Deep Learning in Radiology

Morgan P. McBee, MD, Omer A. Awan, MD, MPH, CIIP, Andrew T. Colucci, MD, Comeron W. Ghobadi, MD, Nadja Kadom, MD, Akash P. Kansagra, MD, MS, Srini Tridandapani, MD, PhD, MSEE, MSCR, MBA, William F. Auffermann, MD, PhD

As radiology is inherently a data-driven specialty, it is especially conducive to utilizing data processing techniques. One such technique, deep learning (DL), has become a remarkably powerful tool for image processing in recent years. In this work, the Association of University Radiologists Radiology Research Alliance Task Force on Deep Learning provides an overview of DL for the radiologist. This article aims to present an overview of DL in a manner that is understandable to radiologists; to examine past, present, and future applications; as well as to evaluate how radiologists may benefit from this remarkable new tool. We describe several areas within radiology in which DL techniques are having the most significant impact: lesion or disease detection, classification, quantification, and segmentation. The legal and ethical hurdles to implementation are also discussed. By taking advantage of this powerful tool, radiologists can become increasingly more accurate in their interpretations with fewer errors and spend more time to focus on patient care.

Key Words: Machine learning; deep learning; machine intelligence; artificial intelligence.

INTRODUCTION

Recent rapid advances in computer hardware and software now allow computers to perform an increasing number of tasks that have historically not been possible (1). The technologies that can in some ways mimic the decision-making abilities of humans are known by several names, depending on the nature of the algorithms used. One of the more sophisticated sets of algorithms is often referred to as deep learning (DL). DL has made great advances in recent years, now performing tasks that only humans could perform just a few years ago. Some may perceive DL algorithms as a threat to medicine and radiology. However, DL is like any other tool, intrinsically neither good nor evil, but rather dependent on the application. In this work, the Association of University Radiologists Radiology Research Alliance Task Force on Deep Learning provides an overview of DL for the radiologist. The goal of this task force was to examine developments in DL and how they will influence the current and future practice of radiology. This article seeks to present an overview of DL in a manner that is understandable to radiologists; to examine past, present, and future applications; and to evaluate how radiologists may benefit from this remarkable new tool. Additional resources providing greater depth of coverage are included for the interested reader.

EVOLUTION OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Most people's knowledge of artificial intelligence (AI) is derived from science fiction movies. AI is defined as "the capacity of computers or other machines to exhibit or simulate intelligent behavior" (2) and is now a thriving field and the focus of a great amount of research and investment. Early on, the focus of AI was to address problems that were difficult for humans but relatively straight forward for computers to solve. Such problems include abstract and formal mathematical problems, such as adjusting the window and the level of a radiographic image on a viewing workstation. Additionally, the early attempts of AI were based on rigid predefined rules, but this approach was largely unsuccessful (3).

An advance in AI was the advent of machine learning (ML), which is the ability of an AI system to extract information from raw data and to learn from experience. This avoids the need for "human operators to formally specify all of the knowledge that the computer needs" (3). For example, an ML algorithm introduced in 1990 utilized logistic regression to determine whether or not cesarean section was appropriate (4).

Acad Radiol 2018; 25:1472–1480

From the Department of Radiology and Medical Imaging, Cincinnati Children's Hospital, Cincinnati, Ohio (M.P.M.); Department of Radiology, Temple University Hospital, Philadelphia, Pennsylvania (O.A.A.); Department of Radiology, Beth Israel Deaconess Medical Center, Boston, Massachusetts (A.T.C.); Department of Radiology, University of Chicago, Chicago, Illinois (C.W.G.); Department of Radiology and Imaging Sciences, Children's Healthcare of Atlanta (Egleston), Emory University School of Medicine, Atlanta, Georgia (N.K.); Mallinckrodt Institute of Radiology and Departments of Neurological Surgery and Neurology, Washington University School of Medicine, Saint Louis, Missouri (A.P.K.); Department of Radiology and Imaging Sciences, Emory University School of Medicine, Electrical & Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia (S.T.); Department of Radiology and Imaging Sciences, Emory University School of Medicine, 1365 Clifton Road NE, Atlanta, GA 30322 (W.F.A.). Received August 13, 2017; revised February 22, 2018; accepted February 23, 2018. **Address correspondence to:** W.F.A. e-mail: william.auffermann@hsc.utah.edu

https://doi.org/10.1016/j.acra.2018.02.018

Broadly speaking, ML comprises a set of algorithms that aim to allow computers to receive an assortment of input data and to generate complex inferences that are based on potentially obscure relationships between inputs. For example, ML algorithms may play an important role in combining financial data (such as company and industry earnings reports) and nonfinancial data (including information of geopolitical events and weather patterns) to generate nuanced recommendations about whether to buy or sell an equity stock position.

WHAT IS DEEP LEARNING?

DL has received a great deal of attention lately both in the consumer world and throughout the medical community, whereas ML algorithms have been a focus of research for many years. There has been a renewed interest in DL algorithms lately since they reduced the top-5 error by 10% in 2012 at the ImageNet Large Scale Visual Recognition Challenge (5). Top-5 error is defined as "the fraction of test images for which the correct label is not among the five labels considered most probable by the model." Every year since then, DL models have dominated the challenges, significantly reducing the top-5 error, and in 2015, human performance was surpassed by DL algorithms.

