

Deep Learning Applications in Chest Radiography and Computed Tomography

Current State of the Art

Sang Min Lee, MD,* Joon Beom Seo, MD,* Jihye Yun, PhD,†

Young-Hoon Cho, MD,* Jens Vogel-Claussen, MD,‡

Mark L. Schiebler, MD,§ Warren B. Geftter, MD,|| Edwin J.R. van Beek, MD,¶

Jin Mo Goo, MD,# Kyung Soo Lee, MD,** Hiroto Hatabu, MD,††

James Gee, PhD,‡‡ and Namkug Kim, PhD*†

Abstract: Deep learning is a genre of machine learning that allows computational models to learn representations of data with multiple levels of abstraction using numerous processing layers. A distinctive feature of deep learning, compared with conventional machine learning methods, is that it can generate appropriate models for tasks directly from the raw data, removing the need for human-led feature extraction. Medical images are particularly suited for deep learning applications. Deep learning techniques have already demonstrated high performance in the detection of diabetic retinopathy on fundoscopic images and metastatic breast cancer cells on pathologic images. In radiology, deep learning has the opportunity to provide improved accuracy of image interpretation and diagnosis. Many groups are exploring the possibility of using deep learning-based applications to solve unmet clinical needs. In chest imaging, there has been a large effort to develop and apply computer-aided detection systems for the detection of lung nodules on chest radiographs and chest computed tomography. The essential limitation to computer-aided detection is an inability to learn from new information. To overcome these deficiencies, many groups have turned to deep learning approaches with promising results. In addition to nodule detection, interstitial lung disease recognition, lesion segmentation, diagnosis and patient outcomes have been addressed by deep learning approaches. The purpose of

this review article was to cover the current state of the art for deep learning approaches and its limitations, and some of the potential impact on the field of radiology, with specific reference to chest imaging.

Key Words: chest imaging, machine learning, deep learning, radiography, computed tomography, magnetic resonance imaging

(*J Thorac Imaging* 2019;34:75–85)

Deep learning is a genre of machine learning that allows computational models to learn representations of data with multiple levels of abstraction through the use of a number of unique processing layers.¹ The most distinctive feature of deep learning, compared with the conventional machine learning methods, is that it can extract fully automated features and generate appropriate models for tasks directly from the raw data on its own, removing the need for human-led feature extraction.

In recent years, deep learning methods have shown breakthroughs in various fields including image recognition,² speech recognition,³ as well as information technology. However, in the medical field, the application of deep learning is currently in its infancy. Medical images and their respective patient electronic medical records are well suited for analysis by deep learning. Some of the first successful demonstrations of deep learning techniques were reported in the detection of lymph node metastasis from hematoxylin and eosin-stained pathologic micrographs, analysis of skin cancer from photographs of the lesion, and the diagnosis of diabetic retinopathy from fundoscopic images.^{4–7}

In radiology, deep learning will help to improve efficiency by automated image interpretation and generation of an appropriate differential diagnosis. Data mining of the patient's electronic medical record data (big data) combined with deep learning applied to the patient's medical images should help to improve patient outcomes. Cloud-based applications also allow the deep learning algorithm to continuously learn on data sets that are not restricted to a single institution. Many groups are now exploring deep learning-based applications for solutions to unmet clinical needs. In chest imaging, significant effort has been directed at developing and applying computer-aided detection (CAD)⁸ systems for the detection of nodules on chest radiographs and chest computed tomography (CT).^{9,10} Although many CAD systems are being used in clinical practice, the implementation of CAD has not been widely accepted

From the *Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center; †Department of Convergence Medicine, University of Ulsan College of Medicine, Asan Medical Center; #Department of Radiology, Seoul National University College of Medicine, and Institute of Radiation Medicine, Seoul National University Medical Research Center; **Department of Radiology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea; ‡Institute of Diagnostic and Interventional Radiology, German Center for Lung Research, Hannover Medical School, Hannover, Germany; §Department of Radiology, University of Wisconsin-Madison School of Medicine and Public Health, Madison, WI; ||Department of Radiology, University of Pennsylvania Perelman School of Medicine; ‡‡Penn Image Computing and Science Laboratory, Department of Radiology, University of Pennsylvania, Philadelphia, PA; ¶Department of Radiology, University of Edinburgh, and Edinburgh Imaging, Queen's Medical Research Institute, Edinburgh, Scotland, UK; and ††Department of Radiology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

Edwin J.R. van Beek: Advisory boards: Imbio, Aidence; Owner/founder: Quantitative Clinical Trials Imaging Services. The remaining authors declare no conflicts of interest.

Correspondence to: Joon Beom Seo, MD, Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, 88 Olympic-ro 43 Gil, Songpa-gu, Seoul 138-736, South Korea (e-mail: seobj@amc.seoul.kr).

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.
DOI: 10.1097/RTI.0000000000000387

due to its poor performance (ie, frequent false-positive and false-negative cases). Deep learning approaches have the potential to overcome the limitations of existing CAD systems, with several studies showing promising results.^{11,12} Moreover, disease pattern recognition, lesion segmentation, diagnosis, and survival prediction have been successfully studied using deep learning in chest imaging.¹³ However, there is still concern on this technology in terms of clinical application.

In this review article, we introduce the principle methods of deep learning, their potential applications, and clinical promise in chest imaging.

DEEP LEARNING AND CONVOLUTIONAL NEURAL NETWORKS (CNNs)

Machine learning is defined as a set of methods that can automatically detect patterns in data, and then utilize the uncovered patterns to classify, predict, or conduct various types of decision making under uncertain conditions.¹⁴ Conventional machine learning techniques rely on extensive data engineering and considerable domain expertise to design a “feature extractor” algorithm that converts the raw data into suitable representations for computational analysis. A CNN is a special type of deep learning and is quite similar to the overall learning process (eg, neuronal pruning) of the mammalian visual cortex,¹⁵ and it is responsible for the recent improvements in the field of computer vision (eg, self-driving automobiles). With the availability of large data sets and increased computing power, CNNs have produced promising results for many tasks including image classification, correct image detection, correct image segmentation, and understanding speech (eg, natural language processing).

The architecture of a CNN can be composed of convolutional, pooling, and fully connected layers (Fig. 1): (1) the convolutional layers detect distinctive local motifs by applying multiple filters and generating multiple feature maps, (2) the pooling layers effectively reduce the

dimensions of feature maps (other techniques such as dilated convolutions and convolutions with stride can also be used), (3) the fully connected layers integrate all feature responses and eventually project them onto an output layer, which serves to answer the task at hand. By using deep CNN architecture (repeating the convolutional and pooling layers several times) to mimic the natural neuromorphic multilayer network, deep learning can automatically and adaptively learn a hierarchical representation of patterns and consequently identify the most significant features for a given task.² In order to deal with complex tasks, networks with many layers—so called Deep Networks—are required. However, adding additional layers increases the number of parameters in the model and can make it more difficult to train it for a specific task without overfitting the data.

