Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

CHAPTER

14

Artificial intelligence in radiology

Dakai Jin, Adam P. Harrison, Ling Zhang, Ke Yan, Yirui Wang, Jinzheng Cai, Shun Miao and Le Lu

Abstract

The interest in artificial intelligence (AI) has ballooned within radiology in the past few years primarily due to notable successes of deep learning. With the advances brought by deep learning, AI has the potential to recognize and localize complex patterns from different radiological imaging modalities, many of which even achieve comparable performance to human decision-making in recent applications. In this chapter, we review several AI applications in radiology for different anatomies: chest, abdomen, pelvis, as well as general lesion detection/identification that is not limited to specific anatomies. For each anatomy site, we focus on introducing the tasks of detection, segmentation, and classification with an emphasis on describing the technology development pathway with the aim of providing the reader with an understanding of what AI can do in radiology and what still needs to be done for AI to better fit in radiology. Combining with our own research experience of AI in medicine, we elaborate how AI can enrich knowledge discovery, understanding, and decision-making in radiology, rather than replacing the radiologist.

Keywords: Radiology; artificial intelligence; deep learning; lesion; pulmonary; abdomen; pelvis; classification; segmentation; detection; characterization

14.1 Introduction

Computers have revolutionized the field of diagnostic and quantitative imaging and are imperative in radiology workflow nowadays. Early milestones of computer technology include imaging acquisition inventions, such as computerized tomography (CT), nuclear medicine, and magnetic resonance imaging (MRI), and the developments of digitized picture archiving and communication systems (PACSs). Significant advances in "intelligent" image analysis have been achieved in recent years with the booming of artificial intelligence (AI) technology due to the emergence of deep learning. In certain very specific and limited applications, computers are now able to perform tasks that previously only physicians could accomplish. For instance, a deep-learning empowered segmentation and classification system for optical coherence tomography achieves clinically applicable performance, that is, comparable or exceeding the performance by professional experts, on

DOI: https://doi.org/10.1016/B978-0-12-821259-2.00014-4

a range of sight-threatening retinal diseases.[1] With the appropriate integration of deep-learning technologies paired with suitable medical imaging tasks, effective, and efficient AI systems can be developed to help radiologists reduce workloads and increase accuracy and consistency. It may eventually change the radiology workflow for some tasks. Hence, a better understanding of the strengths and limitations of new technology is of great benefit for radiologists. In this chapter, we review several important medical imaging tasks for different anatomies, emphasizing applications that we have worked on in our own research. Specifically, we overview and discuss the recent AI advances in thoracic, abdominal, and pelvic applications as well as general lesion analysis, which is not limited to a specific anatomy. Various imaging modalities are included, such as X-rays, CT, and MRI. For each anatomy, we focus on introducing the tasks of detection, segmentation, and classification with AI-based methods and discuss their achievements and also what future work remains to be done. Throughout, a common thread unifies the discussion, which also undergirds our own work—clinically useful AI tools must be developed hand-in-hand with radiologists toward a shared goal of empowering the radiology field.

14.2 Thoracic applications

14.2.1 Pulmonary analysis in chest X-ray

Chest X-rays (CXRs) are the most ordered radiological scan in the United States[2] used to diagnose or screen for a variety of thoracic ailments. Given the challenges in reading CXRs, for example, low sensitives,[3] there is great impetus for AI-based tools to help or enhance interpretation. Work along this line, catalyzed by the release of the CXR14 dataset,[4] has accelerated in recent years. In this subsection, we first overview the history of large-scale CXR datasets for training AI systems. Then we outline some on-going efforts and innovations aimed at pushing forward what is possible in AI-based analysis. Finally, we discuss some challenges for future investigation.

Like all AI applications, a necessary, but not sufficient, condition for an effective AI system for CXR analysis is an extensive and curated data source. Prior to the advent of deep learning, there was a paucity of large-scale CXR datasets. The one exception was the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial[5] whose CXR screening arm includes roughly 200,000 manually annotated CXRs. Annotated disease patterns include masses and nodules along with nononcological patterns, such as opacities and pleural abnormalities. However, because the PLCO is a screening trial, disease prevalence is low. Moreover, PLCO CXRs are film radiographs that were later digitized, so they may differ in appearance from digital radiographs.