Another relatively recent advancement is the application of graphics processing units (GPUs) in ML algorithms. GPUs have been used for decades in the video game market and are now available at a relatively low cost. GPUs excel at the types of computations needed for DL applications and, in fact, speed up DL algorithms. The usage of GPUs has been a significant contributing factor in advancements in pattern recognition, image segmentation, and object detection (1), all of which are highly relevant to radiology.

DL is not a specific algorithm but is rather a technique that involves many layers. The DL algorithms most applicable to radiology are called convolutional neural networks (CNNs) as these are very efficiently applied to image segmentation and classification (6).

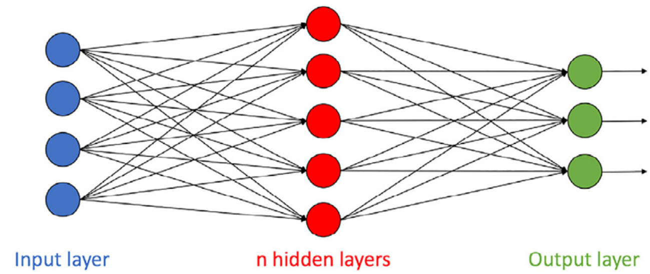

Figure 1. Basic representation of an artificial neural network with *neurons* similar to those within a brain. The left layer of the neural network is called the *input* layer and contains neurons that encode the values of the input pixels. The rightmost layer is called the *output layer*, which contains the output neurons. The middle contains "n" number of *hidden layers*, which perform mathematical transformations or *convolutions* of the data. (Color version of figure is available online.)

CNNs are called "neural networks" based on the fact that their structure is analogous to biological nervous systems (Fig 1). Lower level information inputs, akin to cutaneous sensory nerves, form synaptic connections to the next level or "layer" of neurons. Each neuron in this second layer can combine the inputs from lower level neurons to form a newer, more complex output. As the number of intermediate or hidden layers increases, so too does the allowable complexity and richness of the output from the highest layer. Simple neural network-based ML algorithms typically include only a small number of these layers (Fig 1). DL algorithms may include many, many more. Having more layers has been shown to increase test accuracy (5) (Fig 2).

Several factors contribute to the accuracy of DL algorithms. One factor that has been shown to increase accuracy is the number of times a dataset is passed through the ML algorithm (epoch), as shown in Figure 3. Having more layers has been shown to consistently increase test accuracy; a representative figure of this relationship is illustrated in Figure 2.

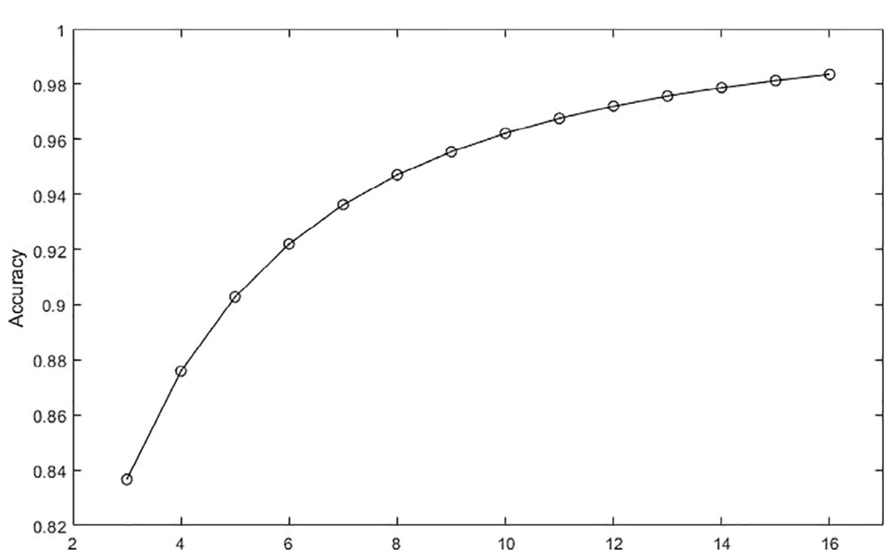

Figure 2. A representative example of how increasing the number of layers (*x* axis) increases the test accuracy (*y* axis).

