectively with the computer will become a distinct competitive advantage. The futurist Kevin Kelly suggests that we cannot race against the machines, but that we can race with the machines. His conclusion is even more succinct: "You'll be paid in the future based on how well you work with robots".[63] This whole concept is also being embraced in various industries, as well as in medicine, including the explicit re-definition of "AI" by the American Medical Association as standing for "augmented intelligence" rather than "artificial intelligence".[64]

We believe it likely that machine-based learning systems will need oversight for a great many years because of the potential for many different kinds of errors on various kinds of imaging studies. In addition, most current medical imaging algorithms are not equipped with the basic knowledge and skills in human anatomy, physiology, and pathology. If we do reach a point when we might expect that machine-based systems approach the accuracy and reliability of a practicing radiologist, then it will become a societal issue as to whether or not diagnoses based solely upon machine learning are acceptable. If this is viewed solely as a technological upgrade, and if society has already accepted other innovations such as self-driving cars, then this change may not be controversial. On the other hand, if there is significant adverse public reaction to the loss of human interaction in the realm of medicine, then it is possible that radiologists may not be displaced for a very long time, if at all. Along these lines, Verghese et al have issued a strong call for the computer and the physician to be working together for the foreseeable future and have given a warning about the unintended consequences of the implementation of new technology.[65]

Given the expected retirement of increasing numbers of baby-boomer radiologists over the next two decades and the growing emphasis on screening and maintenance of health, it is likely that there will be a need for more radiologists over the next 20 years, and that computers will increasingly be regarded by those radiologists as trusted partners. The ML systems will be able to help create preliminary reports and note additional findings that may not make it into the final report, but, as is true of CAD today, computers would not be primarily responsible for the final reports. There will be a requirement for much more academic work to be done by human radiologists, including knowledge sharing and transfer learning, even before reaching the stage where the machine-based learning programs can become true partners in the imaging interpretation process.[66,67] Over the last few years, the RSNA R&E Foundation has received an increasing number of submissions of research and education grant applications (1 in 2015, 3 in 2016, 9 in 2017, and 27 in 2018) which involve the development of artificial intelligence in radiology, including machine learning.[68] In addition to educational offerings at various universities in the US and around the world (whether as part of degree-granting programs, certificate-based programs, or online training), there are also several developmental opportunities for physicians (whether internships or jobs) at various technology-based corporations in the US. The involvement of radiologists in machine-based learning in radiology will be critical in assuring that the care of future patients is not compromised by errors of commission or omission. While not yet part of the radiology curriculum for trainees, it is not hard to imagine that training in radiology informatics is likely to become an even more central component of radiology residency education. The first step has been taken by organized radiology with the development of a specific training program in radiology informatics geared towards fourth-year radiology

residents that was funded by the Association of University Radiologists (AUR) and is co-sponsored by RSNA and the Society for Imaging Informatics in Medicine (SIIM).[69] In the US, curricular issues in radiology education are still governed by the American Board of Radiology (ABR), but educational initiatives to incorporate informatics training for all radiology trainees are likely to be in line with future developments in diagnostic imaging. Current practicing radiologists will also need to be proactive in ensuring that they are full partners in these endeavors, rather than serving as "hand-maidens" to the other investigators who "just want their images labeled". For those individuals wishing to learn more about machine learning without having to abandon busy clinical careers for any length of time, there are several recommended online courses offered by various academic institutions (including Stanford, MIT, and Columbia) and by certain corporate entities (including Google and Nvidia) that have been available to the public at no charge.[70] Academic medical centers and other radiology organizations will need to provide environments where radiologists, machine learning experts, and other computer scientists can interact on a continual basis. As Davenport and Dreyer point out: "If the predicted improvements in deep learning image analysis are realized, then providers, patients, and payers will gravitate toward the radiologists who have figured out how to work effectively alongside AI".[71] Along those lines, we find that the creation of the ACR Data Science Institute is a strong indication that radiology organizations have recognized machine learning as a potential disruptive technology and are getting prepared to respond to this threat by investing resources to help develop, adapt, and deploy this new technology in the radiology workspace over the coming years.[72] In addition, several radiology-based organizations have started collaborations with major technology companies to develop ML algorithms and platforms.[73–75] We believe that this is just the beginning of a major trend in radiology, and that it behooves radiologists to participate in such endeavors for the betterment of radiology practice and the welfare of the patients that we serve.