While the PLCO remains invaluable, it was collected at enormous expense by executing a multisite clinical trial. Clearly, alternative data collection strategies are needed. Fortunately, the data housed in hospital PACSs offers a preexisting source of large-scale CXR data. The CXR14 dataset[4] was the first to exploit large-scale PACS CXRs. The authors collected $\sim$110K CXRs by retrospectively mining the National Institutes of Health Clinical Center PACS. Labels for each CXR were generated by automatically text mining the accompanying radiological reports written during daily clinical workflows. Once released, the CXR14 dataset quickly became a core dataset for AI training and kicked off a trend of additional groups

releasing their own PACS-mined data, such as CheXpert,[6] MIMIC-CXR,[7] and PadChest.[8] Fig. 14.1 depicts the numbers of released CXRs from each dataset.

Mining PACS is a highly promising source of data, but the aforementioned studies all rely on natural language processing to extract labels. Apart from any errors in the text mining, radiologist reports are written by considering many other factors outside of the CXR appearance, for example, lab tests, prior scans, and patient history.[3] Thus mentioned disease patterns may not actually be present in the image and disease patterns present in the image may not actually be mentioned in the report, for example, an "unchanged" assessment. This can cause serious issues[9] and AI specialists must work hand-in-hand with clinicians to most effectively use PACS-mined data. Despite these challenges, PACS-mined data still represents the most promising source of large-scale data for CXR AI and, deployed carefully, models trained on PACS-mined data can generalize well.[10] Furthermore, enhanced data collection efforts, such as more robust evaluation subsets[6] and more ontological approaches to label extraction,[8] will continue to strengthen the value of PACS-mined data.

The most straightforward application of CXR AI is predicting the scan- or study-wise labels. This is essentially a multilabel classification problem, and many initial efforts focused on this task.[4,13,14] However, another key aim is to localize each disease pattern being predicted. This enhances explainability and is a beneficial end in and of itself. The key challenge is that CXR datasets typically only possess scan- or study-wise labels that do not specify the disease pattern's location. This means training an AI-based localizer requires using weak-supervision techniques. For the most part, CXR localizers are all built off class-activation map techniques[15] which exploits the implicit localization properties within convolutional neural networks (CNNs). Promising approaches include generating

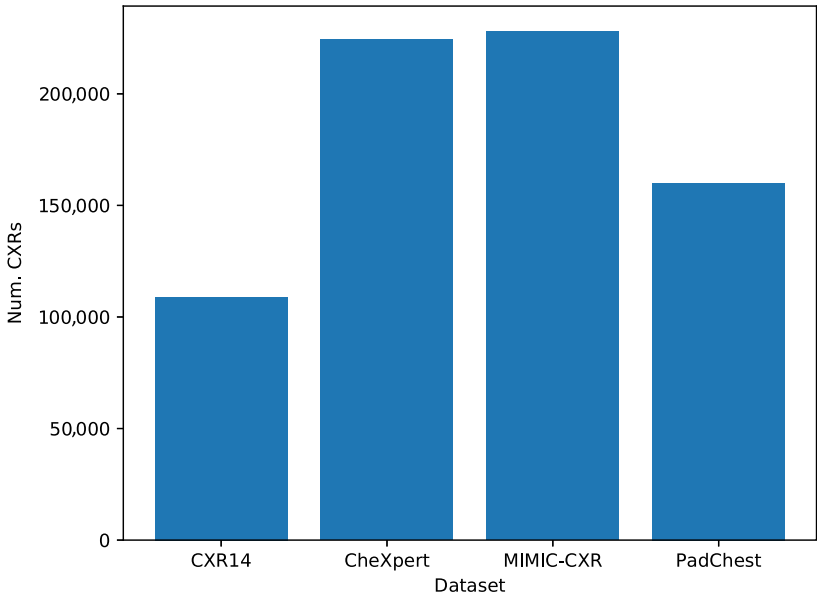


FIGURE 14.1 The plot of the numbers of publicly released CXRs by recent PACS-mined CXR datasets. *CXRs*, Chest X-rays; *PACS*, picture archiving and communication system.

pseudolabels to supervise an AI-localizer[11,12,16] and developing techniques that can work well with only a small batch of localization labels.[17] Another direction is to force the CNN to use as many regions of the image as possible when making its prediction.[18] Fig. 14.2 depicts some example localizations derived from these weakly supervised techniques. Weakly supervised localization shows promise, but challenges remain to ensure the model captures the entire extent of the disease pattern and does not focus on spurious regions.