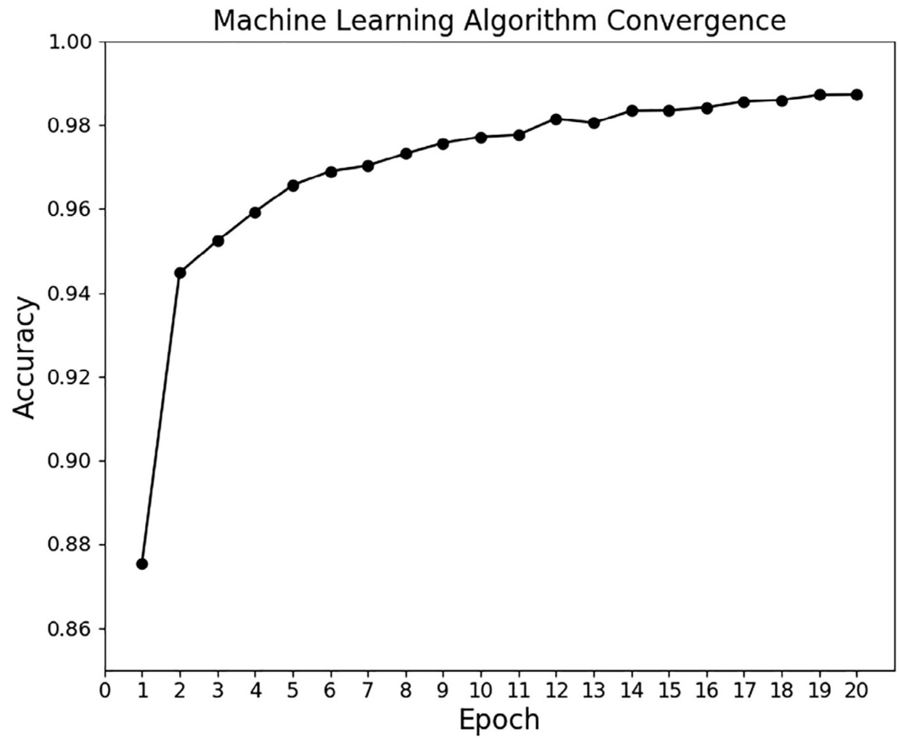

Figure 3. A curve showing the convergence of accuracy for a machine learning algorithm for chest radiograph data as a function of the number of iterations (epochs).

The historical challenge of DL has been the need for tremendous computational power. To make the problem more mathematically tractable within the limits of computational reality, researchers have had to simplify and constrain the problems they address in various ways. One common strategy is to curate and label input data to reduce the number of neuronal layers needed to generate meaningful output. Another strategy is supervision of the learning process, wherein the algorithm is allowed to infer certain relationships only in specific, predefined ways. These strategies are still in common use, but with recent dramatic advances in computational power, researchers are increasingly able to utilize unsupervised learning of large-scale, unlabeled data. A familiar example from this decade is IBM's Watson, which relies on DL algorithms to produce meaningful answers to natural language queries based on free, unstructured data available on the Internet and a variety of other information databases.

Creating a CNN and training it from conception can take a significant amount of time and resources. An alternative to this approach is *transfer learning*, which involves transferring the knowledge gained by a CNN in one dataset to another dataset with a completely different kind of data. In the context of medical imaging, this method usually means training the algorithm on a large variety of nonmedical images. In fact, CNNs trained in this manner outperform or, at the very least, perform just as well as those trained with purely medical images (7−9).

DL may proceed using two basic approaches: supervised learning and unsupervised learning. In supervised learning, the computer is given labeled datasets in which objects have been preclassified, and the algorithm looks for features differentiating the objects in each class. In unsupervised learning, the computer algorithms are given unlabeled data (objects that have not been prepartitioned into classes). The unsupervised DL algorithm is then tasked with both determining the labels of the different classes of objects and separating the objects into their appropriate classes. As an example, an unsupervised DL algorithm may be tasked with both identifying features that differentiate benign and malignant nodules and classifying the nodules into their respective class or category.

The goal of DL is an intelligent computer system that can disentangle massive amounts of unlabeled and unstructured data to produce complex and meaningful insights. In a sense, this is the workflow of the modern radiologist—translating a large, digital dataset containing pixel intensity values into an accurate diagnosis. For a compressive overview of the details of how DL works (under the hood), the interested reader is directed to the text by Goodfellow et al. (10).

CURRENT ADVANCES IN DEEP LEARNING OUTSIDE OF MEDICINE

DL has found many applications outside of medicine. Some of these applications include, but are not limited to, the following areas: gaming, language (written and spoken), financial analysis, and imaging (processing and analysis). Some of these applications are discussed further in greater detail.

Gaming

The first computer games were developed in the 1950s and have evolved rapidly as computer and software technology has advanced. Two specific games that have undergone much examination have been Chess and Go.

Chess

Computer chess algorithms were first developed in the late 1950s but had limited capabilities and were not very successful against human players. As technology advanced, so did the abilities and performance of chess playing algorithms. For many years, even the best computer algorithms could not defeat a grand master in chess. However, in 1997, Deep Blue defeated Gary Kasparov, then the reigning world champion, in a chess match. Many years of development and training were needed to achieve this result (11). Subsequently, in 2015, Matthew Lai developed the DL chess algorithm, Giraffe, which played thousands of games against itself and in 72 hours essentially trained itself to play at the master level (12).

Go

Go is a game played on a 19 × 19 board and has an immense game space. For reference, there are approximately 10^{360} possible moves in a game of Go, compared to approximately 10^{123} possible moves in a game of chess, and an estimated 10^{80} atoms in the known universe (13). Until recently, computers could not defeat professional Go players. In October 2015, DL-based AlphaGo (Google, Mountain View, CA) defeated Fan Hui in five of five games to become the first computer to defeat a professional Go player without handicaps (14).