Classification

One key task for radiologists is creating an appropriate differential diagnosis for each patient's medical images. This job can be computationally defined as a typical classification task using input from medical images and any available clinical information. There are many different CNN network architectures for classifying images. In order to improve the efficiency of the training procedure and reduce the number of parameters, the deeper networks have introduced more effective subroutines—“building blocks.” These building blocks are small branching/spanning convolution blocks with pooling and batch normalization layers, which can be repeated to construct deeper architectures.¹⁶ VGG19¹⁷ used small and fixed size kernels in each layer to win the ImageNet challenge of 2014. Another CNN named GoogLeNet (ie, “Inception”)¹⁸ made use of the building block that is a multilevel feature extractor with a set of convolutions of different sizes. ResNet,¹⁹ which won the ImageNet challenge of 2015, introduced the subroutine of a “residual building block,” which was designed to learn the residual (eg, features that remain important) in order to make it easier to train deeper neural networks. This residual block was implemented by

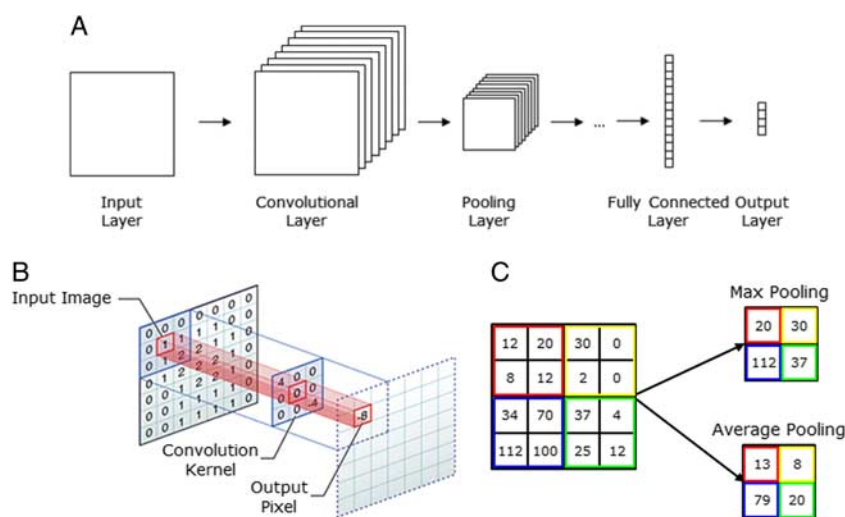


FIGURE 1. Architecture of convolutional neural network (CNN): (A) A CNN is comprised of one or more convolutional layers (often with a pooling layer) and then followed by one or more fully connected layers, (B) the convolutional layers compute the sum of the element-wise multiplication between the input image and multiple filters (ie, kernel) to detect distinctive local motifs, (C) the pooling layers, which are a form of nonlinear downsampling (eg, max and average pooling), serve to progressively reduce the spatial size of the representation and reduce the number of parameters and amount of computation in the network; this process helps to limit overfitting of the data.

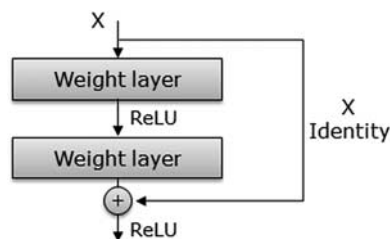


FIGURE 2. By a skip connection (ie, identity mapping) and element-wise addition, a residual block makes it easier to train a deeper network without extra parameter and computational complexity.

adding the input of the block to the output of the layers within the block (Fig. 2). Since 2014, the performance on the ImageNet benchmark has saturated, but the use of these architectures remains popular for medical image processing.

Moreover, these days, various combinations of these building blocks are used to construct deep learning architectures for the desired tasks, instead of relying on the bare previously proposed architecture.

Segmentation

The innovations of object classification have now shifted to semantic segmentation. This is a common task for both natural and medical image analysis, whereby each voxel is classified in an image to determine the boundary conditions that define a specific object. The fully convolutional network (FCN)²⁰ represents a critical breakthrough for deep learning-based semantic segmentation. In an FCN, the fully connected layers in the standard CNNs are replaced by convolutions with large receptive fields. This method achieves this segmentation by using coarse class score maps obtained by feed-forwarding an input image. U-Net,²¹ which is the most well-known segmentation architecture in medical image analysis, combines an equal amount of upsampling and downsampling layers with skip connections between opposing convolution and deconvolution layers. Mask region-based CNN (R-CNN)²² detects objects in an image while simultaneously generating segmentation masks for each instance. This method has achieved the state-of-the-art performance on Microsoft Common Objects in Context.²³

Detection

The detection of objects of interest (ie, lesions) is a key part of diagnosis and is one of the most labor-intensive tasks for radiologists. Several CNN network architectures have been shown to be able to detect a variety of objects quickly and accurately. R-CNN²⁴ combines region proposals (from a defined set of candidate detections) with CNNs and has then been improved to Fast R-CNN²⁵ and Faster R-CNN²⁶ with better performance. There are a few methods that computationally approach the problem of image detection by using multivariate regression; 2 of the most popular CNNs are You only look once (YOLO)²⁷ and single shot multibox detector (SSD),²⁸ successfully predicting bounding boxes and classification probability.

Generative Model

Generative adversarial networks (GANs) have the advantage of automatically producing new images (eg, synthetic image data) similar to samples from the training set by using 2 competing CNNs, wherein one generates

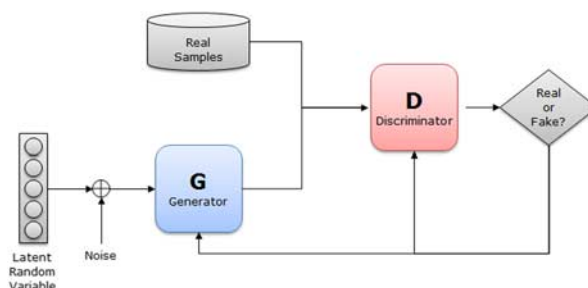


FIGURE 3. Generative adversarial networks consist of a generator and a discriminator, wherein the generator aims to generate samples (synthetic images) that resemble those in the training data while the discriminator tries to distinguish between the 2. [full color online](#)

artificial samples and the other discriminates artificial from real samples²⁹ (Fig. 3). These GANs could be trained end-to-end and learn representative features in a completely unsupervised manner. The representations learned by GANs are utilized in various applications including medical image syntheses,^{30,31} image normalization,^{32,33} and super-resolution.^{34,35}

APPLICATIONS IN CHEST IMAGING

Chest Radiography

Chest radiography is the most commonly performed diagnostic imaging procedure, and over 35 million chest radiographs are performed each year in the United States alone, and the average radiologist reads > 100 chest radiograph examinations per day.³⁶ Although these examinations are clinically useful, efficient, and cost-effective, chest radiography consists of complex 3-dimensional anatomic information condensed in a 2-dimensional projection. Accurate interpretation of chest radiographs requires a great deal of experience and medical knowledge on the part of the radiologist. Increased radiologist work-loads combined with the intrinsic challenges of interpreting chest radiographs is associated with considerable interreader and intrareader variability, missed lesions, and reporting delays in today's medical practice.⁸ Deep learning technology has the potential to automatically detect abnormalities or assist radiologists in reading chest radiographs. Such technology would be very attractive for rural areas with few radiologists as well as for state-of-the-art medical centers to help support high-volume workflows and improve the efficiency of the radiology departments.¹²

Lung Nodule Detection

Lung nodule detection from chest radiographs is another promising area for the application of deep learning technology. Lung cancer is the leading cause of cancer-related death worldwide, and chest radiography has been the most widely adopted screening and imaging tool to detect lung cancer. However, unfortunately, due to the confounding effects of anatomic complexity on chest radiographs, lung cancer screening using plain chest film has yielded unsatisfying results, with reports of missed nodules being as high as 40%.^{37,38}

CAD systems have been developed to help radiologists detect lung nodules. Recently, this method has shown a sensitivity of 71% with 1.3 false-positive CAD marks per image.³⁹

Using bone-suppressed dual-energy chest radiographs, another stand-alone CAD system achieved a sensitivity of 74% with a 1.0 false-positive CAD mark per image.⁹ In the setting of follow-up of patients with previous cancer (of any type), another CAD system showed promise by improving sensitivity from 63% to 92% for detection of lung nodules while only slightly decreasing specificity (from 98% to 96%).⁴⁰ Although there has been improvement in CAD nodule detection on chest radiograph, these methods still need better accuracy before they are routinely accepted.