CONCLUSION

We agree that machine learning will continue to make major advances in radiology over the next 5 to 10 years, but we completely disagree that there is any real possibility that radiologists will be replaced in that time frame, or even during the careers of our current trainees. In spite of all the advances of machine learning in the fields of self-driving cars, robotic surgery, and language translation, we believe that the work performed by radiologists is more complex than is thought by non-radiologists, and therefore more difficult to replicate by machine learning. The emergence of deep learning algorithms will help radiologists broaden the kinds of activities that establish their value in clinical care (such as routinely providing quantitative analysis), and to enhance the proportion of cognitive work (*e.g.* formulation of diagnosis) relative to visual search work (*e.g.* detection of imaging abnormalities). Imaging modalities will increasingly utilize deep learning to reduce image noise and enhance image quality overall. Since the potential for disruption of the radiology industry by machine learning does remain latent, it would be wise for various radiology organizations–especially academic institutions - to participate in research and development of this technology, and not leave the arena solely to corporate entities in the information technology sphere. While the economic environment of healthcare will continue to bring change to the practice of medicine and radiology, we believe that machine learning will not bring about the imminent doom of the radiologist. Instead, we foresee an intellectually vibrant future in which radiologists will continue to thrive professionally and benefit substantially from increasingly sophisticated and useful ML systems over the next few decades. Therefore, we would certainly encourage medical students and others interested in radiology as a profession–especially those with expertise in computer science–to pursue, enjoy, and look forward to a long career in diagnostic radiology, nuclear medicine, molecular imaging, and/or interventional radiology. This would provide benefits not only for the practitioners of diagnostic radiology, but even more importantly for our patients and for society.

REFERENCES

1. Lewis-Kraus G. The great A.I. awakening. 2016. Available from: https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html?_r=0.
2. Muhkerjee S. A.I. versus M.D. What happens when diagnosis is automated? 2017. Available from: http://www.newyorker.com/magazine/2017/04/03/ai-versus-md?utm_term=0_e44ef774d4-947de57fec-92238377&mc_cid=947de57fec&mc_eid=6e3f8f4920&utm_content=buffera21af&utm_medium=social&utm_source=linkedin.com&utm_campaign=buffer [Published on April 3, 2017. Accessed in pre-publication on March 30, 2017].
3. Christiansen CM. When new technologies cause great firms to fail. In: *The innovator's dilemma*: Boston; 1997.
4. Christiansen CM, Raynor ME, McDonald R. What is disruptive innovation? *Harvard Bus Rev* 2015;.
5. Chan S. Strategy development for anticipating and handling a disruptive technology. *J Am Coll Radiol* 2006; **3**: 778–86. doi: https://doi.org/10.1016/j.jacr.2006.03.014
6. Chockley K, Emanuel E. The end of radiology? Three threats to the future practice of radiology. *J Am Coll Radiol* 2016; **13**(12 Pt A): 1415–20. doi: https://doi.org/10.1016/j.jacr.2016.07.010
7. Dargan R. Not so elementary: experts debate the takeover of radiology by machines. 2017. Available from: http://www.rsna.org/News.aspx?id=21017 [PublishedJanuary 1. Accessed January 2, 2017].
8. Video from RSNA 2016: AI and radiology -- separating hope from hype. 2016. Available from: https://www.youtube.com/watch?v=4b0t7TzjZRE [Published November 29, 2016. Accessed August 26, 2018].
9. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vis* 2015; **115**: 211–. doi: https://doi.org/10.1007/s11263-015-0816-y
10. Koch C. How the computer beat the Go master. Scientific American. 2016. Available from: https://www.scientificamerican.com/article/how-the-computer-beat-the-gomaster/. [Accessed May 5, 2018].
11. Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, et al.