Language

DL has been proven to be useful for several applications in the realm of language. For instance, DL algorithms have recently been used for automatic translation of text documents between languages (15). Another application of ML in language is the Chatbot, also known as an Artificial Conversational Entity. These programs use both natural language processing (NLP) and DL to analyze human input and to generate a response (16). The exact scope of Chatbot use is difficult to quantify, but it is likely that many of us have communicated with a Chatbot online and have not even realized we were interacting with a computer.

DL has also been used in speech recognition. Virtual assistants may be thought of as the computer analog of an administrative assistant. Today there are several virtual assistants available, including Siri (Apple Inc, Cupertino, CA), Alexa (Amazon, Seattle, WA), Cortana (Microsoft, Redmond, WA), and Google Assistant (Google, Mountain View, CA). The algorithms underlying these applications rely on ML for both recognition of speech and interpretation of the actions the user desires to execute. Applications in radiology are discussed later.

Imaging

DL has been used in image processing and analysis for a variety of nonmedical applications. For example, DL has been used to colorize black and white images. Even with computer film colorization in the late 1980s and 90s, colorization was still a long and labor intensive process, often requiring significant human intervention. Now DL learning algorithms are able to recognize the correct color for many objects and to colorize images with little human input (17). Optical character recognition refers to the ability of computers to identify and translate handwritten human characters into machine-encoded text. While recognition of written characters had been a difficult task for computers to perform, DL algorithms have successfully converted written language into machine-encoded text (18).

DEEP LEARNING IN MEDICINE

The implementations of DL in other areas of medicine most relevant to radiology not surprisingly involve imaging. Examples include visible light images—photographs, such as those taken of skin lesions (particularly malignancies)—and ophthalmologic funduscopic images. Such images are particularly suited for DL techniques because they are typically only a single image as opposed to the thousands of images common in advanced imaging studies.

Utilizing two different validation sets with two different set points (high specificity or high sensitivity), researchers were able to achieve 90.3% and 87.0% sensitivity and 98.1% and 98.5% specificity for detecting referable diabetic retinopathy at the high specificity set point, and 97.5% and 96.1% sensitivity and 93.4% and 93.9% specificity at the operating point selected for high sensitivity. This finding was complementary to the low false-positive rates of ophthalmologists (19).

A CNN developed to classify skin lesions (keratinocyte carcinomas vs benign seborrheic keratoses, and malignant melanomas vs benign nevi) achieved performance equivalent to dermatologists (20).

A framework developed to detect and localize metastatic disease on gigapixel microscopy images utilizing a CNN architecture achieved image-level area under the curve (AUC) scores above 97% (21). In a similar vein, researchers achieved an AUC of 89% when utilizing a DL algorithm to automatically identify metaphase chromosomes using scanning microscopic imaging (22).

CURRENT APPLICATIONS IN RADIOLOGY

Radiology differs from other image recognition applications of DL algorithms in that a computed tomography or magnetic resonance imaging examination can consist of thousands of images as opposed to a single image. This greatly increases the complexity of required computational algorithms. Additionally, other applications such as facial recognition deal with a relatively homogenous set of images of faces, whereas images in radiology can vary widely, depending on patient factors and pathologies, which further increase the complexity of the problem.

Currently, most of the applications within radiology are narrowly focused on achieving a specific task. Areas of active focus within radiology can be divided into several different categories: lesion or disease detection, classification and diagnosis, segmentation, and quantification. These categories are somewhat arbitrary and have a significant amount of overlap, but provide a useful framework for discussing current applications of DL in radiology. A large portion of DL research in radiology, to date, has been in the fields of cardiothoracic imaging and breast imaging, although the range of applications is rapidly expanding.

Additionally, there are many ML applications in radiology performance improvement and health policy, but these are beyond the scope of this article.

LESION OR DISEASE DETECTION

There are differences in medical and nonmedical data, and one of the most notable may be the importance of relatively small findings on images. Interestingly, DL systems utilizing the ImageNet training data (ie, nonmedical images) have been shown to be effective at categorizing findings on chest radiographs, such as pleural effusion, cardiomegaly, and mediastinal enlargement (8). Another more recent study classifying tuberculosis on chest radiographs showed that utilizing a DL system pretrained with AlexNet and GoogLeNet nonmedical data was the most effective, with an AUC of 0.99. With a radiologist-augmented approach, the achieved sensitivity was 97.3% and the specificity was 100% (23). This finding suggests that the algorithms are able to handle images from a wide variety of sources and are not restricted to the image domains for which they were originally developed.

Computer-aided detection (CAD) applications utilizing DL systems are significantly more effective than traditional systems (24−26). Several different applications have shown that DL is highly effective at identifying pulmonary nodules (8,24,27−29). Such applications can enable radiologists to practice "at the top of their license" instead of spending time tediously searching for small lesions.

DL has been very successfully applied to the identification and characterization of imaging abnormalities. One particular area where DL has been proven to be useful is in the diagnosis of pulmonary nodules. Note that the diagnosis component of CAD includes characterization and classification of lesions. The historical model for both detection and characterization or diagnosis involved using predefined rules. These rules were often predefined by humans and few in number. DL algorithms allow computer algorithms to define features of interest themselves based on the characteristics of the dataset and may be far greater in number than the feature sets selected by humans (30). As a result of the use of DL algorithms, the sensitivity and true positives per examination for the detection of pulmonary nodules have increased. Consider that an early example of a CAD algorithm for nodule detection on computed tomography had a sensitivity of 72% with 4.6 false positives per study (31), with newer DL algorithm demonstrating sensitivities of up to 92% with 4 false positives per scan (32).

