

Will Machine Learning End the Viability of Radiology as a Thriving Medical Specialty?

Authors: Stephen Chan and Eliot L. Siegel

Corresponding Author:

Stephen Chan, MD, MBA, MPH

Department of Radiology

Harlem Hospital Center and Columbia University

506 Lenox Avenue

New York, NY 10037

Cell: 201-240-1902

Fax: 201-767-2341

e-mail: sc56md@gmail.com

Mailing address for correspondence: 29 Trautwein Crescent

Closter, NJ 07624

Co-Author:

Eliot L. Siegel, MD

Department of Diagnostic Radiology and Nuclear Medicine

VA Maryland Health Care System

Baltimore, MD

Presented in part at the 2016 meeting of the Radiological Society of North America

Article type: Review Article

Word Count: 8080 words (Introduction to Conclusion)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Is Machine Learning a Disruptive Technology That Will End the Viability of Radiology as a Thriving Medical Specialty?

Review Article

Short title: Is Machine Learning a Disruptive Technology for Radiologists?

BJR UNCORRECTED PROOFS

1
2
3
4
5 Abstract
6
7
8

9 There have been tremendous advances in artificial intelligence and machine learning
10 within the past decade, especially in the application of deep learning to various
11 challenges. These include advanced competitive games (such as Chess and Go), self-
12 driving cars, speech recognition, and intelligent personal assistants. Rapid advances in
13 computer vision for recognition of objects in pictures have led some individuals,
14 including computer science experts and health care system experts in machine
15 learning, to make predictions that machine learning algorithms will soon lead to the
16 replacement of the radiologist. However, there are complex technological, regulatory,
17 and medicolegal obstacles facing the implementation of machine learning in radiology
18 that will definitely preclude replacement of the radiologist by these algorithms within the
19 next two decades and beyond. While not a comprehensive review of machine learning,
20 this article is intended to highlight specific features of machine learning where it faces
21 significant technological and health care systems challenges. Rather than replacing
22 radiologists, machine learning will provide quantitative tools that will increase the value
23 of diagnostic imaging as a biomarker, increase image quality with decreased acquisition
24 times, and improve workflow, communication, and patient safety. In the foreseeable
25 future, we predict that today's generation of radiologists will be replaced not by machine
26 learning algorithms, but by a new breed of data science-savvy radiologists who have
27 embraced and harnessed the incredible potential that machine learning has to advance
28 our ability to care for our patients. In this way, radiology will remain a viable medical
29 specialty for years to come.
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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 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

1
2
3
4
5 leading radiology organizations have begun to adopt strategies for handling this
6
7
8 potentially disruptive technology [5].
9

10
11
12
13
14 In 2016, Chockley (a medical student) and Emanuel (an internal medicine physician and
15
16
17 “Obamacare” architect) identified three threats to the future practice of diagnostic
18
19 imaging, with machine learning singled out as the “ultimate threat” [6]. They made the
20
21 following two assertions: 1) “machine learning will become a powerful force in radiology
22
23 in the next 5 to 10 years, not in multiple decades”; and 2) “indeed, in a few years there
24
25 may [be] no specialty called radiology” [6]. If they meant that the computer will largely
26
27 replace the radiologist in 5 to 10 years (as implied in their work), then we completely
28
29 disagree and believe that this is an ill-informed prediction borne out of a lack of domain
30
31 knowledge of radiology. Their view reflects a fundamental misunderstanding of the
32
33 nature of the work performed by radiologists, as well as a lack of appreciation of exactly
34
35 how difficult it will be for machine learning to replace the wide variety of imaging
36
37 interpretation and patient care tasks inherent in the practice of radiology. We note that
38
39 the specter of a future in which radiologists are no longer needed to provide image
40
41 interpretation services has seriously alarmed forward-thinking medical students,
42
43 radiology residents and fellows, impelling some to ask if they should quit or avoid
44
45 radiology residency because of the risk of not getting a job after residency [7,8].
46
47
48
49
50
51
52

53 Indeed, that fear could potentially damage the radiology profession by discouraging
54
55 talented medical students from choosing radiology as their future career. We seek to
56
57
58 allay such fear by careful examination of the recent developments in machine learning,
59
60

1
2
3
4
5 and by detailed evaluation of the kind of technological development necessary to render
6
7 the broad range of radiologic diagnosis. Specifically, there are two fundamental
8
9 sources of misunderstanding that lead many individuals to conclude that radiologists
10
11 can be easily replaced by machine learning.
12
13
14
15
16
17
18

19 Misunderstanding #1: Machine learning can easily absorb and process the wide
20
21 variation of information and ambiguity inherent in interpretation of medical images.
22
23
24
25
26
27

28 Remarkable achievements have been made in machine learning such as the impressive
29
30 computer vision performance on identification of objects in everyday pictures from the
31
32 Stanford ImageNet challenge [9] and the victory of Google's AlphaGo over the 2016
33
34 human champion of Go [10]. Computer scientists cite these accomplishments to assert
35
36 that unsupervised machine learning will soon be rendering medical imaging findings and
37
38 diagnoses. However, board games such as "Go" focus on a very "narrow" artificial
39
40 intelligence task where a winning versus losing status can be assessed, whereas
41
42 medical imaging is associated with far greater amount of ambiguity, and a larger variety
43
44 of features, classifications, and outputs. It is also likely that thousands of "narrow"
45
46 algorithms based on separate large, well-annotated databases will be required for a
47
48 computer to begin to compete with a radiologist for comprehensive diagnostic
49
50 assessment of even a single modality covering a single anatomic region of the body.
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6 Advances in self-reinforcement learning have led to substantial further improvements in
7
8 “AlphaGo” resulting in “AlphaGoZero” which utilizes an approach in which the computer
9
10 is provided with the basic rules of the game and learns by playing itself large numbers
11
12 of games rather than learning by analyzing the play of human experts [11]. Although
13
14 possible in games with simple defined rules such as Go or chess, analogous self-
15
16 reinforcement learning is not so easily attainable in radiology, given the lack of a simple
17
18 set of rules of the “radiology game” to allow this sort of self-play. Barring an unforeseen
19
20 major technological breakthrough, it is likely that human annotation and guidance will
21
22 likely be necessary at multiple stages in the development of machine learning in
23
24 medical imaging, augmented by increases in computing power and conceptual
25
26 advances in artificial intelligence. This pattern is exemplified by the technological
27
28 development of the Google Translate app in which significant conceptual advances in
29
30 machine learning-based language translation needed to be made by computer
31
32 scientists, who then were able to render the sequential and contextual information
33
34 inherent in languages far more amenable to deep neural networks [1].
35
36
37
38
39
40
41
42
43
44
45

46 Misunderstanding #2: Computer-aided detection and computer-aided diagnosis is an
47
48 immediate technological precursor to machine learning algorithms.
49
50
51
52
53
54

55 Chockley and Emanuel cite the current performance of computer-aided detection
56
57 (CADe) and computer-aided diagnosis (CADx) in various areas of radiology - including
58
59
60
61

1
2
3
4
5 the field of mammography – as evidence of success stories, with machine learning
6
7
8 “working as well as or better than the experienced radiologist”. Indeed, many papers
9
10 and presentations describing CAD systems in mammography have claimed a
11
12 performance level in lesion detection similar to that of an experienced radiologist
13
14 [12,13,14]. Based on that research, CAD was approved by the Food and Drug
15
16 Administration (FDA) for use with mammography and has been widely introduced into
17
18 radiology practices across the U.S. as an adjunctive technology for mammography [15].
19
20 However, in spite of its widespread use for the past decade, it has not been shown to
21
22 improve detection rates in academic settings, and it is unclear whether or not CAD
23
24 improves the detection rate of invasive breast carcinoma in community practice [16]. In
25
26 addition, the use of CAD can be detrimental if its limitations are not understood [17].
27
28 While review of mammographic images with adjunctive CAD would likely be considered
29
30 the de facto standard of care in community mammography practice [18], we note that
31
32 CAD systems have not replaced the practicing radiologist. In practice, survey data
33
34 suggests that more than half (~62%) of radiologists have never or rarely changed their
35
36 report as a result of CAD findings in mammography, and about a third of radiologists
37
38 never or rarely use the findings generated by CAD [19]. There has been initial
39
40 demonstration of a machine learning tool to help separate high-risk breast lesions that
41
42 are more likely to become cancerous from those that are at lower risk [20]. However, we
43
44 know of no CAD program in clinical use that continually receives feedback about its
45
46 diagnostic performance – a task that is essential to learning from experience. Finally,
47
48 we are not aware of any mammography CAD/machine learning software program that
49
50 formally compares a prior mammogram to a current mammogram, just as a human
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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 inter-neurons of the human nervous system.