Mastering the game of Go without human knowledge. *Nature* 2017; **550**: 354–9. doi: https://doi.org/10.1038/nature24270
12. Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology* 2001; **220**: 781–6. doi: https://doi.org/10.1148/radiol.2203001282
13. Birdwell RL. The preponderance of evidence supports computer-aided detection for screening mammography. *Radiology* 2009; **253**: 9–16. doi: https://doi.org/10.1148/radiol.2531090611
14. Noble M, Bruening W, Uhl S, Schoelles K. Computer-aided detection mammography for breast cancer screening: systematic review and meta-analysis. *Arch Gynecol Obstet* 2009; **279**: 881–90. doi: https://doi.org/10.1007/s00404-008-0841-y
15. Rao VM, Levin DC, Parker L, Cavanaugh B, Frangos AJ, Sunshine JH. How widely is computer-aided detection used in screening and diagnostic mammography? *J Am Coll Radiol* 2010; **7**: 802–5. doi: https://doi.org/10.1016/j.jacr.2010.05.019
16. Fenton JJ, Abraham L, Taplin SH, Geller BM, Carney PA, D'Orsi C, et al. Effectiveness of computer-aided detection in community mammography practice. *J Natl Cancer Inst* 2011; **103**: 1152–61. doi: https://doi.org/10.1093/jnci/djr206
17. Philpotts LE. Can computer-aided detection be detrimental to mammographic interpretation? *Radiology* 2009; **253**: 17–22. doi: https://doi.org/10.1148/radiol.2531090689
18. Mezrich JL, Siegel EL. Legal ramifications of computer-aided detection in mammography. *J Am Coll Radiol* 2015; **12**: 572–4. doi: https://doi.org/10.1016/j.jacr.2014.10.025
19. Jackson WL, Jackson WL. *The state of CAD for mammography*. 2014. Available from: Imaging.http://www.diagnosticimaging.com/cad/state-cad-mammography. [PublishedFebruary 18. Accessed February 4, 2017].
20. Bahl M, Barzilay R, Yedidia AB, Locascio NJ, Yu L, Lehman CD. High-risk breast lesions: a machine learning model to predict pathologic upgrade and reduce unnecessary surgical excision. *Radiology* 2018; **286**: 810–8. doi: https://doi.org/10.1148/radiol.2017170549
21. Masarie FE, Miller RA, Myers JD. INTERNIST-I properties: representing common sense and good medical practice in a computerized medical knowledge base. *Comput Biomed Res* 1985; **18**: 458–79. doi: https://doi.org/10.1016/0010-4809(85)90022-9
22. Domingos P. A few useful things to know about machine learning. *Commun ACM* 2012; **55**: 78–87. doi: https://doi.org/10.1145/2347736.2347755
23. Erickson BJ, Korfiatis P, Kline TL, Akkus Z, Philbrick K, Weston AD. Deep Learning in Radiology: Does One Size Fit All? *J Am Coll Radiol* 2018; **15**(3 Pt B): 521–6. doi: https://doi.org/10.1016/j.jacr.2017.12.027
24. Tanz J. *Soon we won't program computers. We'll train them like dogs*. 2016. Available from: https://www.wired.com/2016/05/the-end-of-code [Published May 17, 2016. Accessed August 25, 2018].
25. Ruder S. Transfer learning – machine learning's next frontier. 2017. Available from: http://ruder.io/transfer-learning [Published March 21, 2017. Accessed August 25, 2018].
26. Benjamini Y Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 1995; **57**: 289–300.
27. NIH Clinical Center. NIH Clinical Center provides one of the largest publicly available chest x-ray datasets to scientific community. 20172018. Available from: https://www.nih.gov/news-events/news-releases/nih-clinical-center-provides-one-largest-publicly-available-chest-x-ray-datasets-scientific-community [Published September 27, 2017. Accessed October 4, 2018].