CLASSIFICATION

ML algorithms excel at solving linear and logistic regression problems. Classifying images into one of two or more categories based on imaging findings represents a logistic regression problem.

There have been many applications of DL within chest imaging. Unlike conventional ML classification, which requires predefined features, DL algorithms are able to create or identify their own features for classification (Fig 4). Several researchers have demonstrated the ability of a CNN to classify lung nodules as benign or malignant (33,34). DLs with CNNs have accurately classified tuberculosis on chest radiographs with an AUC of 0.99. Utilizing a radiologist-augmented approach further improved accuracy (23). Other ML

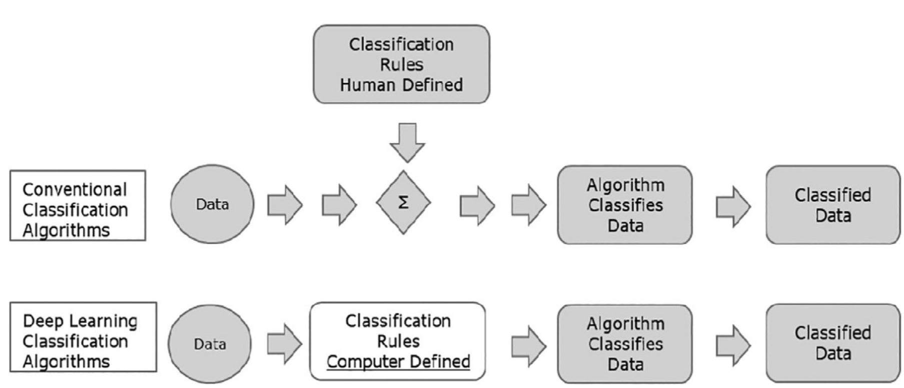

Figure 4. Differences between classification using conventional algorithms and deep learning algorithms. Note that Σ indicates combination of data inputs.

algorithms, such as Bayesian and support vector machines (SVMs), have successfully characterized different obstructive lung diseases on high-resolution CT (35). A CNN with five convolutional layers was able to achieve an 85.5% accuracy rate for classifying interstitial lung diseases (36). CNNs have also been effective at classifying the presence or the absence of an endotracheal tube on chest radiograph with an AUC of 0.99 (37).

There is a great deal of overlap between detection and classification with CNNs in mammography as many CNNs designed for detection also ultimately aim to classify lesions. Additional applications in breast imaging include accurate classification of breast density on mammograms (38) and classification of tumors (9).

Applications of DL have also been demonstrated in musculoskeletal imaging. A highly accurate, fully automated system of determining bone age was developed with an interpretation time of less than 2 seconds (39).

There have been mixed results with transfer learning in medical imaging applications. Initial results showed that using nonmedical image databases to train CNNs later used in medical image analysis can increase accuracy (22), and transfer learning with CNNs can be used to effectively classify abdominal ultrasound images (40), but the use of transfer learning with images of natural scenes did not improve the estimation of treatment response in patients with bladder cancer. However, transfer learning with images of bladders did improve classification (41).

Exploiting radiology report databases by using modern information processing technologies may improve report search and retrieval and help radiologists in diagnosis. Compared to search reports using keywords, NLP and natural language understanding provide a more efficient way to organize and retrieve relevant information from radiology reports.

QUANTIFICATION

Many areas in radiology may benefit from improved tools for quantification, for example, lung nodule volumes, liver iron, brain atrophy, and brain tumors (Response Evaluation Criteria in Solid Tumors and other measurement systems). The amount of cerebral edema following stroke can be accurately quantified automatically using ML algorithms (42).

"Radiomics," a field of study involved with the extraction of large numbers of features from medical imaging examinations using data-characterization algorithms, is undergoing a rapid revolution with the introduction of ML techniques. Traditionally, radiological diagnosis involved the extraction of high-level imaging features by experts, for example, radiologists, who through experience noted the relationships between clinical factors and imaging features. Typically, these involved a limited set of features, such as diameter, volume, attenuation and signal intensities, and enhancement values and trajectories. However, with quantitative ML approaches, a rich and large set of radiomic features, which in many cases are not perceived by human eyes, can be employed to correlate imaging features with clinical factors, diagnosis, and outcomes. These radiomic features include

1. the more traditional but limited features based on size and shape;
2. descriptors of the relationship between image voxels (eg, gray-level co-occurrence matrix (43) and run-length matrix (44), among others;
3. descriptors of histograms of image intensity;
4. textures extracted from filtered images; and
5. complex fractal features.