1
2
3
4
5
6 Convolutional neural networks (CNNs) are a special type of neural network that is
7
8 optimized for image pattern recognition [22]. Unlike other types of artificial neural
9
10 networks (ANNs), the majority of nodes (neurons) in a CNN are only connected to a
11
12 subset of other nodes, particularly those in closer proximity in an image which enhances
13
14 their ability to recognize local features of an image. In brief, a CNN consists of multiple
15
16 layers between the input and output layers [Figure 1]. The main building blocks are the
17
18 convolution layer which can be thought of a series of adjustable image filters that can
19
20 emphasize or de-emphasize certain aspects of an image such as borders, colors, noise,
21
22 and texture. Each of multiple convolution layers within a CNN is followed by a pooling
23
24 layer which serves to reduce the number of parameters. For example, the commonly
25
26 used technique referred to as “max pooling” simply chooses the maximum pixel value
27
28 within each small portion of an image and assigns all pixels to that value. The final
29
30 layer of a CNN is a fully connected one similar to other types of artificial neural
31
32 networks.
33
34
35
36
37
38
39
40
41
42
43

44 The term “AI” is currently commonly utilized in medical imaging in both the lay and
45
46 scientific literature to refer to machine learning (ML) in general and CNNs specifically.
47

48 Although the architecture of both CAD programs and ML-based algorithms are designed
49
50 by humans, the essential discriminatory functions of the AI algorithms emerge directly
51
52 from the data, and, unlike CAD, do not require humans to identify and compute the
53
54 critical features [23]. This emergence of algorithms from the data is what prompted
55
56

57
58 Wired magazine to suggest that machine learning may represent "the end of code" [24].
59
60

1
2
3
4
5 Specifically, the algorithms to predict such things as the presence of an intracranial
6 bleed, or malignancy in a prostate MRI study will emerge directly from the “learned”,
7
8 iteratively adjusted values of the nodes in a CNN. Those values themselves represent
9
10 the trained model and the “training” continues with the introduction of each annotated
11
12 dataset. Although inputs to CNNs are not always raw images (and may be segmented
13
14 and co-registered prior to classification), the many steps such as feature extraction,
15
16 segmentation, registration, and statistical analysis utilized by the previous generation of
17
18 so called CAD (Computer Aided Detection or Computer Aided Diagnosis) software are
19
20 not required. Both AI and CAD techniques can be utilized to develop medical imaging
21
22 software, but AI algorithms typically require much more annotated data but then
23
24 subsequently take much less time to develop using fewer steps. Most developers
25
26 previously utilizing these more human understandable basic steps have made the
27
28 transition to the use of CNNs. Creation of a machine learning model can be performed
29
30 much faster (e.g. 5 to 6 days rather than 5 to 6 months or years) than traditional
31
32 computer aided detection and diagnosis (CAD_e/CAD_x).
33
34
35
36
37
38
39
40
41
42
43
44
45

46 The nature of learning in machine learning can be confused with that associated with
47
48 humans. Machine learning has been defined as “algorithm-driven learning by
49
50 computers such that the machine’s performance improves with greater experience” and
51
52 indicated that it “involves the construction of algorithms that learn from data and make
53
54 predictions on the basis of those data [6].” Although this definition may imply a type of
55
56 self-reinforcement learning “with greater experience”, in actuality, these algorithms in
57
58
59
60
61

1
2
3
4
5
6 diagnostic imaging have improved largely by the addition of annotated data based on
7
8 human review or patient outcome information, rather than repetitive application of work-
9
10 in-progress. This is a fundamental difference between the use of machine learning in
11
12 strategy games such as Chess and Go in which there are well defined parameters of
13
14 success. The current regulatory constraint on progressive learning of these algorithms
15
16 is that clearance of these products has been based on a well-defined training and test
17
18 set. Within a clinical practice, there is not currently agreement by the FDA from a
19
20 regulatory perspective for the system to continue to improve using additional local
21
22 patient data as would be the case with a radiologist. As is true of other related
23
24 statistical techniques such as linear regression, additional “processed” data are required
25
26 to enhance the model rather than simply “learning from experience” per se. Therefore,
27
28 computers remain far less efficient than humans at learning and generalizing concepts
29
30 from a relatively small dataset. This has been mitigated somewhat by techniques such
31
32 as transfer learning which takes patterns learned from a related task as a starting point
33
34 effectively kickstarting the training (also known as “one-shot learning”) that takes
35
36 advantage of existing knowledge to train using just a single or very few examples [25].
37
38
39
40
41
42
43
44
45
46
47
48

49 Will a machine learning system soon be capable of replicating the work of a radiologist?

50
51 Major technological developments in machine learning have been made over the last
52
53 few years, including advances in deep learning algorithms, further advances in GPU
54
55 (Graphics Processing Units) speed and memory, and the exponential growth of
56
57 corporate investment. However, there are several independent factors which suggest
58
59
60

1
2
3
4
5
6 that successful replacement of the radiologist's work is likely to be substantially more
7
8 difficult than is currently envisioned by some non-radiologist health care experts and
9
10 computer science futurists. These potential challenges are based upon unique aspects
11
12 of the radiologic image, the visual processing capability of the radiologist, and the role of
13
14 the radiologist in maximizing and maintaining clinical relevance in image interpretation.
15
16 In particular, the job of the radiologist is not simply to detect findings related to various
17
18 imaging studies, but to determine "what is wrong with this picture" and help determine
19
20 the future course of action in the diagnostic evaluation and therapeutic decision-making.
21
22 Determining "what is wrong with this picture" is a much harder task that extends far
23
24 beyond the capabilities of the current generation of computer vision systems.
25
26 Contextualization of the imaging information in diagnostic evaluation and therapeutic
27
28 decision-making may be an even more difficult task to replicate.
29
30
31
32
33
34
35
36
37
38

39 A major challenge for machine learning algorithms is the greater technical complexity of
40
41 the radiologic image as compared to those images typically used in object-recognition
42
43 tests for computer "visualization". In addition to the differing imaging modalities, this
44
45 complexity includes a wide variety of manifestations of normal and pathological findings,
46
47 multiple sequential images in a cross sectional/volumetric dataset, with much higher
48
49 complexity of data and raw number of pixels/voxels in medical images. It also includes
50
51 a high level of ambiguity and difficulty in annotation that is not inherent in the ImageNet
52
53 challenges that have used common objects such as dogs, cats, bikes, cars, etc.
54
55
56
57