Recently, CAD systems using deep learning techniques have shown improved accuracy for nodule detection on chest radiograph. A deep learning-based technique identified by Wang and colleagues extracted deep learning features by transfer learning and combined them with traditional hand-crafted features. This CAD system achieved a higher sensitivity (69.3%) for nodule detection at a significantly lower false-positive rate (1.19 false-positive marks per examination).⁴¹ There is a recent report using CNN, with visual attention networks generating an accuracy of 0.76 for nodule detection and an accuracy of 0.65 for nodule localization on chest radiographs.⁴²

Diagnosis of Tuberculosis

Another specific field of research with great potential benefit to public health is utilizing deep learning technology for the diagnosis of pulmonary mycobacterium tuberculosis (TB) on the basis of chest radiography. TB is an infectious disease caused by the acid fast (ie, outer capsule appears red on hematoxylin and eosin staining) bacillus mycobacterium tuberculosis. In western countries, it is often thought of as a “disease of the past,” but TB is still a major health problem in the developing world, with millions of new cases encountered every year.

Although the diagnosis of TB can be confirmed by bacteriology or by the whole blood gamma interferon release assay (Quantiferon Gold, Qiagen), chest radiography is a highly sensitive imaging tool for triaging and screening for

current and previous pulmonary TB infection. This organism is very difficult to culture in vitro, and it is often not possible to confirm it bacteriologically.⁴³ In locations where the prevalence of TB is high, there are a limited number of experienced chest radiologists available to use chest radiography as a method to confirm this disease. This shortage impairs screening efficacy and limits the opportunity to start medical therapy for a complete recovery.⁴⁴ Therefore, considerable effort to develop CAD systems for the detection of pulmonary TB on chest radiographs has been extended. Traditional CAD systems without deep learning technology have shown acceptable TB detection performance with an area under the curve ranging from 0.71 to 0.84.⁴⁵

Recently, Lakhani and Sundaram¹² reported the performance of CAD using CNN for detection of pulmonary TB, and, in that study, the CAD system reached an AUC of 0.99, which is greater than any previously reported CAD system. Although external validation is still needed to determine the true clinical benefit, CNN-based CAD is a feasible and promising approach in this clinical scenario.

Multiple Abnormal Pattern (MAP) Detection

Although the detection of lung nodules and TB on chest radiographs has gained attention for many deep learning researchers, these findings can be relatively rare. Each chest radiograph may contain many abnormalities, for example, pneumonia, pleural effusions, pneumothorax, medical devices, and cardiomegaly (Figs. 4, 5). Therefore, facilitating deep learning technology to detect MAP, rather than concentrating simply on nodules or TB, would be more clinically practical.

The emergence of deep learning has drastically improved the performance of machine learning for object recognition, detection, and localization when compared with previous methodologies. Critical to the success of these methods are well-annotated (strongly labeled) large data sets for effective system training. Recently, 2 large data sets of chest radiographs, Open-I and chest x-ray 14, consisting of > 110,000 chest radiographs from 30,805 patients, have been

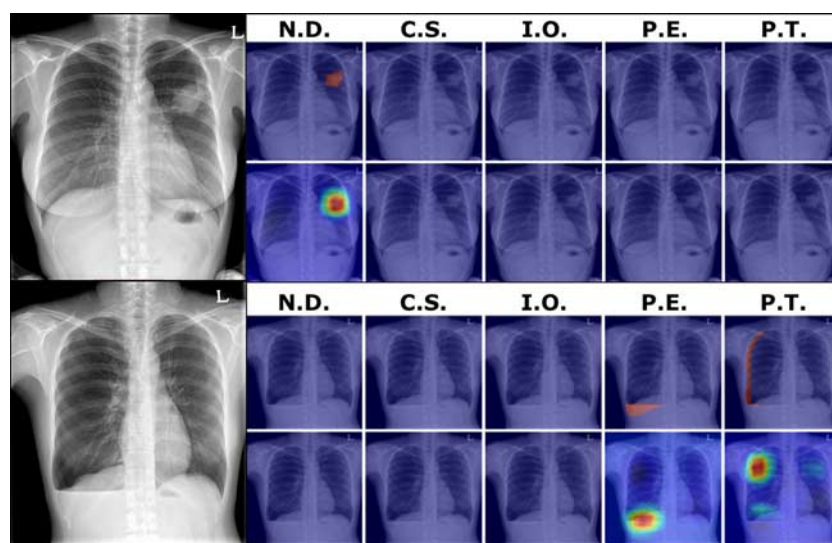


FIGURE 4. A model for detecting 5 kinds of pulmonary abnormalities (including nodule [ND], consolidation [CS], interstitial opacity [IO], pleural effusion [PE], and pneumothorax [PT]) on chest radiograph with weak labeled data, which indicate the presence of abnormalities' labels only. Upper rows of each case depict the regions of interest labeled by radiologists, and lower rows show the class activation map (CAM). Localization of trained abnormal patterns through CAM could assist radiologists to diagnose lung diseases much easier. [full color online](#)

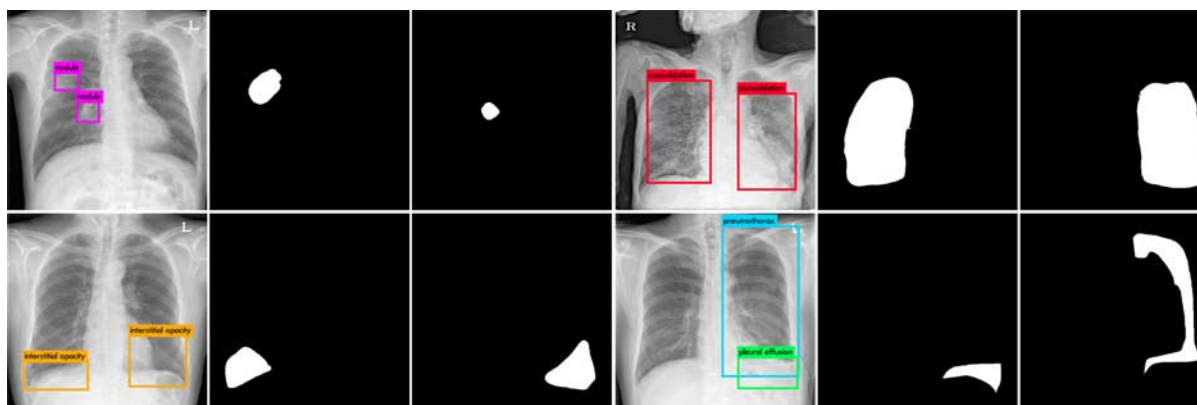


FIGURE 5. Detection of multiple abnormal lesions on chest radiograph. A CNN model was trained with strongly labeled data, which indicates not only the type of abnormalities, but also their locations and boundaries. Multiple lesions were detected in the whole-lung images, and detected regions matched with those of interest delineated by radiologists. This approach could assist radiologists to diagnose and monitor multiple lesions of whole lung. [full color online](#)

publicly released and have attracted considerable attention in the deep learning community. These publicly available data are external validation sets for any deep learning application using chest radiographs.