28. The Cancer Imaging Archive (TICA). TICA Collections. 2018. Available from: http://www.cancerimagingarchive.net/ [Accessed October 4, 2018].
29. Kohli MD, Summers RM, Geis JR. Medical Image Data and Datasets in the Era of Machine Learning-Whitepaper from the 2016 C-MIMI Meeting Dataset Session. *J Digit Imaging* 2017; **30**: 392–9. doi: https://doi.org/10.1007/s10278-017-9976-3
30. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 2017; **284**: 574–82. doi: https://doi.org/10.1148/radiol.2017162326
31. McBee MP, Awan OA, Colucci AT, Ghobadi CW, Kansagra AP. Deep learning in radiology. Academic Radiology 2018 in press. 2018. Available from: https://doi.org/10.106/j.acra.2018.02.018 [Accessed August 25, 2018].
32. Lusted LB. Signal detectability and medical decision-making. *Science* 1971; **171**: 1217–9. doi: https://doi.org/10.1126/science.171.3977.1217
33. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* 1978; **8**: 283–98. doi: https://doi.org/10.1016/S0001-2998(78)80014-2
34. McNeil BJ, Hanley JA. Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. *Med Decis Making* 1984; **4**: 137–50. doi: https://doi.org/10.1177/0272989X8400400203
35. Oakden-Rayner L. The philosophical argument for using ROC curves. 2018. Available from: https://lukeoakdenrayner.wordpress.com/2018/01/07/the-philosophical-argumentfor-using-roc-curves [Published January 7, 2018. Accessed August 25, 2018].
36. Keats TE, Anderson MW.Elsevier, ed. *Atlas of normal roentgen variants that may simulate disease*. 9th ed. Philadelphia, PA: Elsevier; 2013.
37. Ding J, Bashashati A, Roth A, Oloumi A, Tse K, Zeng T, et al. Feature-based classifiers for somatic mutation detection in tumour-normal paired sequencing data. *Bioinformatics* 2012; **28**: 167–75. doi: https://doi.org/10.1093/bioinformatics/btr629
38. Knight W. *The dark secret at the heart of AI. MIT Technology Review*. 2018. Available from: https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ [Accessed August 26, 2018].
39. Zhang Q, Zhu SC. Visual interpretability for deep learning. 2018. Available from: https://arvix.org/abs/1802.00614 [Published February 2, 2018. Accessed August 26, 2018].
40. Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1998; **20**: 1254–9. doi: https://doi.org/10.1109/34.730558
41. Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: visualizing image classification models and saliency maps. 2013. Available from: https://arvix.org/abs/312.6034 [Published December 20. Accessed August 26, 2018].
42. Zhang Q, Cao F, Shi F, YN W, Zhu SC. Interpreting CNN knowledge via an explanatory graph. 2017. Available from: https://arxiv.org/pdf/1708.01785 [Published November 22. Accessed October 4, 2018].
43. Polikar R. Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine* 2015; **6**: 21–45. doi: https://doi.org/10.1109/MCAS.2006.1688199
44. Acemoglu D, Autor D. Skills, tasks and technologies: implications for employment and earnings. Working paper, National Bureau of Economic Research. 2017. Available from: http://www.nber.org/papers/w16082 [Issued June 2010. Accessed January 2, 2017].