58 Another major technical challenge is the development of a "reasonable" detection rate
59
60
61

1
2
3
4
5 of abnormalities without an excessive rate of false-positive findings as compared with
6 human performance. For more than 20 years, CAD and CADx programs, such as those
7 used to detect lung nodules or breast masses, have been fraught with the issue of
8 frequent false-positive findings (i.e., low specificity) and we suggest that this problem
9 may also be an intrinsic problem for deep learning algorithms [22,26]. This problem is
10 further complicated by: 1) the multiple classes of imaging abnormalities detected on
11 diagnostic imaging studies; 2) the time and expense associated with the collection of
12 large annotated datasets (such as ChestX-ray14 and the Cancer Imaging Archive)
13 required for deep learning, of which a fair number are available in the public domain [27,
14 28], but many more are needed [29]; 3) the difficulties associated with ensuring
15 sufficient, detailed image annotation; and 4) the rapid changes in imaging technology
16 (e.g., 2D to 3D mammography tomosynthesis) that makes a multi-year annotation effect
17 obsolete due to major technological improvements in imaging modalities. All of these
18 challenges must be addressed before machine learning could replicate the work of a
19 radiologist.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

46 What are the technical details underlying challenges in object recognition and
47 identification of abnormalities on diagnostic imaging studies?
48
49

50
51
52 First, it is true that computers with deep learning algorithms have approached human
53 levels of performance in object recognition – as demonstrated in the Stanford ImageNet
54 Large Scale Visual Recognition Competition (ILSVRC) [9]. However, object recognition
55 is a necessary but not sufficient prerequisite to performing this task on medical imaging
56
57
58
59
60

1
2
3
4
5
6 studies. The set of validation images used in the ILSVRC are characterized by lower
7
8 resolution, fewer classes and instances of objects per image, and larger objects, as
9
10 compared to those features on the typical medical image used for diagnostic purposes.
11
12 Stated in another way, the task of object recognition on medical images is far more
13
14 difficult because the objects (i.e., imaging findings) are more numerous, more varied,
15
16 and far more complex than those on the standard test images for the ILSVRC. The
17
18 issues of greater resolution, increased frequency of objects per unit space, and wider
19
20 variety of object shapes and characteristics on medical images together pose a far
21
22 greater challenge for computer-based object recognition than those posed by simple
23
24 recognition of discrete objects. Medical evaluation of imaging findings typically requires
25
26 analysis of multiple features, requiring several levels of analysis beyond object detection
27
28 and classification (extending beyond the classic visual task of discriminating “dog
29
30 versus cat”). Unless this learning algorithm can be trained with hundreds or thousands
31
32 of additional algorithms to distinguish varying features of a recognized object, it will not
33
34 yield any useful information about such questions. In the medical imaging realm, many
35
36 kinds of imaging pathology require detailed analysis of a combination of features, likely
37
38 requiring a greater degree of testing and validation, as well as an ensemble of multiple
39
40 narrow algorithms. However, we recognize that focused applications of deep learning
41
42 to specific medical imaging problems have already been devised and evaluated,
43
44 especially in the fields of cardiothoracic imaging and breast imaging [30,31].
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6 In order for a machine learning system to replicate fully the multi-factorial nature of the
7
8 radiologist's assessment of an image, for example, a chest radiograph, it will likely need
9
10 to be trained not by a single large dataset (containing many disparate types of
11
12 radiographic abnormalities), but by the presentation of multiple datasets that specifically
13
14 reinforce the learning associated with each class of imaging abnormalities (such as
15
16 cardiac, mediastinal, pulmonary, and osseous) as well as additional datasets with
17
18 various important subclasses of imaging abnormalities (for example, congenital heart
19
20 disease). The final aggregate of the multiple datasets for chest radiographic images will
21
22 need to be extremely large and extensively annotated, in order to ensure that the
23
24 computer's experience matches both the depth and breadth of the radiologist's
25
26 knowledge. Of course, a less ambitious training approach could be devised to ascertain
27
28 whether a radiograph is normal or abnormal for triage purposes, but this approach
29
30 would not replicate the bandwidth and detailed accuracy of expert performance.
31
32
33
34
35
36
37
38
39
40

41 Another major problem is the establishment of a gold standard. For example, within a
42
43 large dataset of chest radiographs in patients suspected to have tuberculosis, there may
44
45 be variability among several clinical radiologists in image interpretation. In clinical
46
47 practice, one individual radiologist may want to not miss a case of tuberculosis due to its
48
49 high clinical impact and thus would annotate cases as positive with subtle/non-specific
50
51 findings of TB, while another radiologist may not want to overcall tuberculosis and may
52
53 instead look for the more classical signs specific to the disease. Thus, when creating a
54
55 predictive machine learning model, does one attempt to create different radiologist
56
57
58
59
60

1
2
3
4
5 “personas” (e.g., high sensitivity versus high specificity profiles), or predict what a
6
7 specific radiologist will report, or somehow create a middle-of-the-road report or
8
9 “consensus” report? Alternatively, does the annotation of the final outcome of an
10
11 imaging study get labeled as the actual sputum lab result or the actual clinical outcome?
12
13 If so, then cases that are obviously normal or obviously strongly suggestive of TB will be
14
15 labeled differently due to the clinical outcome. (In a recent academic study on this topic,
16
17 the combination of sputum results, original radiologist interpretations, and confirmation
18
19 by a single over-reading, expert radiologist was required for inclusion into the pulmonary
20
21 TB database [30].) Finally, is the task to predict how a specific radiologist performs or
22
23 how an “average” radiologist performs in interpretation of a radiograph or in prediction of
24
25 the clinical outcome? If the goal is to predict clinical outcome, then issues such as
26
27 prevalence of disease in a particular population may weight too heavily on the
28
29 performance of the system. All of these questions raise important, clinically relevant
30
31 issues that have not yet been resolved.
32
33
34
35
36
37
38
39
40
41
42
43

44 In machine learning, the computer’s greatest strength - its abilities to process data
45
46 endlessly and to repeat the same steps without tiring – could also represent a type of
47
48 Achilles’ heel. This problem is due to the issue of overfitting- defined as the functioning
49
50 of a learning model (or prediction model) that fits so well with its training dataset to the
51
52 extent that it models the statistical noise, fluctuations, biases and errors inherent in the
53
54 dataset, negatively impacting the performance on new data (i.e., diagnostic imaging
55
56 studies not previously presented). This is more likely to occur in medical imaging than
57
58
59
60

1
2
3
4
5
6 in other computer vision applications due to the relatively large number of categories of
7
8 normal and abnormal findings and limited numbers of annotated training sets. More
9
10 succinctly, Domingos indicates that overfitting has occurred “when your learner outputs
11
12 a classifier that is 100% accurate on the training data but only 50% accurate on test
13
14 data, when in fact it could have output one that is 75% accurate on both.” [22] While
15
16 the notion of accuracy in machine learning was relatively simple in the reported studies
17
18 of object recognition, we note that radiology has a rich scientific history of measurement
19
20 of diagnostic accuracy, including the development of receiver-operating characteristic
21
22 of diagnostic accuracy, including the development of receiver-operating characteristic
23
24 (ROC) analysis [32,33,34]
25
26
27
28
29
30
31

32 Classifier performance is central to making informed decisions about machine learning
33
34 and yet, the typical use of a single measure of diagnostic accuracy, while simple, is
35
36 inadequate for technical evaluation. Publications of medical machine learning studies
37
38 are much more informative and rigorous when they utilize ROC analysis because its
39
40 measures of sensitivity and specificity are not dependent on prevalence of disease (as
41
42 is true of accuracy). In addition, the measure of diagnostic accuracy is typically derived
43
44 from use of a single arbitrary threshold, whereas ROC analysis demonstrates the
45
46 performance using all known threshold values. However, since the prevalence of a
47
48 disease does affect the performance of any diagnostic classifier, it would also be helpful
49
50 to know the prevalence of the disease in the test population, so that the false-positive
51
52 and false-negative rates could be determined. Precision, which roughly translates as
53
54 the likelihood that a positive test means that the disease or finding is truly present
55
56
57
58
59
60
61