In 2017, Wang and colleagues trained various known CNN models to detect 8 abnormal patterns (atelectasis, cardiomegaly, effusion, infiltration, mass, nodules, pneumonia, and pneumothorax) on chest radiographs and achieved accuracy ranging from 0.56 to 0.78.⁴⁶ Another study by Cicero and colleagues reported that, in a retrospective analysis of 35,038 chest radiographs from a single medical center using CNN (GoogLeNet), they were able to obtain MAP classification accuracy of 0.88.¹¹ MAP detection in CXR with deep learning technology is still an area of active ongoing research, and different methodologies are being tested and validated, and overall accuracy will likely improve.

Chest CT

Unlike chest radiographs, chest CT provides cross-sectional images, allowing for direct 3-dimensional visualization of anatomic structures. Chest CT has a much higher sensitivity and lower interreader variability for detection of lung abnormalities and is frequently utilized in the diagnosis and follow-up of most pulmonary diseases. In addition, enhanced clinical availability, decreased cost, reduced radiation dose, and overall technical improvements of CT machines have resulted in a progressive increase in numbers of CT examinations performed each year. Therefore, an effective CAD system for chest

CT interpretation would promote overall workflow for radiologists, by reducing the time required to read each CT exam and enhance reading accuracy.

Nodule Detection/Screening

Accurate nodule detection on chest CT has become a recent point of emphasis for efficient lung cancer screening. Despite advances in cancer treatment and screening programs, most lung cancer patients are still initially diagnosed at an advanced stage of the disease, which is associated with a <20% 5-year survival.⁴⁷

Since the National Lung Screening Trial (NLST) announced a significant improvement (20%) in lung cancer mortality in high-risk populations when screened with low-dose chest CT (LDCT),⁴⁸ LDCT for cancer screening has been widely accepted.⁴⁹ Potentially this will lead to an increased number of LDCT, which will require expert analysis from a radiologist for the detection and classification of nodules into either benign or malignant diagnoses.

A CAD system could aid radiologists in both detection and classification of lung nodules (Fig. 6). Although traditional CAD systems have provided solid results, they often consist of complex pipelines of algorithms that depend heavily on manual human input such as preprocessing, segmentation, feature extraction, and model training, potentially hindering their performance.⁵⁰ Application of deep learning technology, on the other hand, can potentially remove innate challenges in traditional CAD systems by providing seamless feature

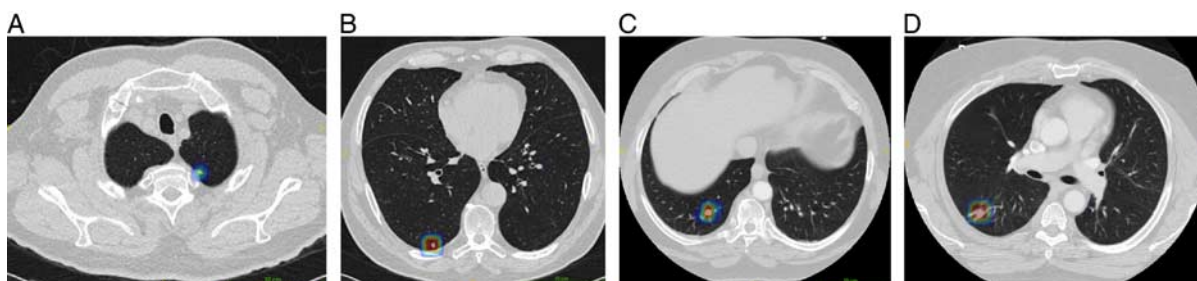


FIGURE 6. A method for detecting multiscale nodules. By training with RGB color images that were comprised of 3 adjacent slices in the axial plane, we detected nodules of various sizes. A–D, Depict the detected nodules with sizes of 3.4, 5.6, 9.8, and 14.4 mm, respectively. [full color online](#)

identification and classification and removing the need for complex human-led feature extraction pipelines.

In 2011, the Lung Image Database Consortium (LIDC) database, containing 1018 cases of thoracic CT scans and image annotations by 4 thoracic radiologists, was released and has motivated deep learning researchers to develop CAD systems for chest CT nodule detection and classification.⁵¹ CNNs are the most commonly utilized deep learning technology for lung nodule detection on CT images, and they achieve good nodule detection sensitivity while maintaining an acceptable false-positive rate. The first report of a CAD system with deep learning technology for lung nodule detection on CT was that of Hua and colleagues in 2015, achieving a sensitivity of 73% and a specificity of 80%, which was superior to any other available conventional CAD system.⁵² In 2016, Setio and colleagues trained CNN to detect pulmonary nodules and achieved 85.4% sensitivity with only one false-positive lesion per scan.⁵³ Studies that are more recent have shown the ability of CNNs to boost nodule detection sensitivity on CT to a higher level (95%) but were associated with a wide range (1.17 to 22.4) of false-positive rates.^{54–56}

Classification of detected lung nodules is also a potential area that could benefit from the use of CAD systems. CT characteristics of a lung nodule, mainly nodule type and size, are closely associated with the likelihood of malignancy. These CT features are important determinants for planning treatment and follow-up strategy. However, there is considerable observer variability in the classification of pulmonary nodules among radiologists, and this can lead to redundant follow-up examinations, unnecessary invasive procedures, or neglected malignancy.⁵⁷ In 2017, Ciompi and colleagues introduced a deep learning system that achieved good performance for nodule-type classification based on lung-RADS system and was even within the interobserver variability among 4 experienced human readers.⁵⁸ Furthermore, one study found that nodule classification accuracy of the CAD system was improved by combining deep residual learning, curriculum learning, and transfer learning.⁵⁹ Other studies using different CNN models have achieved a classification accuracy as high as 87.1%.^{60,61}

Interstitial Lung Disease (ILD)

ILD pattern classification is another area of research for deep learning technology. ILD is characterized by progressive fibrosis or inflammation of lung tissue and eventual deterioration of respiratory function.⁶² Accurate diagnosis of ILD presents a challenge for the multidisciplinary medical panel at each institution that cares for these disorders because most ILDs have similar clinical manifestations, despite the fact that they are a histologically heterogeneous group of diseases with distinct prognoses.