Texture-based radiomic features were useful in predicting which patients with esophageal cancer would best respond to therapy (45). Radiomic features have also been successfully used to predict clinical outcomes and in predicting the risk of distant metastases in lung cancer. For example, Coroller et al. (46) identified 35 radiomic features of lung cancers that were useful in distinguishing patients at high risk of developing distant metastases. Historically, such texture features were analyzed with relatively standard statistical methods (including measures of the mean and analysis of variance). Although these results have been promising, DL algorithms have the potential to identify subtle or complex patterns that may elude humans and conventional statistical methods. Such texture features may be analyzed by a DL algorithm, either in isolation or in conjunction with radiological and endoscopic datasets, and allow for identification of patterns not previously perceived.

SEGMENTATION

Segmentation of brain MRIs used to be a tedious task that required a great deal of manual intervention. However, DL algorithms have been highly effective at automatically segmenting brain anatomy (6,47). Furthermore, one study was able to achieve accurate automatic organ segmentation with CNNs with training based on only a single manually segmented image (48). Automated segmentation applications in prostate imaging have also been successful (49). A representative example of a segmented chest image is shown in Figure 5.

Speech Recognition

DL algorithms are used in speech recognition in radiology, particularly in Europe and the United States. Before computer speech recognition, human transcriptionists would transcribe audio recordings of radiologist reports. However, human transcription of a report could take minutes to days, depending on the setting. Computer speech recognition offered an efficient means of dictating studies in busy academic and private practice settings. Specific algorithms using neural networks are utilized for voice user interfaces, NLP, and speech-to-text transcription for radiology reporting. Speech recognition software such as Dragon (Nuance Communications, Burlington, MA) and

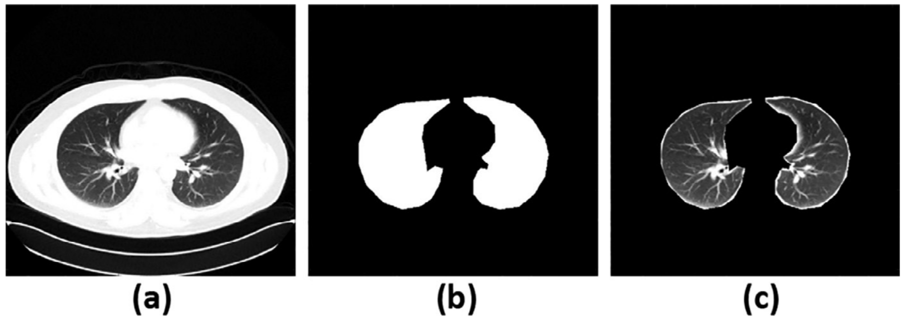

Figure 5. A representative example of chest and lung segmentation. **(a)** Original computed tomography image of the chest in lung windows. **(b)** Region of the image corresponding to the lungs. **(c)** Segmented image with the chest wall and the mediastinum removed and the lungs isolated.

SpeechRite (Capterra, Arlington, VA) have been using ML algorithms to develop successful voice recognition software to aid the Radiologist. A recent study showed that using speech recognition may result in decreased turnaround time from 35% to over 90% (50). A variety of algorithms are currently used for voice recognition; please refer to the review article by Deng and Li (51) for a more comprehensive survey.

CURRENT LIMITATIONS

Although great promise has been shown with DL algorithms in a variety of tasks across radiology and medicine as a whole, these systems are far from perfect. Neural networks can be "statistically impressive, but individually unreliable" and can make mistakes that humans would not (52).

Reverse engineering of a DL system can allow someone to subtly alter the input data in imperceptible ways. Researchers were able to modify images in a way that is undetectable to the human eye but renders a DL image classification algorithm ineffective. Examples include mistakenly classifying a flagpole for a Labrador or a joystick for a Chihuahua (53). Another example outside of medicine is a Twitter bot created by a large technology company that utilized DL algorithms, which was successfully manipulated by Internet users to make offensive remarks (52).

A potential challenge is that consistent data for training are needed, and this can create problems with annotated images. With specific regard to radiology, problems exist with datasets, in that companies often want to protect their intellectual property and keep the datasets proprietary (54).

Additionally, there may not be a consensus on proper annotations for image review, diagnosis, and decision-making. Obtaining high-quality annotated datasets will remain a challenge for DL. An additional challenge from a clinical standpoint will be the time to test how well DL techniques perform vs human radiologists. This challenge will require a large study, time to perform the study, and a large cost for the duration of the study. Another challenge could be the lack of collaboration between physicians and ML scientists. The high degree of complexity of human physiology will also be a challenge for ML techniques.

Another challenge is the requirements to validate a DL system for clinical implementation. Such a validation process would likely require multi-institutional collaboration and large datasets. The historical model for health care was for health-care systems to practice with relatively little sharing of large datasets. As health systems grow, larger datasets will become available for DL training sets. Coupled with the emergence of cloud health-care analytics, obtaining training sets large enough for algorithm validation is becoming more feasible technically.

Ethical and legal challenges revolve around who will take responsibility for the images if DL and ML techniques perform interpretation of studies. This can have legal ramifications for lawsuits and ethical issues are raised as well. For example, if a DL algorithm failed to identify a pulmonary nodule, would the algorithm vendor or radiologist be responsible? Based on the historic way CAD and diagnosis has been treated legally, it would likely be the radiologist. However, this may change as DL algorithms have greater independence and autonomy with regard to medical image interpretation.