1
2
3
4
5
6 (otherwise known as the positive predictive value), can demonstrate the relative
7
8 strength or weakness in a classifier for findings or diseases that are low prevalence
9
10 [26,35]
11
12
13
14
15
16

17 The problem of overfitting in medical imaging is also magnified by the wide variety of
18
19 “odd” shapes of normal structures, and the myriads of anatomic variants related to extra
20
21 or missing anatomical structures (such as accessory ossicles or congenitally absent or
22
23 hypoplastic structures). This problem is made most evident by considering the
24
25 problems faced by a radiology researcher who is collecting and classifying the many
26
27 types of anatomic structures and abnormalities that are found on chest radiography.
28
29
30

31 That researcher would have to obtain images and related data for the computer to
32
33 demonstrate abnormalities of the heart, mediastinum, lungs, bones, pleura, and various
34
35 other structures. Distinguishing anatomic variants from pathologic entities has been an
36
37 important function of the practicing radiologist, with a whole atlas devoted to helping
38
39 them avoid making a false-positive diagnosis [36]. In other scientific fields, such as the
40
41 field of genomics, there has been recognition of the unacceptably “false-positive” rate
42
43 associated with various kinds of “wild-type” variations that mimic findings associated
44
45 with genetic mutations associated with cancer [37]. In one study of machine learning
46
47 algorithms devoted to this problem, they characterized the types of false-positive errors
48
49 into six different groups and suggested that “feature-based analysis of ‘negative’ or wild-
50
51 type positions can be helpful to guide future developments in software” [37]. This is
52
53 akin to the problem with anatomic variants in diagnostic radiology.
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9 Because the deep learning approach is highly complex, and because no method has
10 been developed that allows a given algorithm to “explain” its reasoning, technology
11 experts are generally not able to understand fully the reasons for the algorithm’s
12 conclusions, and not able to predict the occurrence and frequency of failure or error in
13 performance of the algorithm [38]. Therefore, validation and regulatory approval could
14 take more time due to the “black box” nature of machine learning approaches.
15
16
17
18
19
20
21
22

23 Fortunately, major advances have been made in recent years in illuminating the
24 contents of the CNN black box [39]. One such advance, saliency maps, was originally
25 proposed in 1998, and is based on the “feature-integration theory” of human visual
26 attention [40]. In 2013, two image visualization techniques for visualization inside deep
27 convolutional networks were demonstrated, one of which involved saliency maps [41].
28 For a given output category value (e.g., a type of interstitial lung disease), saliency
29 maps display the pixels of the image (e.g., CT of the thorax) that were most important
30 for image classification. More recently, other more sophisticated techniques have been
31 developed that organize non-human interpretable convolution layers into an explanatory
32 and potentially interactive graph or image that can be used to speed up the learning
33 process and identify inaccuracies or important areas of an image ignored by a CNN
34 allowing refinement of the model and improving performance [39,42]
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56

57 In contrast, CAD algorithms have been developed over several decades, many of which
58 are focused on specific clinical imaging problems, and therefore have relatively narrow
59
60

1
2
3
4
5 imaging applications. Examples of these applications include: 1) fracture detection,
6 bone age determination, and bone mineral density quantitation in orthopedic radiology;
7
8 2) brain hemorrhage detection, multiple sclerosis detection and quantitation, and
9 regional brain segmentation and volumetry in neuroradiology; and 3) coronary and/or
10 carotid artery stenosis evaluation, and cardiac function assessment in cardiovascular
11 radiology. In order for a machine learning system to replicate the performance of a
12 radiologist, it would have to incorporate large portfolios of narrow machine learning
13 algorithms, each of which has been devised to answer a specific clinical question. The
14 use of combinations of algorithms to solve a single narrow machine learning problem or
15 problems has been referred to as ensemble methods in machine learning and has been
16 successful in winning machine learning competitions on classification of complex
17 datasets. Yet, the integration and orchestration of such a wide and varied array of
18 learning algorithms - possibly from several different developers - into a single clinical
19 system would likely require substantial amounts of time and effort in validation and
20 testing (according to the “no free lunch” theorem of ensemble learning) [43], not to
21 mention the potential regulatory challenges. In the field of artificial intelligence, the “holy
22 grail” is to devise a form of “general artificial intelligence,” which could replicate average
23 human intelligence. General artificial intelligence, as opposed to a collection of narrow
24 artificial intelligences, could help overcome this technological hurdle. Unfortunately, the
25 majority of computer scientists do not believe that generalized artificial intelligence will
26 emerge in the next 20 years, if ever. However, there are other ways that narrow
27 artificial intelligence can help to improve the radiology work process, aside from
28 diagnostic interpretation. There is a wide range of opportunities to increase operational

1
2
3
4
5 efficiency, improve the radiology workflow, and provide decision support to clinicians
6
7
8 and radiologists.
9

10
11
12
13
14 Is it likely that the job of the practicing radiologist is going to be completely displaced by
15 artificial intelligence in the near future?
16
17

18
19
20 Acemoglu and Autor devised a 3 x 2 x 2 matrix model by which “work” can be classified,
21
22 according to whether it is based upon 1) low, medium, or high skills; 2) cognitive or
23
24 manual labor; and 3) routine or non-routine tasks [44]. Based upon their analysis, they
25
26 found two interesting results relevant to a radiologist. First, the rapid diffusion of new
27
28 technologies which substitute capital for labor – such as computerization - resulted in
29
30 decreased demand for work based upon routine tasks. This effect was present whether
31
32 the work is cognitive or manual, but was predominantly found among workers with
33
34 medium skill levels. Interestingly, the types of workers found to be more resistant to job
35
36 displacement included financial analysts (a non-routine, cognitive job) and hairdressers
37
38 (a non-routine, manual job). (We do note that the asset management industry is
39
40 devoting substantial economic resources – even more than that devoted to radiology -
41
42 to incorporate artificial intelligence into financial analysis [45].) With respect to routine
43
44 interpretation tasks performed by the practicing radiologist, it is likely that a machine
45
46 learning system will soon perform some of the routine image interpretation tasks (for
47
48 example, lung nodule screening or preoperative chest radiography). However, much of
49
50 radiologists’ highly skilled work tasks, especially in complex image pattern recognition,
51
52 will be more difficult to replicate over at least the next two decades, and therefore will
53
54
55
56
57
58
59
60