High-resolution CT is currently the diagnostic imaging tool of choice for the diagnosis and evaluation of ILDs. However, ILDs have similar appearance on CT, and CT readings are prone to high interobserver and intraobserver variability.⁶³ Therefore, automatic identification and classification of different ILD patterns on chest CT may be helpful even for experienced chest radiologists, and application of deep learning technology could play an eminent role in developing such CAD systems. Segmentation of the lung with ILD could be enhanced by semantic segmentation with CNN. In 2016, deep learning technology with CNN showed an accuracy of 85% for classifying 6 different ILD patterns in a data set of 14,696 image patches.⁶⁴ In 2017, Kim and colleagues compared shallow and deep learning methods on classifying 6 ILD patterns on CT; they found that deep learning methods showed significantly better accuracy, and that accuracy was further increased with the addition of more convolution layers⁶⁵ (Fig. 7). More recently, a new CNN method achieved an ILD pattern classification accuracy of 87.9% using the holistic input of the entire CT data set.⁶⁶ Moreover, CAD methodology demonstrated a prognostic ability of lung function decline using quantifiable ILD on CT studies.⁶⁷

Chronic Obstructive Pulmonary Disease

A more basic field of application for deep learning technology is the segmentation and reconstruction of organs-of-interest from chest CT scans. Organ segmentation

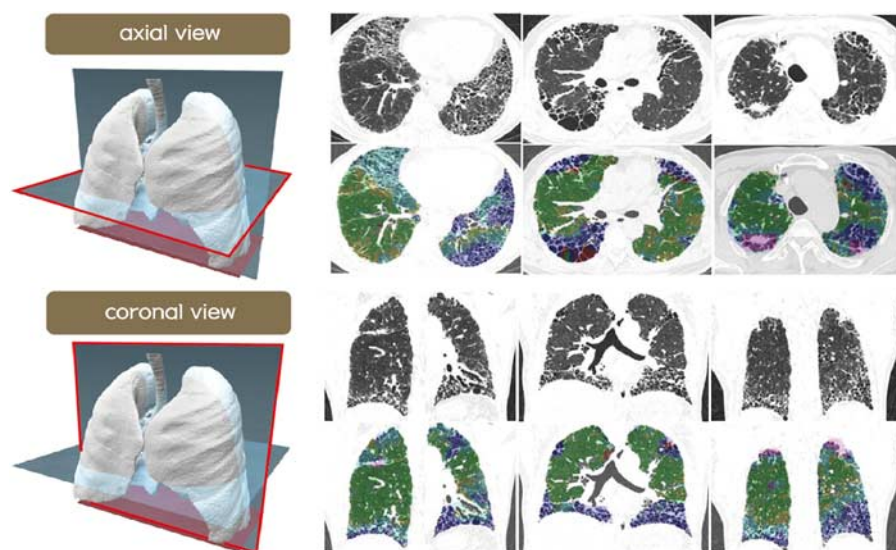


FIGURE 7. Regional image patterns of diffuse interstitial lung disease (DILD) using 3D CNN. As the diagnosis of DILD shows significant variation in interobserver and intraobserver interpretation due to a lack of standard criteria and a burden of reviewing a large amount of data, CNN-based automated classification on voxel-by-voxel basis is necessary for the quantification of disease extent and distribution of DILD.

full color
online

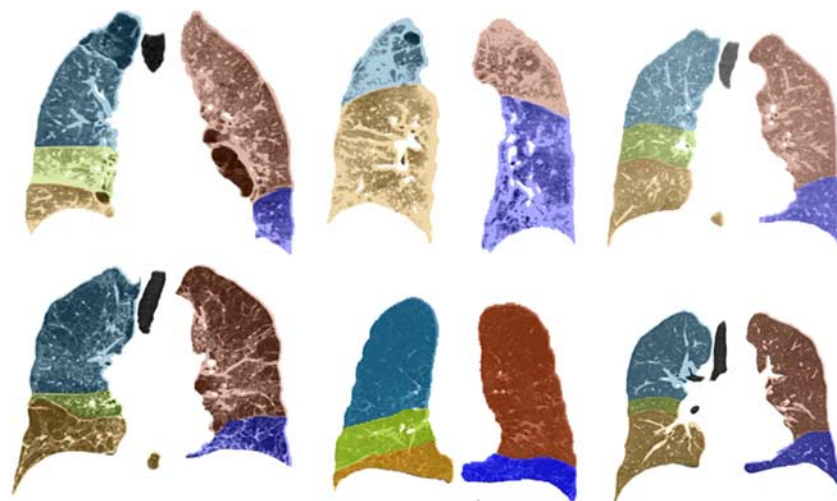


FIGURE 8. We utilized 3D U-Net (a typical type of 3D CNN) to develop a robust lobe segmentation. This approach also performed well in the fake and incomplete fissures, as this network was trained on lobe-by-lobe expert human training set. [full color online](#)

usually is the first step of many CAD systems, even those using deep learning methods, and accuracy of the segmentation process is critical because any errors in this process would affect all the subsequent analysis. Various methods for organ segmentation have been developed and tested, showing promising results, but deep learning-based models could potentially improve methodological robustness and generalizability across imaging platforms, thus providing outcomes that are more reliable.

In 2017, Harrison and colleagues developed a deep model called progressive holistically nested networks (P-HNNs) and reported that their P-HNNs model showed significant improvements in lung segmentation performance compared with previous segmentation approaches.⁶⁸ As for lobar segmentation, traditional methods are semiautomatic at best and largely rely on airway or vessel anatomy to delineate the lobar borders, with only few exceptions.⁶⁹ To address these problems, a deep learning method for lobe segmentation was introduced in 2017, and this method achieved high accuracy without reliance on prior airway or vessel segmentations, even when tested in lungs that had an underlying disease⁷⁰ (Fig. 8).

Aside from lung tissue segmentation, robust and reliable airway segmentation is also essential for quantitative evaluation of various diseases involving the airways, such as

chronic obstructive pulmonary disease (Fig. 9). A large number of prior methods have common limitations—they are substantially influenced by morphologic changes in airway trees and by measurement errors, such as airway leaks that are most prevalent at smaller (or more peripheral) airways.⁷¹ In fact, 15 different traditional algorithms were evaluated at an airway segmentation challenge in 2009 (EXACT 09), and precise delineation of a small bronchus without airway measurement leaks remained a common unsolved problem from this challenge.⁷² In 2017, a deep learning method was developed and tested using a data set from EXACT 09, and it was found that CNN significantly decreased airway leaks during segmentation process, resulting in higher sensitivity and specificity, compared with all the other algorithms that participated in the EXACT 09 challenge.⁷³ In another study, even with incompletely annotated data, 3D deep FCNs demonstrated considerable improvements in airway segmentation while maintaining acceptable quantity of airway leaks.⁷⁴

Image Normalization

The reconstruction kernel is one of the most important technical parameters that determine the trade-off between spatial resolution and image noise in CT.⁷⁵ As the selection of kernel affects the quantitative analysis,⁷⁶ CT images with different

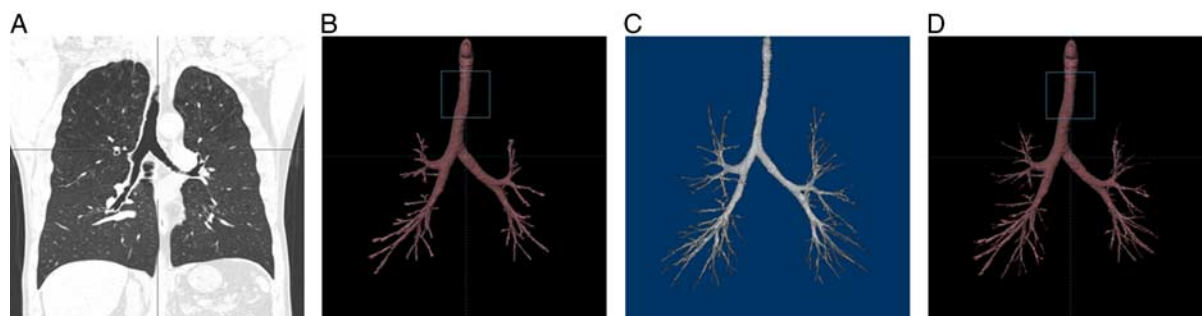


FIGURE 9. The fully automated airway segmentation method in a patient with chronic obstructive lung disease (A), which started from (B) the initial airways by using the region growing method. C, Our method achieved a high sensitivity at a low false-positive rate with fast execution time (2 to 8 min). D, Manual segmentation usually required 1 to 2 hours by an experienced research assistant. [full color online](#)