FUTURE APPLICATIONS

There are many potential future applications of DL in radiology in which practically every aspect of image interpretation could see potential uses. Additionally, future applications include worklist optimization to triage studies with life-threatening findings to be read earlier by the radiologist (eg, subdural hematoma, stroke, and aortic dissection) (54), NLP, novel diagnostic applications, prognostication, automated tracking of imaging findings, and automated preliminary report generation.

Exploiting radiology report databases by using modern information processing technologies may improve report search and retrieval and help radiologists in diagnosis. Compared to search reports using keywords, NLP and natural language understanding provide a more efficient way to organize and retrieve relevant information hidden in the radiology reports (55).

Novel diagnostic applications are a possibility. For example, patients with schizophrenia can be distinguished from controls based on the connectivity of the anterior insula (56). Additional applications include neurodegenerative disorders such as Alzheimer disease (57,58).

Traditional prognostic risk assessment in patients undergoing noninvasive imaging is based on a limited selection of clinical and imaging findings. ML can consider a greater number and complexity of variables. For example, it is feasible that ML can predict 5-year all-cause mortality in patients undergoing coronary computed tomographic angiography (59).

Automatic tracking of tumor markers, such as maximum standard uptake values and tumor size, can be imagined as a future application of DL based on work showing that automated tumor volume segmentation is possible (60).

CONCLUSION

There has been a great amount of research within medicine and especially radiology in regard to DL, which has shown great promise. Perhaps DL algorithms can one day delve deeper into unsupervised learning territory and show us patterns that humans are unable to perceive. Imagine utilizing existing modalities in brand new ways, such as diagnosing appendicitis on radiographs or predicting heart attack risk based on an ultrasound. DL offers exciting opportunities for radiologists to improve safety by providing more accurate diagnoses, increasing efficiency by automating tasks, and helping to generate data on imaging features that were not previously used as diagnostic criteria.

REFERENCES

1. Schmidhuber J. Deep learning in neural networks: An overview. Neural Netw 2015; 61:85–117.
2. Dictionary OE. Artificial intelligence, n. Oxford University Press; Available at: http://www.oed.com. Accessed December 8, 2017.
3. Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press, 2016.
4. Mor-Yosef S, Samueloff A, Modan B, et al. Ranking the risk factors for cesarean: logistic regression analysis of a nationwide study. Obstet Gynecol 1990; 75:944–947.
5. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks Lake Tahoe, Nevada, Curran Associates Inc. 2012.
6. Moeskops P, Viergever MA, Mendrik AM, et al. Automatic segmentation of MR brain images with a convolutional neural network. IEEE Trans Med Imaging 2016; 35:1252–1261.
7. Tajbakhsh N, Shin JY, Gurudu SR, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans Med Imaging 2017;doi:10.1109/TMI.2016.2535302.
8. Bar Y, Diamant I, Wolf L, et al. Chest pathology detection using deep learning with non-medical training In 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI). 2015, 294–297.
9. Huynh BQ, Li H, Giger ML. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. J Med Imaging (Bellingham) 2016; 3:034501.
10. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge, MA: The MIT Press, 2016.
11. Campbell M, Hoane AJ, Hsu F-H. Deep blue. Artif Intell 2002; 134:57–83.
12. Deep learning machine teaches itself chess in 72 hours, plays at international master level. Available at: https://www.technologyreview.com/s/541276/deep-learning-machine-teaches-itself-chess-in-72-hours-plays-at-international-master/#comments. Accessed June 7, 2017.
13. Koch C. How the computer beat the Go master Nature America, Inc. 2016.
14. Gibney E. Go players react to computer defeat. 2016.
15. Perez S. Google's smarter, A.I.-powered translation system expands to more languages AOL Inc. 2017.
16. French K. Your New Best Friend: AI Chatbot. Futurism. Available at: https://futurism.com/ai-chatbot-meaningful-conversation/; 2011.
17. Coldewey D. This neural network “hallucinates” the right colors into black and white pictures TechCrunch. 2016.
18. Sebastiani F. Classification of text, automatic. In: Brown K, ed. Encyclopedia of language & linguistics, Oxford: Elsevier; 2006.
19. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016; 316:2402–2410.
20. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017; 542:115–118.
21. Liu Y, Gadepalli K, Norouzi M, et al. Detecting cancer metastases on gigapixel pathology images arXiv. 2017.
22. Bar Y, Diamant I, Wolf L, et al. Deep learning with non-medical training used for chest pathology identification In SPIE Medical Imaging. SPIE, 9414, 7 .
23. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology 2017; 284:162326.
24. Cheng JZ, Ni D, Chou YH, et al. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in CT scans. Sci Rep 2016; 6:24454.
25. Kooi T, Litjens G, van Ginneken B, et al. Large scale deep learning for computer aided detection of mammographic lesions. Med Image Anal 2017; 35:303–312.
26. Wang J, Yang X, Cai H, et al. Discrimination of breast cancer with microcalcifications on mammography by deep learning. Sci Rep 2016; 6:27327.
27. Hua KL, Hsu CH, Hidayati SC, et al. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. Onco Targets Ther 2015; 8:2015–2022.
28. Rajkomar A, Lingam S, Taylor AG, et al. High-throughput classification of radiographs using deep convolutional neural networks. J Digit Imaging 2016: 1–7.
29. Wang C, Elazab A, Wu J, et al. Lung nodule classification using deep feature fusion in chest radiography. Comput Med Imaging Graph 2017; 57:10–18.
30. van Ginneken B. Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning. Radiol Phys Technol 2017; 10:23–32.
31. Armato 3rd SG, Altman MB, Wilkie J, et al. Automated lung nodule classification following automated nodule detection on CT: a serial approach. Med Phys 2003; 30:1188–1197.
32. Dou Q, Chen H, Yu L, et al. Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection. IEEE Trans Biomed Eng 2017; 64:1558–1567.
33. Nibali A, He Z, Wollersheim D. Pulmonary nodule classification with deep residual networks. Int J Comput Assist Radiol Surg 2017;doi:10.1007/s11548-017-1605-6. LID.
34. Ciompi F, Chung K, van Riel SJ, et al. Towards automatic pulmonary nodule management in lung cancer screening with deep learning. Sci Rep 2017;doi:10.1038/srep46479.
35. Kim N, Seo JB, Lee Y, et al. Development of an automatic classification system for differentiation of obstructive lung disease using HRCT. J Digit Imaging 2009; 22:136–148.
36. Anthimopoulos M, Christodoulidis S, Ebner L, et al. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE Trans Med Imaging 2017; 35:1207–1216.
37. Lakhani P. Deep convolutional neural networks for endotracheal tube position and x-ray image classification: challenges and opportunities. J Digit Imaging 2017;doi:10.1007/s10278-017-9980-7. LID.
38. Wang J, Kato F, Yamashita H, et al. Automatic estimation of volumetric breast density using artificial neural network-based calibration of full-field digital mammography: feasibility on Japanese women with and without breast cancer. J Digit Imaging 2016: 1–13.
39. Lee H, Tajmir S, Lee J, et al. Fully automated deep learning system for bone age assessment. J Digit Imaging 2017: 1–15.
40. Cheng PM, Malhi HS. Transfer learning with convolutional neural networks for classification of abdominal ultrasound images. J Digit Imaging 2016: 1–10.
41. Cha KH, Hadjiiski LM, Chan H-P, et al. Bladder cancer treatment response assessment using deep learning in CT with transfer learning In SPIE Medical Imaging. SPIE, 10134, 6 .