1
2
3
4
5
6 require more time for adequate dataset generation and training, validation, and
7
8 performance testing. This suggests that those radiologists who have acquired higher
9
10 levels of skills (such as higher degrees of sub-specialization, or greater experience in
11
12 narrow, focused areas of clinical imaging) would be even more resistant to job
13
14 displacement. Second, “technical change that makes highly skilled workers uniformly
15
16 more productive” results in a lowering of the threshold for task difficulty that separates
17
18 the medium-skill worker and the high-skill worker [44]. Therefore, in the face of
19
20 potential displacement of radiologists from some image interpretation tasks, many
21
22 radiologists will increasingly spend a higher percentage of time on other valuable
23
24 radiology-based tasks. These radiology-based tasks include those listed in the ACR 3.0
25
26 Initiative, such as: consultation with referring physicians; timely oversight of ongoing
27
28 complex imaging studies; direct patient contact including discussion about test results;
29
30 verification of adherence to national imaging guidelines for proper test ordering;
31
32 participation and data collection for radiology quality initiatives; and timely review of
33
34 radiology-based patient outcomes [46]. The potential shift in the proportion of imaging
35
36 interpretation activities in the daily work of the radiologist is also in keeping with the
37
38 findings of the 2017 McKinsey report on the effects of automation on employment and
39
40 productivity. While over half of all occupations have at least 30% work activities that
41
42 could be automated, no more than 5% of all occupations could be entirely automated;
43
44 this indicates that far more jobs will change than will be eliminated by automation [47].
45
46 In particular, the report states “high-skill workers who work closely with technology will
47
48 likely be in strong demand, and may be able to take advantage of new opportunities for
49
50 independent work.” For radiologists, this could potentially include renewed focus on the
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6 entire spectrum of patient care in imaging. It is likely that new kinds of jobs for
7
8 radiologists will arise as a result of machine learning, similar to the way whereby online
9
10 retail activities led to both a decreased need for marketers and sales staff, and a
11
12 tremendous, high demand for data scientists able to perform the data-mining activities
13
14 needed to assess consumer wants and satisfaction.
15
16
17
18
19
20
21

22 For some expert radiologists, particularly those situated along the frontiers of their
23
24 radiologic subspecialties, there is also the possibility of being involved with a higher
25
26 proportion of non-routine clinical work, including the interpretation of more complex
27
28 imaging technologies that are found to be much more difficult to encode into a machine
29
30 learning system. It is far less likely that sufficiently large datasets could be generated to
31
32 provide neural networks the “experience” to answer questions about less common
33
34 clinico-pathologic entities, or to deal with non-routine clinical issues that often arise in
35
36 medical practice. Therefore, there will remain an important role for the expert
37
38 radiologist who can deal with the non-routine clinical work. This viewpoint is expressed
39
40 by two experts in information systems and economics: “While computer reasoning from
41
42 predefined rules and inferences from existing examples can address a large share of
43
44 cases, human diagnosticians will still be valuable even after Dr. Watson finishes its
45
46 medical training because of the idiosyncrasies and special cases that inevitably arise.
47
48 Just as it is much harder to create a 100-percent self-driving car than one that merely
49
50 drives in normal conditions on a highway, creating a machine-based system for
51
52
53
54
55
56
57
58
59
60
61

1
2
3
4
5 covering all possible medical cases is radically more difficult than building one for the
6
7
8 most common situations.” [48]
9

10
11
12
13
14 Finally, there is significant uncertainty as to whether or not certification by governmental
15
16 regulatory agencies would initially allow these systems to operate autonomously, as
17
18 opposed to requiring oversight by human radiologists. Similar to the steps established
19
20 for CAD in mammography almost two decades ago, we believe that machine learning
21
22 systems will, for the foreseeable future, be approved only for adjunctive use with
23
24 radiologist oversight, over which time it could become the norm for machines and
25
26 humans to work together in imaging study interpretation. At first, this may manifest as
27
28 “worklist triage” in which cases suspected to be more likely to be abnormal by a
29
30 machine learning algorithm will be prioritized for human interpretation.
31
32
33
34
35
36
37
38
39
40

41 Obtaining regulatory (FDA) clearance will continue to be an arduous process during the
42
43 initial introductory phase of machine learning systems into the clinical care environment,
44
45 because of all the intricate details involved in validation and approval of a plethora of
46
47 machine learning systems. The FDA will likely need greater time, resources and
48
49 expertise to evaluate a completely different kind of imaging-based technology, and to
50
51 understand the ramifications of a system wherein the underlying work processes – the
52
53 learning algorithms themselves – are relatively opaque (i.e., a “black box”). Even after
54
55
56
57
58 FDA approval, user acceptance of machine learning systems could be adversely
59
60
61
62
63
64
65

1
2
3
4
5 affected if systemic errors or deviations are detected than cannot be explained. This
6
7 suggests that post-market surveillance could become a more important feature with
8
9 these systems.
10
11
12
13
14
15
16

17 The FDA's "Clinical and Patient Decision Support Software draft guidance" issued in
18
19 December 2017 exempts software that provides decision support that merely makes it
20
21 easier to perform simple calculations or retrieval of accessible data. However, deep
22
23 learning applications are thought to be "black box" and thus must be FDA regulated
24
25 [49]. In response to these challenges, the FDA has recently begun to make significant
26
27 strides toward making the clearance process less onerous. One FDA guidance draft
28
29 document, "Expansion of the Abbreviated 510(k) Program: Demonstrating Substantial
30
31 Equivalence through Performance Criteria" [50] makes 510(k) clearance easier by
32
33 allowing manufacturers to establish "substantial equivalence" functionally using
34
35 performance metrics rather than requiring direct comparison testing and the same
36
37 technology [51]. A few companies have managed to obtain FDA clearance for their
38
39 deep learning-based algorithms related to diagnostic imaging and diagnostic testing.
40
41 Arterys (San Francisco, CA) was the first company to receive clearance by the FDA for
42
43 a deep learning application (for a suite of oncology software for automated
44
45 segmentation of solid tumors on liver CT and MRI scans, and lung CT scans), thereby
46
47 setting a precedent for other applications using CNNs [52]. Also, as of August 2018, the
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

FDA has recently approved clinical decision support software for alerting providers of a

1
2
3
4
5 potential stroke in patients [53], an algorithm for detection of wrist fractures [54] and an
6
7
8 artificial intelligence-based device to detect certain diabetes-related eye problems [55].
9

10
11
12
13
14
15 What is the likely pathway of incorporating machine learning into radiology practice?
16

17
18 Even if the use of machine learning technology throughout society continues to increase
19
20 exponentially, it is not at all clear that machine learning algorithms in a relatively well-
21
22 defined field such as medical imaging will necessarily experience such astronomical
23
24 growth. Advances in computational speed may only guarantee that the same answer -
25
26 including the wrong answer – could be provided 1000 times faster, unless there are new
27
28 techniques or new insights that emerge with approaches to deeper neural networks or
29
30 future approaches such as Bayesian deep learning networks. Currently, machine
31
32 learning for various image recognition algorithms requires presentation of many well-
33
34 annotated imaging studies by human researchers, who then periodically test each
35
36 algorithm for reliability and accuracy. Large imaging datasets will need to be developed
37
38 and shared across institutions and radiology practices; this is an activity that requires
39
40 work and trust to overcome technological, institutional, and regulatory barriers. The
41
42 longstanding requirements of the medical field for high levels of diagnostic accuracy (as
43
44 measured by sensitivity and specificity) and precision in differential diagnosis, will likely
45
46 serve both as important benchmarks by which to judge the usefulness of these
47
48 computer-aided diagnostic algorithms, and as essential “brakes” to the otherwise
49
50 headlong rush to introduce labor-saving technology to reduce costs. The incorporation
51
52 of these machine learning programs into the medical arena will likely be more gradual
53
54
55
56
57
58
59
60