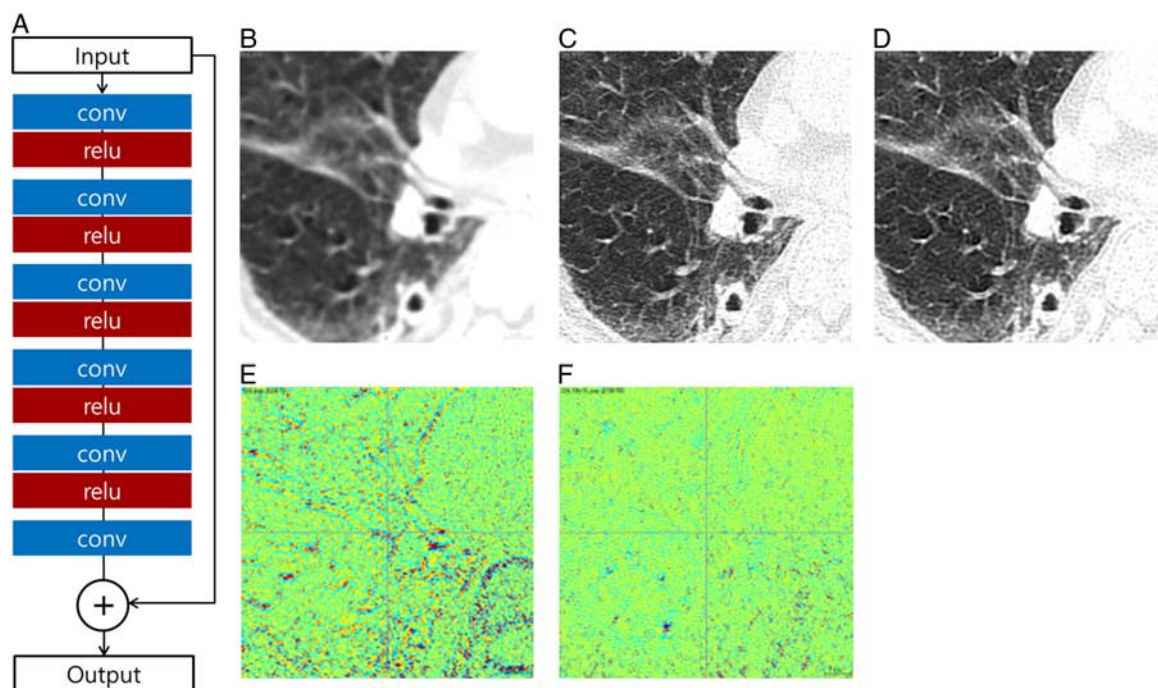


FIGURE 10. Conversion of CT images reconstructed with one kernel to images with different kernels without using a sinogram: (A) CNN architecture for CT kernel conversion. B and C, CT images reconstructed with B10f and B70f, respectively. D, A CT image interconverted from B10f to B70f using (A). E and F, Difference images between (B-C) and (B-D), respectively. [full color online](#)

reconstruction kernels are necessary for various diagnostic or quantitative purposes. To overcome the limitation of difficulty in saving the raw data before reconstruction with various kernels, postprocessing techniques have been developed to permit interconversion among CT images obtained with different kernels. Kim et al⁷⁷ recently demonstrated that CNNs could be taught differences between high-resolution and low-resolution images (residual images), and then they could be used to accurately and rapidly convert low-resolution images to high-resolution images. This approach is also applicable for interconverting CT images obtained using different kernels (Fig. 10).

Radiomics and Deep Survival

Radiomics and prediction of patient outcomes (a.k.a. “deep survival”) are also active areas of research for the application of deep learning technology. Radiomics, which has gained substantial interest from researchers around the globe, involves the high-throughput extraction of quantitative features from medical images to develop reliable models to predict genomic information, clinical outcomes, and survival.⁷⁸ Extraction of radiomics features is a critical process in radiomics research, and the majority of previous studies use hand-crafted features, which are limited by current medical knowledge and human observation. In contrast, CNN and transfer learning can be incorporated into radiomics models to extract more diverse features (deep features), which are free from prerequisite medical knowledge and biases. In this context, Lao and colleagues extracted 98,304 deep features (this would qualify as an example of overfitting of the data) from images of glioblastoma multiforme and found 6 deep features that could predict overall survival with a concordance index of 0.71.⁷⁹

In chest imaging, Paul and colleagues combined deep features of lung nodules detected on chest CT with

traditional radiomics features to predict the probability of a malignant nodule and reported an overall accuracy of 76.8% and an AUC of 0.87.⁸⁰ Another group used CNN to predict patient outcomes in a large cohort of smokers and COPD patients, and the CNN model predicted mortality with fair discrimination.¹³ Deep radiomics and deep survival are promising new fields for study.

PERSPECTIVE, CHALLENGES, AND LIMITATIONS

In this manuscript, we reviewed the basic concepts of deep learning and its various applications in chest radiography and CT. In comparison with CT, MRI is more challenging for deep learning application because there is no pulse sequence-dependent standardized intensity scale like the Hounsfield units in CT.⁸¹ The application of this new technology to radiology has barely started, but it has shown remarkable results when compared with previous studies. We believe that these improvements in performance will soon offer new possibilities for the clinical practice of radiology.

The first deep learning-based CAD application may be used to find critical findings on chest radiograph and triage the workload before a radiologist’s read. In brain CT, Prevedello et al⁸² already demonstrated that deep learning-based algorithm could automatically identify critical findings and notify the interpreting radiologist. Furthermore, if the performance of CAD can be clinically acceptable in terms of prioritization of chest radiograph, it implies that deep learning-based CAD has the potential to differentiate normal chest radiographs from grossly abnormal examinations. Thus, deep learning-based CAD should improve the workflow and efficiency of radiology departments.

Second, CAD can help in the diagnosis of diseases such as ILD and generate a preliminary quantitative report based

on CAD results. This CAD report is repeatable with the same results and has no “intrareader” variability. CAD combined with big data technology may retrieve similar images or diagnosis when radiologists require them during interpretation of CT. It can also help to reduce the reading time.

Third, automation of lesion detection, segmentation, quantification by deep learning techniques facilitates reporting of the quantitative analysis of medical images more easily. Deep learning–based segmentation tool improves accuracy and decreases image interpretation time. Furthermore, these data will likely provide an improvement in the prediction of patient outcomes and risk stratification.

However, there are still many challenges to overcome. Currently, training deep learning algorithms requires large, strongly labeled, and anonymized image data sets. These data sets are very challenging to acquire. Although some abnormalities such as pneumothorax and malpositioned lines/tubes can be based on imaging findings alone, most diseases require clinical documentation and/or pathologic confirmation. Ambiguous or overlapping radiographic terms such as “consolidation” and “infiltrate” should not be used as surrogates for pneumonia to label training cases. This has been recognized as a limitation of some publicly available data sets. National organizations (eg, ACR, RSNA) and subspecialty radiology societies can play important roles in defining appropriate tasks for deep learning algorithms, as well as assisting in making publicly available strongly labeled training data sets and validation data sets.