45. Kim C. How will AI manage your money? *Barron's – Lipper Mutual Fund Quarterly* 2018; L7–L12.
46. McGinty G, Allen Jr B, Wald C. Imaging 3.0TM. American College ofRadiology. 2013. Available from: https://www.acr.org/~/media/ACR/Documents/PDF/Advocacy/IT%20Reference%20Guide/IT%20Ref%20Guide%20Imaging3.pdf [Accessed January 2, 2017].
47. McKinsey Global Institute. A future that works: automation, employment, and productivity. January 2017. Available from: http://www.mckinsey.com/global-themes/digital-disruption/harnessing-automation-for-a-future-that-works [Accessed July 2, 2017].
48. Brynjolfson E, McAfee A. Work, progress, and prosperity in a time of brilliant technologies. In: *The second machine age*. New York: Norton; 2014. pp. 192–3.
49. Edwards B. FDA guidance on clinical decisions: Peering inside the black box of algorithmic intelligence. 2017. Available from: https://www.chilmarkresearch.com/fda-guidance-clinical-decision-support/ [Published December 19, 2017. Accessed August 26, 2018].
50. U.S. Food and Drug Administration. Expansion of the abbreviated 510(k) program: demonstrating substantial equivalence through performance criteria. Draft guidance for industry and Food and Drug Administration. 2018. Available from: https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM604195.pdf [Published April 12, 2018. AccessedAugust 26, 2018].
51. U.S. FoodDrug Administration. *In Brief: FDA to offer a voluntary, more modern 510(k) pathway for enabling moderate risk devices to more efficiently demonstrate safety and effectiveness*. 2018. Available from: https://www.fda.gov/NewsEvents/Newsroom/FDAInBrief/ucm604348.htm [Published April 12, 2018. Accessed August 26, 2018].
52. News FDA. Arterys wins FDA clearance for oncology imaging software. 2018. Available from: https://www.fdanews.com/articles/185673-arterys-wins-fda-clearance-for-oncology-imaging-software [Published February 20, 2018. Accessed August 26,2018].
53. U.S. Food and Drug Administration. FDA permits marketing of clinical decision support software for alerting providers of a potential stroke in patients. 2018. Available from: https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm596575.htm [Published February 13, 2018. Accessed August 26, 2018].
54. U.S. Food and Drug Administration. FDA permits marketing of artificial intelligence algorithm for aiding providers in detecting wrist fractures. 2018. Available from: https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm608833.htm [Accessed August 26].
55. U.S. Food and Drug Administration. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. 2018. Available from: https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm604357.htm [Published April 11, 2018. Accessed August 26, 2018].
56. Enzmann DR, Feinberg DT. The nature of change. *J Am Coll Radiol* 2014; **11**: 464–70. doi: https://doi.org/10.1016/j.jacr.2013.12.006
57. Domingos P. *The master algorithm; how the quest for the ultimate learning machine will remake our world. New York*. NY: Basic; 2015. pp. 25–6.
58. Grove AS. *Only the paranoid survive; how to exploit the crisis points that challenge every company*. New York, NY: Currency; 1996.
59. Bloch A. *Murphy's Law and other reasons why things go wrong. Los Angeles*. **41**. CA: Price/Stern/Sloan; 1979.
60. Henson J. The Paperwork Explosion. The Jim Henson Company (for IBM). 196720102018. Available from: www.youtube.com/watch?v=_IZw2CoYztk [Produced October 1967. PublishedMarch 30, 2010. Accessed August 26, 2018].
61. Dreyer K. When machines think: radiology's next frontier. In: *Presented at the 2016 annual meeting of the Radiological Society of North America*. Chicago, Ill; 2016.
62. Thompson C. In: Smarter than you think; how technology is changing our minds for the better. In: *The rise of the centaurs. NY*: Penguin; 2013. pp. 1–5.
63. Kelly K. *The inevitable; understanding the 12 technological forces that will shape our future*. New York, NY: Viking; 2016.
64. American Medical Association. AMA Passes First Policy Recommendations on Augmented Intelligence. 2018. Available from: https://www.ama-assn.org/ama-passes-first-policy-recommendations-augmented-intelligence [Published June 14, 2018. AccessedAugust 26, 2018].
65. Verghese A, Shah NH, Harrington RA. What This Computer Needs Is a Physician: Humanism and Artificial Intelligence. *JAMA* 2018; **319**: 19–20. doi: https://doi.org/10.1001/jama.2017.19198
66. Gunderman R, Chan S. Knowledge sharing in radiology. *Radiology* 2003; **229**: 314–7. doi: https://doi.org/10.1148/radiol.2292030030
67. Kohli M, Prevedello LM, Filice RW, Geis JR. Implementing machine learning in radiology practice and research. *AJR Am J Roentgenol* 2017; **208**: 754–60. doi: https://doi.org/10.2214/AJR.16.17224
68. Walter S. Private communication: unpublished observation on trends in RSNA R&E Foundation grants applications 2014-2018. February 2018
69. Cook T. https://sites.google.com/view/imaging-informatics-course/resources. 2017. Available from: https://sites.google.com/view/imaging-informatics-course/resources [Start date: October 2, 2017. Accessed August 26, 2018].
70. Marr B. The 6 best free online artificial intelligence courses for 2018. 2018. Available from: https://www.forbes.com/sites/bernardmarr/2018/04/16/the-6-best-free-online-artificial-intelligence-courses-for-2018/#6836668959d7 [Published April 16, 2018.Accessed August 26, 2018].
71. Davenport TH, Dreyer KJ. AI will change radiology, but it won't replace radiologists. 2018. Available from: https://hbr.org/2018/03/ai-will-change-radiology-but-it-wont-replace-radiologists [Rev March 27, 2018, Accessed April 19, 2018].
72. American College of Radiology (ACR) Data Science Institute. 2018. Available from: http://www.acrdsi.org/index.html [Accessed May 5, 2018.].
73. Healthcare P. Partners HealthCare and GE Healthcare launch 10-year collaboration to integrate Artificial Intelligence into every aspect of the patient journey. 2017. Available from: https://www.partners.org/Newsroom/Press-Releases/Partners-GE-Healthcare-Collaboration.aspx [Published May 17, 2017. Accessed August 26,2018].
74. Terdiman T. Facebook and NYU believe AI can make MRIs way, way faster. 2018. Available from: https://www.fastcompany.com/90220180/facebook-and-nyu-believe-ai-can-make-mris-way-way-faster [Published August 20, 2018. Accessed August 26, 2018].
75. Partners R. Radiology Partners to join the IBM Watson Health Medical Imaging Collaborative. 2017. Available from: http://www.radpartners.com/events/j4ss9xuz22/Radiology-Partners-to-join-the-IBM-Watson-Health-Medical-Imaging-Collaborative [Accessed August 26, 2018.].