1
2
3
4
5
6 than in other sectors such as industrial, financial, chemistry, astronomy, etc., with a
7
8 reasonable likelihood of a monotonic increase in the rate of progress over the years.
9
10 Our healthcare system is a complex adaptive system and change in portions of that
11
12 system – such as in the radiology industry – is typically characterized by “punctuated
13
14 equilibrium” – i.e., relatively long periods of incremental change, interrupted by relatively
15
16 short bursts of intense change [56]. Thus, the one caveat that we make to our
17
18 prediction of gradual incorporation of machine learning would be the advent of an earth-
19
20 shattering technological innovation in generalized artificial intelligence – such as the
21
22 invention of “the master algorithm,” which is that universal learning algorithm that can be
23
24 applied to disparate fields of knowledge, and yet still make robust, accurate predictions,
25
26 when supplied with sufficient, appropriate data [57]. Only in that case, would we
27
28 suggest that machine learning has become a “10X force” – a change in the business
29
30 force so large that it exceeds the usual competitive influences by an order of magnitude
31
32 [58]. This “sea-change” would then motivate radiologists to prepare for an upcoming
33
34 “strategic inflection point” – that point in time when the old ways of doing business and
35
36 competing in the marketplace are no longer favored, and a new strategic paradigm
37
38 takes over [58]. However, the history of science indicates that the timing of such an
39
40 invention cannot be predicted in advance and likely will not occur any time soon.
41
42
43
44
45
46
47
48
49
50
51
52
53

54 In order to support their cognitive processes, current practicing radiologists have
55
56 already learned to incorporate various kinds of technology, including quantitative
57
58 analysis, 3-dimensional imaging display tools, collaborative tools for consultation, and
59
60
61
62
63
64
65

1
2
3
4
5 digital imaging resources. Future artificial intelligence tools hold the promise of further
6
7 expanding the work that radiologists can do, including in the realms of precision
8
9 (personalized) medicine, and in population management. Rather than replacing
10
11 radiologists, future AI tools could advance the kind of work that radiologists perform; this
12
13 would be in line with the classic IBM Pollyanna Principle: “Machines should work;
14
15 humans should think.” [59,60] At the RSNA 2016 meeting, Keith Dreyer proposed that
16
17 the future model of the radiologist is the “centaur diagnostician”; such a physician would
18
19 team up with the machine learning system to optimize patient care [61]. This idea
20
21 follows the observation that the performance of human-machine teams in playing chess
22
23 could exceed that of a human, or a machine system alone [62]. This partnership would
24
25 yield greater precision and detail in their imaging-based report, including more
26
27 quantitative information and evidence-based recommendations [61]. In addition, this
28
29 could help facilitate advanced visualization techniques, refine clinical-radiologic work
30
31 procedures, and improve the timeliness and quality in communication between the
32
33 radiologist and referring physician, as well as between the radiologist and patient. By
34
35 viewing machine learning systems as a collaborator, not as a competitor, future
36
37 radiologists could benefit from a partnership where the combined performance of the
38
39 radiologist-computer team would likely to be superior to either one alone, and feel
40
41 enriched by the “luxury” of working with the advanced technological support offered by
42
43 machine learning. In addition, the computer could allow the human to do more of what
44
45 he or she does best – such as judicious use of the cognitive abilities associated with
46
47 curiosity, experimentation, and insight. Just as in the example of Advanced Chess, it
48
49 seems likely that the ability to work effectively with the computer will become a distinct
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5 competitive advantage. The futurist Kevin Kelly suggests that we cannot race against
6
7 the machines, but that we can race with the machines. His conclusion is even more
8
9 succinct: “You’ll be paid in the future based on how well you work with robots.” [63] This
10
11 whole concept is also being embraced in various industries, as well as in medicine,
12
13 including the explicit re-definition of “AI” by the American Medical Association as
14
15 standing for “augmented intelligence” rather than “artificial intelligence” [64].
16
17
18
19
20
21
22
23

24 We believe it likely that machine-based learning systems will need oversight for a great
25
26 many years because of the potential for occurrence of many different kinds of errors on
27
28 various kinds of imaging studies. In addition, most current medical imaging algorithms
29
30 are not equipped with the basic knowledge and skills in human anatomy, physiology
31
32 and pathology. If we do reach a point when we might expect that machine-based
33
34 systems approach the accuracy and reliability of a practicing radiologist, then it will
35
36 become a societal issue as to whether or not diagnoses based solely upon machine
37
38 learning are acceptable. If this is viewed solely as a technological upgrade, and if
39
40 society has already accepted other innovations such as self-driving cars, then this
41
42 change may not be controversial. On the other hand, if there is significant adverse
43
44 public reaction to the loss of human interaction in the realm of medicine, then it is
45
46 possible that radiologists may not be displaced for a very long time, if at all. Along
47
48 these lines, Verghese et al. have issued a strong call for the computer and the physician
49
50 to be working together for the foreseeable future and has given a warning about the
51
52 unintended consequences of the implementation of new technology [65].
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9 Given the expected retirement of increasing numbers of baby-boomer radiologists over
10 the next two decades and the growing emphasis on screening and maintenance of
11 health, it is likely that there will be a need for more radiologists over the next 20 years,
12 and that computers will increasingly be regarded by those radiologists as trusted
13 partners. The machine learning systems will be able to help create preliminary reports
14 and note additional findings that may not make it into the final report, but, as is true of
15 CAD today, computers would not be primarily responsible for the final reports. There
16 will be a requirement for much more academic work to be done by human radiologists,
17 including knowledge sharing and transfer learning, even before reaching the stage
18 where the machine-based learning programs can become true partners in the imaging
19 interpretation process [66,67]. Over the last few years, the RSNA R&E Foundation has
20 received an increasing number of submissions of research and education grant
21 applications (one in 2015, 3 in 2016, 9 in 2017, and 27 in 2018) which involve the
22 development of artificial intelligence in radiology, including machine learning [68]. In
23 addition to educational offerings at various universities in the U.S. and around the world
24 (whether as part of degree-granting programs, certificate-based programs, or online
25 training), there are also several developmental opportunities for physicians (whether
26 internships or jobs) at various technology-based corporations in the U.S. The
27 involvement of radiologists in machine-based learning in radiology will be critical in
28 assuring that the care of future patients is not compromised by errors of commission or
29 omission. While not yet part of the radiology curriculum for trainees, it is not hard to
30 imagine that training in radiology informatics is likely to become an even more central

1
2
3
4
5 component of radiology residency education. The first step has been taken by
6
7 organized radiology with the development of a specific training program in radiology
8
9 informatics geared towards fourth-year radiology residents that was funded by the AUR
10
11 and is co-sponsored by RSNA and SIIM [69]. In the U.S., curricular issues in radiology
12
13 education are still governed by the ABR, but educational initiatives to incorporate
14
15 informatics training for all radiology trainees are likely to be in line with future
16
17 developments in diagnostic imaging. Current practicing radiologists will also need to be
18
19 proactive in ensuring that they are full partners in these endeavors, rather than serving
20
21 as “hand-maidens” to the other investigators who “just want their images labeled”. For
22
23 those individuals wishing to learn more about machine learning without having to
24
25 abandon busy clinical careers for any length of time, there are several recommended,
26
27 online courses offered by various academic institutions (including Stanford, MIT and
28
29 Columbia) and by certain corporate entities (including Google and Nvidia) that have
30
31 been available to the public at no charge [70]. Academic medical centers and other
32
33 radiology organizations will need to provide environments where radiologists, machine
34
35 learning experts, and other computer scientists can interact on a continual basis. As
36
37 Davenport and Dreyer point out: “If the predicted improvements in deep learning image
38
39 analysis are realized, then providers, patients, and payers will gravitate toward the
40
41 radiologists who have figured out how to work effectively alongside AI.” [71] Along those
42
43 lines, we find that the creation of the ACR Data Science Institute is a strong indication
44
45 that radiology organizations have recognized machine learning as a potential disruptive
46
47 technology and are getting prepared to respond to this threat by investing resources to
48
49 help develop, adapt, and deploy this new technology in the radiology workspace over
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5 the coming years [72]. In addition, several radiology-based organizations have started
6
7 collaborations with major technology companies to develop machine learning algorithms
8
9 and platforms [73,74,75]. We believe that this is just the beginning of a major trend in
10
11 radiology, and that it behooves radiologists to participate in such endeavors for the
12
13 betterment of radiology practice and the welfare of the patients that we serve.
14
15
16
17
18
19
20
21