In contrast, the deep learning–based system should integrate information in the different domains. In real clinical practice, the differential diagnosis is greatly dependent on factors other than the imaging features of the examination at hand, such as reasons for the examination, medical history, laboratory results, etc. However, this adds a much greater complexity to the task that deep learning algorithms need to master. At present, efforts to add clinical or pathological information to imaging features using deep learning network and determine the output have been made. Suk and Shen⁸³ showed a deep learning–based method for Alzheimer Disease and Mild Cognitive Impairment diagnosis using multimodality information including MRI, PET, CSF findings, and Minimum Mental State Examination. We hope to see more advanced and flexible deep learning networks dealing with various medical data in the near future.

Furthermore, the challenges with regard to the ethical and legal aspects of data sharing and patient privacy are also paramount. There are severe monetary penalties (ie, fines) in the United States⁸⁴ for any medical facility that allows compromise of personal health information/images. In the United States, the Health Insurance Portability and Accountability Act (HIPPA) governs any use of a patient’s health information; as such, it is of critical importance that these imaging and medical data that are used for training, testing, and validation of deep learning methods are fully anonymized and comply with this law. New data protection laws have also been introduced throughout Europe. As deep learning requires an enormous amount of high-quality data, the laws governing the safe handling of medical images and medical record data need to be followed. New technology, such as Blockchain, may be helpful in guaranteeing secure data sharing.

Lastly, we should demand a thorough and systematic clinical validation of any deep learning–based applications

as a prerequisite to commercial application. A well-known problem with these methods is overfitting and lack of utility when asked to analyze other data sets (eg, poor interoperability). Most machine learning publications have shown their results in carefully preselected and enriched test sets (eg, spiked to favor that algorithm with a higher prevalence of the condition than is found clinically). Thus, beyond just determining the feasibility of using any deep learning application in a test set chosen by the author, each deep learning application should be tested by a publicly available external validation set. In addition to methods of clinical validation, the validity of the reference standard should also be carefully considered. Some abnormalities cannot be definitely diagnosed in imaging modalities such as ILD on CT. Therefore, whenever the performance of deep learning algorithm is compared with that of a human reader or other software, it is needed to clarify the task and confirm the appropriateness of the reference standard. We believe that this should be a requirement for any commercially approved deep learning method.

CONCLUSIONS

The application of deep learning methodology to help solve many tasks associated with medical imaging is at its infancy. Although there are problems with every disruptive technological innovation, we believe that deep learning will soon be an indispensable tool for radiology. This is analogous to how the picture archiving communication systems and radiology information systems have transformed medical imaging and improved radiology while at the same time decreasing the cost of medical care. Reasonable expectations for this disruptive technology are needed, along with careful attention to any ethical, legal, and regulatory issues that may arise. This technology will enable radiologists to become more productive and improve patient care. The full potential of this technology will require radiologists to have an active role in governing its successful introduction to the clinic.

REFERENCES

1. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–444.
2. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. 2012;25:1097–1105.
3. Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Proc Mag*. 2012;29:82–97.
4. 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.
5. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115–118.
6. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*. 2017;318:2199–2210.
7. Ting DSW, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*. 2017;318:2211–2223.
8. McAdams HP, Samei E, Dobbins J III, et al. Recent advances in chest radiography. *Radiology*. 2006;241:663–683.
9. Schalekamp S, van Ginneken B, Koedam E, et al. Computer-aided detection improves detection of pulmonary nodules in

- chest radiographs beyond the support by bone-suppressed images. *Radiology*. 2014;272:252–261.
10. Liang M, Tang W, Xu DM, et al. Low-dose CT screening for lung cancer: computer-aided detection of missed lung cancers. *Radiology*. 2016;281:279–288.
 11. Cicero M, Bilbily A, Colak E, et al. Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. *Invest Radiol*. 2017;52:281–287.
 12. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*. 2017;284:574–582.
 13. Gonzalez G, Ash SY, Vegas-Sanchez-Ferrero G, et al. Disease staging and prognosis in smokers using deep learning in chest computed tomography. *Am J Respir Crit Care Med*. 2018;197:193–203.
 14. Robert C. *Machine Learning, A Probabilistic Perspective*. Abington, UK: Taylor & Francis; 2014.
 15. Hubel DH, Wiesel TN. Receptive fields and functional architecture of monkey striate cortex. *J Physiol*. 1968;195:215–243.
 16. Dutta JK, Liu J, Kurup U, et al. Effective building block design for deep convolutional neural networks using search. arXiv preprint arXiv:2018;1801.08577:1–8.
 17. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:2014;1409.1556:1–14.
 18. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015:1–9.
 19. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016:770–778.
 20. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015:3431–3440.
 21. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention, Springer. 2015:3431–3440.
 22. He K, Gkioxari G, Dollár P, et al. Mask r-cnn. Computer Vision (ICCV), 2017 IEEE International Conference. 2017: 2980–2988.
 23. Lin T-Y, Maire M, Belongie S, et al. Microsoft coco: common objects in context. European Conference on computer vision, Springer. 2014:740–755.
 24. Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2014:580–587.
 25. Girshick R. Fast r-cnn. arXiv preprint arXiv:2015;1504.08083:1–9.
 26. Ren S, He K, Girshick R, et al. Faster r-cnn: towards real-time object detection with region proposal networks. *Adv Neural Inf Process Syst*. 2015;28:91–99.
 27. Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779–788.
 28. Huang Y, Liu Z, He L, et al. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. *Radiology*. 2016;281:947–957.
 29. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. Advances in neural information processing systems, 2014:2672–2680.
 30. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:2015;1511.06434:1–16.
 31. Reed S, Akata Z, Yan X, et al. Generative adversarial text to image synthesis. arXiv preprint arXiv:2016;1605.0539:1–10.
 32. Bousmalis K, Silberman N, Dohan D, et al. Unsupervised pixel-level domain adaptation with generative adversarial networks. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017:7.
 33. Tzeng E, Hoffman J, Saenko K, et al. Adversarial discriminative domain adaptation. Computer Vision and Pattern Recognition (CVPR), 2017:4.
 34. Ledig C, Theis L, Huszar F, et al. Photo-realistic single image super-resolution using a generative adversarial network. arXiv preprint. 2016;1609.04802:1–19.
 35. Sønderby CK, Caballero J, Theis L, et al. Amortised map inference for image super-resolution. arXiv preprint arXiv:1610.04490. 2016;1610.04490:1–17.
 36. Kamel SI, Levin DC, Parker L, et al. Utilization trends in noncardiac thoracic imaging, 2002-2014. *J Am Coll Radiol*. 2017;14:337–342.
 37. Finigan JH, Kern JA. Lung cancer screening: past, present and future. *Clin Chest Med*. 2013;34:365–371.
 38. Quekel LG, Kessels AG, Goei R, et al. Miss rate of lung cancer on the chest radiograph in clinical practice. *Chest*. 1999;115:720–724.
 39. Li F, Engelmann R, Armato SG III, et al. Computer-aided nodule detection system: results in an unselected series of consecutive chest radiographs. *Acad Radiol*. 2015;22:475–480.
 40. van Beek EJ, Mullan B, Thompson B. Evaluation of a real-time interactive pulmonary nodule analysis system on chest digital radiographic images: a prospective study. *Acad Radiol*. 2008;15:571–575.
 41. 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.
 42. Pesce E, Ypsilantis P-P, Withey S, et al. Learning to detect chest radiographs containing lung nodules using visual attention networks. arXiv:2017;1712.00996:1–23.
 43. Miller C, Lonnroth K, Sotgiu G, et al. The long and winding road of chest radiography for tuberculosis detection. *Eur Respir J*. 2017;49:1700364.
 44. Melendez J, Sanchez CI, Philipsen RH, et al. An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. *Sci Rep*. 2016;6:25265.
 45. Pande T, Cohen C, Pai M, et al. Computer-aided detection of pulmonary tuberculosis on digital chest radiographs: a systematic review. *Int J Tuberc Lung Dis*. 2016;20:1226–1230.
 46. Wang X, Peng Y, Lu L, et al. ChestX-Ray8: Hospital-scale chest x-Ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. arXiv:2017;1705.02315:1–19.
 47. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68:7–30.
 48. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365:395–409.
 49. Oudkerk M, Devaraj A, Vliegenthart R, et al. European position statement on lung cancer screening. *Lancet Oncol*. 2017;18:e754–e766.
 50. Goo JM. Computer-aided detection of lung nodules on chest CT: issues to be solved before clinical use. *Korean J Radiol*. 2005;6:62–63.
 51. Armato SG III, McLennan G, Bidaut L, et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med Phys*. 2011;38:915–931.
 52. Hua KL, Hsu CH, Hidayati SC, et al. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *Oncotargets Ther*. 2015;8:2015–2022.
 53. Setio AAA, Ciompi F, Litjens G, et al. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. *IEEE Trans Med Imaging*. 2016;35:1160–1169.
 54. Hamidian S, Sahiner B, Petrick N, et al. 3D Convolutional neural network for automatic detection of lung nodules in chest CT. *Proc SPIE Int Soc Opt Eng*. 2017:10134.