42. Chen Y, Dhar R, Heitsch L, et al. Automated quantification of cerebral edema following hemispheric infarction: application of a machine-learning algorithm to evaluate CSF shifts on serial head CTs. Neuroimage Clin 2016; 12:673–680.
43. Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. IEEE Trans Syst Man Cybern 1973; 3: 610–621.
44. Galloway MM. Texture analysis using gray level run lengths. Comput Graph Image Process 1975; 4:172–179.
45. Tixier F, Le Rest CC, Hatt M, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. J Nucl Med 2011; 52:369–378.
46. Coroller TP, Grossmann P, Hou Y, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol 2015; 114:345–350.
47. Akkus Z, Galimzianova A, Hoogi A, et al. Deep learning for brain MRI segmentation: state of the art and future directions. J Digit Imaging 2017; doi:10.1007/s10278-017-9983-4. LID.
48. Gaonkar B, Hovda D, Martin N, et al. Deep learning in the small sample size setting: cascaded feed forward neural networks for medical image segmentation In SPIE Medical Imaging. SPIE, 9785, 8 .
49. Cheng R, Roth HR, Lu L, et al. Active appearance model and deep learning for more accurate prostate segmentation on MRI In SPIE Medical Imaging. SPIE, 9784, 9 .
50. Hammana I, Lepanto L, Poder T, et al. Speech recognition in the radiology department: a systematic review. Health Inf Manag 2015; 44:4–10.
51. Deng L, Li X. Machine learning paradigms for speech recognition: an overview. IEEE Trans Audio Speech Lang Process 2013; 21:1060–1089.
52. Launchbury J. A DARPA perspective on artificial intelligence. 2017.
53. Moosavi-Dezfooli S-M, Fawzi A, Fawzi O, et al. Universal adversarial perturbations. 2016.
54. Summers RM. Progress in fully automated abdominal CT interpretation. AJR Am J Roentgenol 2016; 207:67–79.
55. Wang S, Summers RM. Machine learning and radiology. Med Image Anal 2012; 16:933–951.
56. Mikolas P, Melicher T, Skoch A, et al. Connectivity of the anterior insula differentiates participants with first-episode schizophrenia spectrum disorders from controls: a machine-learning study. Psychol Med 2016; 46:2695–2704.
57. Collij LE, Heeman F, Kuijer JPA, et al. Application of machine learning to arterial spin labeling in mild cognitive impairment and Alzheimer disease. Radiology 2016; 281:865–875.
58. Bryan RN. Machine learning applied to Alzheimer disease. Radiology 2016; 281:665–668.
59. Motwani M, Dey D, Berman DS, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. Eur Heart J 2016; 38:500–507. ehw188.
60. Gaonkar B, Macyszyn L, Bilello M, et al. Automated tumor volumetry using computer-aided image segmentation. Acad Radiol 2015; 22:653–661.