22 Conclusion

23
24 We agree that machine learning will continue to make major advances in radiology over
25
26 the next 5 to 10 years, but we completely disagree that there is any real possibility that
27
28 radiologists will be replaced in that time frame, or even during the careers of our current
29
30 trainees. In spite of all the advances of machine learning in the fields of self-driving
31
32 cars, robotic surgery, and language translation, we believe that the work performed by
33
34 radiologists is more complex than is thought by non-radiologists, and therefore more
35
36 difficult to replicate by machine learning. The emergence of deep learning algorithms
37
38 will help radiologists to broaden the kinds of activities that establish their value in clinical
39
40 care (such as routinely providing quantitative analysis), and to enhance the proportion
41
42 of cognitive work (e.g., formulation of diagnosis) relative to visual search work (e.g.,
43
44 detection of imaging abnormalities). Imaging modalities will increasingly utilize deep
45
46 learning to reduce image noise and enhance image quality overall. Since the potential
47
48 for disruption of the radiology industry by machine learning does remain latent, it would
49
50 be wise for various radiology organizations - especially academic institutions - to
51
52 participate in research and development of this technology, and not leave the arena
53
54
55
56
57
58
59
60

1
2
3
4
5
6 solely to corporate entities in the information technology sphere. While the economic
7
8 environment of healthcare will continue to bring change to the practice of medicine and
9
10 radiology, we believe that machine learning will not bring about the imminent doom of
11
12 the radiologist. Instead, we foresee an intellectually vibrant future in which radiologists
13
14 will continue to thrive professionally and benefit substantially from increasingly
15
16 sophisticated and useful machine learning systems over the next few decades.
17
18

19
20 Therefore, we would certainly encourage medical students and others interested in
21
22 radiology as a profession – especially those with expertise in computer science - to
23
24 pursue, enjoy, and look forward to a long career in diagnostic radiology, nuclear
25
26 medicine, molecular imaging, and/or interventional radiology. This would provide
27
28 benefits not only for the practitioners of diagnostic radiology, but even more importantly
29
30 for our patients and for society.
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61

References

1. Lewis-Kraus G. The great A.I. awakening.
https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html?_r=0. Published December 14, 2016. Accessed on January 2, 2017.
2. Mukherjee S. A.I. versus M.D. What happens when diagnosis is automated? :
http://www.newyorker.com/magazine/2017/04/03/ai-versus-md?utm_term=.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. The innovator's dilemma: when new technologies cause great firms to fail. Boston, Mass: Harvard Business School Press, 1997.
4. Christiansen CM, Raynor ME, McDonald R. What is disruptive innovation? Harvard Bus Rev, December 2015. <https://hbr.org/2015/12/what-is-disruptive-innovation>. Accessed April 19, 2018.
5. Chan S. Strategy development for anticipating and handling a disruptive technology. J Am Coll Radiol 2006; 3:778-786.
6. Chockley K, Emanuel E. The end of radiology? Three threats to the future practice of radiology. J Am Coll Radiol 2016; 13:1415-1420.
7. Dargan R. Not so elementary: experts debate the takeover of radiology by machines. RSNA News. <http://www.rsna.org/News.aspx?id=21017>. Published January 1, 2017. Accessed January 2, 2017.
8. Video from RSNA 2016: AI and radiology -- separating hope from hype. Aunt Minnie. <https://www.youtube.com/watch?v=4b0t7TzjZRE> (starting at 5:21)
Published November 29, 2016. Accessed August 26, 2018.
9. Russakovsky O*, Deng J*, Su H, Krause J, Satheesh S, Ma S, et al. (* = equal contribution) ImageNet Large Scale Visual Recognition Challenge. IJCV, 2015
10. Koch C. How the computer beat the Go master. Scientific American, March 19, 2016. <https://www.scientificamerican.com/article/how-the-computer-beat-the-go-master/>. Accessed May 5, 2018.

11. Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, et al. Mastering the game of Go without human knowledge. *Nature* 2017; 550:354-359.
12. Freer TW, Ulissey, MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology* 2001; 220:781-786.
13. Birdwell RL. The preponderance of evidence supports computer-aided detection for screening mammography. *Radiology* 2009; 253:9-16.
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; 279:881–890.
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; 7:802-805.
16. Fenton JJ, Linn A, Taplin SH, Geller BM, Carney PA, D'Orsi CD et al. Effectiveness of computer-aided detection in community mammography practice *J Natl Cancer Inst* 2011; 103:1152-1161.
17. Philpotts LE. Can computer-aided detection be detrimental to mammographic interpretation? *Radiology* 2009; 253:17–22.
18. Mezrich JL, Siegel EL. Legal ramifications of computer-aided detection in mammography. *J Am Coll Radiol* 2015; 12:572-574.
19. Jackson WL. The state of CAD for mammography. *Diagnostic Imaging*. <http://www.diagnosticimaging.com/cad/state-cad-mammography>. Published February 18, 2014. Accessed February 4, 2017.
20. Bahl M, Barzalay 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; 286:810-818.
21. Masarie FE Jr, Miller RA, Myers JD. INTERNIST-I properties: representing common sense and good medical practice in a computerized medical knowledge base. *Comput Biomed Res* 1985; 18:458-479
22. Domingos P. A few useful things to know about machine learning, *CACM* 2012; 55:78-87. <https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf>

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
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 Rad 2018; 15:521-526
 24. Tanz J. Soon we won't program computers. We'll train them like dogs. Wired. <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. <http://ruder.io/transfer-learning> Published March 21, 2017. Accessed August 25, 2018.
 26. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B, 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. <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. <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; 30:392-399
 30. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology 2017; 284:574-582
 31. McBee MP, Awan OA, Colucci AT, Ghobadi CW, Kansagra AP et al. Deep learning in radiology. Academic Radiology 2018 (in press) <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; 171:1217-1219
 33. Metz CE. Basic principles of ROC analysis. Semin Nucl Med 1978; 8:283-298.
 34. McNeil, BJ, Hanley JA. Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. Med Dec Making 1984; 4:137-150

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
35. Oakden-Rayner L. The philosophical argument for using ROC curves.
<https://lukeoakdenrayner.wordpress.com/2018/01/07/the-philosophical-argument-for-using-roc-curves/> Published January 7, 2018. Accessed August 25, 2018
36. Keats TE, Anderson MW. 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; 28:167-175
38. Knight W. The dark secret at the heart of AI. *MIT Technology Review*.
<https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/>
Published April 11, 2017. Accessed August 26, 2018.
39. Zhang Q, Zhu SC. Visual interpretability for deep learning.
<https://arxiv.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; 20:1254-1259
41. Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: visualizing image classification models and saliency maps.
<https://arxiv.org/abs/312.6034> Published December 20, 2013. Accessed August 26, 2018
42. Zhang Q, Cao, F. Shi F, Wu YN, Zhu SC. Interpreting CNN knowledge via an explanatory graph. <https://arxiv.org/pdf/1708.01785> Published November 22, 2017. Accessed October 4, 2018.
43. Polikar R. Ensemble based systems in decision making. *IEEE Circuit Syst Mag* 2015; 6:21-45
44. Acemoglu D, Autor D. Skills, tasks and technologies: implications for employment and earnings. Working paper, National Bureau of Economic Research. <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*, April 9, 2018; L7-L12

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
46. McGinty G, Allen Jr B, Wald C. Imaging 3.0™. American College of Radiology, published in 2013.
<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. <http://www.mckinsey.com/global-themes/digital-disruption/harnessing-automation-for-a-future-that-works> Accessed July 2, 2017
48. Brynjolfson E, McAfee A. The second machine age: work, progress, and prosperity in a time of brilliant technologies. New York, NY: Norton, 2014;192-193.
49. Edwards B. FDA guidance on clinical decisions: Peering inside the black box of algorithmic intelligence. <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.
<https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM604195.pdf> Published April 12, 2018. Accessed August 26, 2018.
51. U.S. Food and Drug Administration. FDA In Brief: FDA to offer a voluntary, more modern 510(k) pathway for enabling moderate risk devices to more efficiently demonstrate safety and effectiveness.
<https://www.fda.gov/NewsEvents/Newsroom/FDAInBrief/ucm604348.htm> Published April 12, 2018. Accessed August 26, 2018.
52. FDA News. Arterys wins FDA clearance for oncology imaging software.
<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.
<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.