55. Jiang H, Ma H, Qian W, et al. An automatic detection system of lung nodule based on multi-group patch-based deep learning network. *IEEE J Biomed Health Inform.* 2017;22:1227–1237.
56. Masood A, Sheng B, Li P, et al. Computer-assisted decision support system in pulmonary cancer detection and stage classification on CT images. *J Biomed Inform.* 2018;79:117–128.
57. van Riel SJ, Sanchez CI, Bankier AA, et al. Observer variability for classification of pulmonary nodules on low-dose CT images and its effect on nodule management. *Radiology.* 2015;277:863–871.
58. Ciompi F, Chung K, van Riel SJ, et al. Towards automatic pulmonary nodule management in lung cancer screening with deep learning. *Sci Rep.* 2017;7:46479.
59. Nibali A, He Z, Wollersheim D. Pulmonary nodule classification with deep residual networks. *Int J Comput Assist Radiol Surg.* 2017;12:1799–1808.
60. Zhao X, Liu L, Qi S, et al. Agile convolutional neural network for pulmonary nodule classification using CT images. *Int J Comput Assist Radiol Surg.* 2018;13:585–595.
61. Shen W, Zhou M, Yang F, et al. Multi-crop Convolutional Neural Networks for lung nodule malignancy suspiciousness classification. *Pattern Recognit.* 2017;61:663–673.
62. Harari S, Caminati A, Madotto F, et al. Epidemiology, survival, incidence and prevalence of idiopathic pulmonary fibrosis in the USA and Canada. *Eur Respir J.* 2017;49:1602384.
63. Watadani T, Sakai F, Johkoh T, et al. Interobserver variability in the CT assessment of honeycombing in the lungs. *Radiology.* 2013;266:936–944.
64. 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.* 2016;35:1207–1216.
65. Kim GB, Jung KH, Lee Y, et al. Comparison of shallow and deep learning methods on classifying the regional pattern of diffuse lung disease. *J Digit Imaging.* 2017;31:415–424.
66. Gao M, Bagci U, Lu L, et al. Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks. *Comput Methods Biomech Biomed Eng Imaging Vis.* 2018;6:1–6.
67. Humphries SM, Yagihashi K, Huckleberry J, et al. Idiopathic pulmonary fibrosis: data-driven textural analysis of extent of fibrosis at baseline and 15-month follow-up. *Radiology.* 2017;285:270–278.
68. Harrison AP, Xu Z, George K, et al. *Progressive and Multi-path Holistically Nested Neural Networks for Pathological Lung Segmentation from CT Images.* Cham: Springer International Publishing; 2017:621–629.
69. Ross JC, San Jose Estepar R, Kindlmann G, et al. Automatic lung lobe segmentation using particles, thin plate splines, and maximum a posteriori estimation. *Med Image Comput Comput Assist Interv.* 2010;13:163–171.
70. George K, Harrison AP, Jin D, et al. *Pathological Pulmonary Lobe Segmentation from CT Images Using Progressive Holistically Nested Neural Networks and Random Walker.* Cham: Springer International Publishing; 2017:195–203.
71. Pu J, Gu S, Liu S, et al. CT based computerized identification and analysis of human airways: a review. *Med Phys.* 2012;39:2603–2616.
72. Lo P, van Ginneken B, Reinhardt JM, et al. Extraction of airways from CT (EXACT'09). *IEEE Trans Med Imaging.* 2012;31:2093–2107.
73. Charbonnier JP, Rikxoort EMV, Setio AAA, et al. Improving airway segmentation in computed tomography using leak detection with convolutional networks. *Med Image Anal.* 2017;36:52–60.
74. Jin D, Xu Z, Harrison AP, et al. *3D Convolutional Neural Networks with Graph Refinement for Airway Segmentation Using Incomplete Data Labels.* Cham: Springer International Publishing; 2017:141–149.
75. Schaller S, Wildberger JE, Raupach R, et al. Spatial domain filtering for fast modification of the tradeoff between image sharpness and pixel noise in computed tomography. *IEEE Trans Med Imaging.* 2003;22:846–853.
76. Boedeker KL, McNitt-Gray MF, Rogers SR, et al. Emphysema: effect of reconstruction algorithm on CT imaging measures. *Radiology.* 2004;232:295–301.
77. Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1646–1654.
78. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* 2014;5:4006.
79. Lao J, Chen Y, Li ZC, et al. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci Rep.* 2017;7:10353.
80. Paul R, Hawkins SH, Schabath MB, et al. Predicting malignant nodules by fusing deep features with classical radiomics features. *J Med Imaging (Bellingham).* 2018;5:011021.
81. Vovk U, Pernus F, Likar B. A review of methods for correction of intensity inhomogeneity in MRI. *IEEE Trans Med Imaging.* 2007;26:405–421.
82. Prevedello LM, Erdal BS, Ryu JL, et al. Automated critical test findings identification and online notification system using artificial intelligence in imaging. *Radiology.* 2017;285:923–931.
83. Suk HI, Shen D. Deep learning-based feature representation for AD/MCI classification. *Med Image Comput Comput Assist Interv.* 2013;16:583–590.
84. Lee SI, Krishnaraj A, Chatterji M, et al. When does a radiologist's recommendation for follow-up result in high-cost imaging? *Radiology.* 2012;262:544–549.