1
2
3
4
5
6 <https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm608833.htm>
7 Published May 24, 2018. Accessed August 26, 2018.
8

9
10 55. U.S. Food and Drug Administration. FDA permits marketing of artificial
11 intelligence-based device to detect certain diabetes-related eye problems.
12 <https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm604357.htm>
13 Published April 11, 2018. Accessed August 26, 2018.
14

15
16 56. Enzmann DR, Feinberg DT. The nature of change. J Am Coll Radiol
17 2014;11:464-470.
18

19
20 57. Domingos P. The master algorithm; how the quest for the ultimate learning
21 machine will remake our world. New York, NY: Basic, 2015; 25-26.
22

23
24 58. Grove AS. Only the paranoid survive; how to exploit the crisis points that
25 challenge every company. New York, NY: Currency, 1996.
26

27
28 59. Bloch A. Murphy's Law and other reasons why things go wrong. Los Angeles,
29 CA: Price/Stern/Sloan, 1979; 41
30

31
32 60. Henson J. The Paperwork Explosion. The Jim Henson Company (for IBM).
33 www.youtube.com/watch?v=IZw2CoYztk Produced October 1967. Published
34 March 30, 2010. Accessed August 26, 2018.
35

36
37 61. Dreyer K. When machines think: radiology's next frontier. Presented at the 2016
38 annual meeting of the Radiological Society of North America; Chicago, Ill.
39

40
41 62. Thompson C. The rise of the centaurs. In: Smarter than you think; how
42 technology is changing our minds for the better. New York, NY: Penguin,
43 2013;1-5.
44

45
46 63. Kelly K. The inevitable; understanding the 12 technological forces that will shape
47 our future. New York, NY: Viking, 2016.
48

49
50 64. American Medical Association. AMA Passes First Policy Recommendations on
51 Augmented Intelligence. [https://www.ama-assn.org/ama-passes-first-policy-](https://www.ama-assn.org/ama-passes-first-policy-recommendations-augmented-intelligence)
52 [recommendations-augmented-intelligence](https://www.ama-assn.org/ama-passes-first-policy-recommendations-augmented-intelligence) Published June 14, 2018. Accessed
53 August 26, 2018.
54

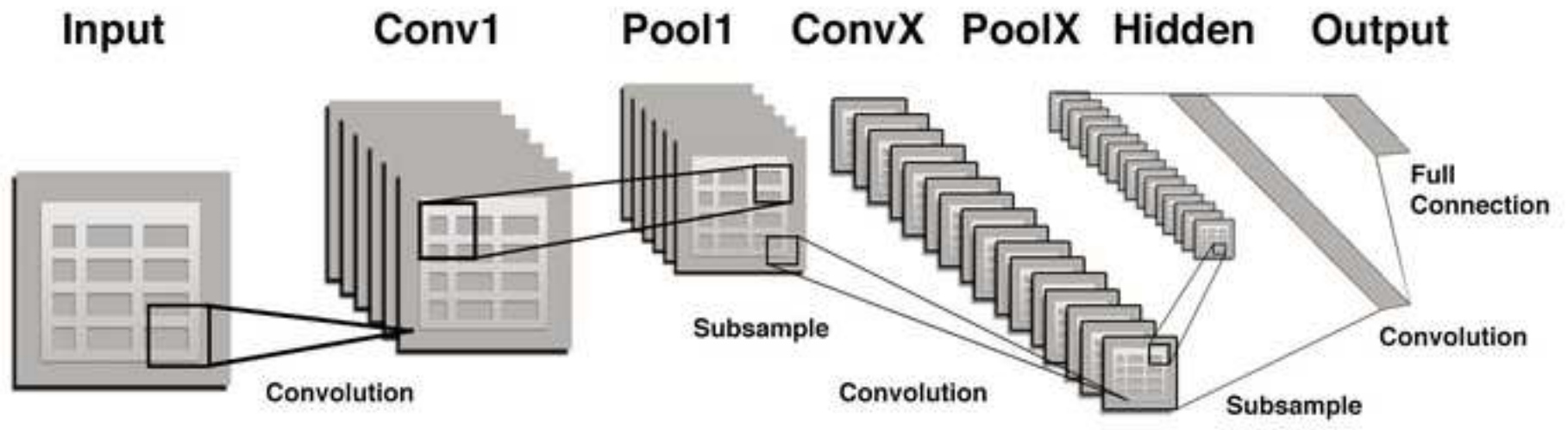
55
56 65. Verghese A, Shah NH, Harrington RA. What This Computer Needs Is a
57 Physician; Humanism and Artificial Intelligence. JAMA 2018; 319:19-20.
58

59
60 66. Gunderman R, Chan S. Knowledge sharing in radiology. Radiology 2003;
61 229:314-317.
62

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
67. Kohli M, Prevedello LM, Filice RW, Geis JR. Implementing machine learning in radiology practice and research. AJR 2017; 208:1-7.
68. Walter S. Private communication: unpublished observation on trends in RSNA R&E Foundation grants applications 2014-2018. (March 2018).
69. National Imaging Informatics Curriculum (NIIC).
<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. Forbes.
<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. Harvard Bus Rev March 27, 2018. <https://hbr.org/2018/03/ai-will-change-radiology-but-it-wont-replace-radiologists> Accessed April 19, 2018.
72. American College of Radiology (ACR) Data Science Institute.
<http://www.acrdsi.org/index.html> Accessed May 5, 2018.
73. Partners Healthcare. Partners HealthCare and GE Healthcare launch 10-year collaboration to integrate Artificial Intelligence into every aspect of the patient journey. <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.
<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. Radiology Partners. Radiology Partners to join the IBM Watson Health Medical Imaging Collaborative.
<http://www.radpartners.com/events/j4ss9xuz22/Radiology-Partners-to-join-the-IBM-Watson-Health-Medical-Imaging-Collaborative> Published July 6, 2017. Accessed August 26, 2018.

1
2
3
4 Figure 1. A portion of the input (image) is passed to each successive pair of convolutional/pooling layers
5 (filter/parameter reducers) with several convolution and pooling layers added before an output
6 (prediction) is made. Initial layers tend to represent general features such as edges and colors and later
7 layers represent features of increasing complexity, such as corners, and then textures, followed by even
8 complex features such as a snout or whiskers, and finally entire objects such as a dog or cat. Finally,
9 there is a “fully connected layer” that flattens input from other layers transforming them into a decision
10 on whether the output belongs to a certain class (e.g., dog vs. cat). Errors in classification in a training
11 set are then “back-propagated” to modify and/or update the filters so that overall errors are minimized.
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

PROOFS



BJR UNC

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 a machine learning 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 machine learning 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 machine learning 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.