Apart from localization, recent works have also focused on providing specialized or enhanced analyses. Mirroring larger trends within deep learning, the use of generative adversarial networks (GANs) to generate synthetic CXRs, has received attention. This includes using realistic synthetic CXRs to simulate image/mask pairs in order to train AI models to segment the lung field.[19,20] GANs have also been used to transfer a model that works well on adult patients to also perform well on pediatric data.[21] Finally, GANs have also been successfully used to flag abnormal CXRs.[16] Moving on from GAN-based analysis, another interesting line of work is using a taxonomy of disease patterns to provide both more meaningful predictions and enhanced performance.[22] As these works suggest, there is a rich set of research directions, beyond just localization, for AI applications in CXR analysis.

The release of recent PACS-mined datasets has spurred an incredibly exciting burst of research activity in CXR analysis. Already much progress has been made, but important challenges remain. One key hurdle is developing AI techniques and models that can better manage the noise and uncertainty that comes with text-mined labels. This could involve integrating clinical domain knowledge to better model the meaning behind text-mined phrases and words found in radiological reports. Relatedly, it would be extraordinarily

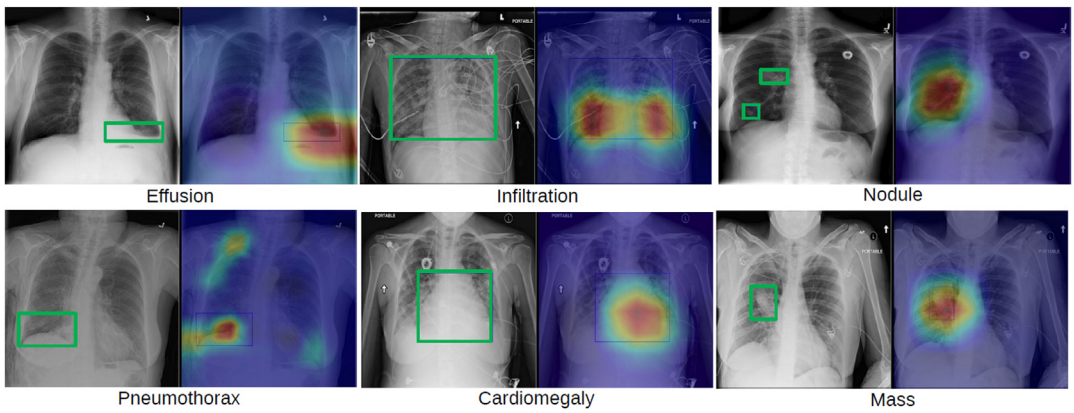


FIGURE 14.2 When properly configured, CNNs can also provide localizations indicating the region of the image that is contributing to the prediction. *CNNs*, Convolutional neural networks. *Source: Credit: Tang Y, Wang X, Harrison AP, Lu L, Xiao J, Summers RM. Attention-guided curriculum learning for weakly supervised classification and localization of thoracic diseases on chest radiographs. In: Shi Y, Suk H-I, Liu M editors.* Machine learning in medical imaging, lecture notes in computer science. *Cham: Springer International Publishing; 2018b. p. 249–58. Available from: https://doi.org/10.1007/978-3-030-00919-9_29.*

beneficial for the AI community to have agreed upon and radiologist-driven ontologies or taxonomies of disease patterns for an AI system to target. Such an ontology would also help incorporate and model the interdependencies across disease patterns. In addition, principled techniques should be developed to consider also prior CXR studies, lab results, and patient history. This would better emulate current radiological practices in the clinic. Along with these improved modeling capabilities, future work should also focus on creating larger and more accurate *manually labeled* evaluation sets, so that performance can be better gauged.

14.2.2 Pulmonary analysis in computerized tomography

CT is the gold standard imaging modality for a broad range of high prevalence pulmonary diseases, such as interstitial lung disease (ILD) and lung cancer.[23,24] Benefiting from its high spatial resolution in three dimensions (3D), CT allows more accurate disease diagnosis and quantification. To effectively detect and analyze pulmonary abnormalities from the large amounts of 3D CT data, automated AI-based tools play a critical role and have been studied for more than two decades.[25–27] In such AI-based systems, typically the first step is to segment the anatomies of interest to facilitate the later steps of disease detection and quantification. In this subsection, we first review the AI-based segmentation methods for three pulmonary anatomies, that is, lung, lobe, and airway. Then we use ILD as a case study for how AI systems can play a role in pulmonary analysis.

14.2.2.1 Lung, lobe, and airway segmentation

An often necessary first step for any computer-aided diagnosis or detection system is to accurately delineate the organs of interest. Measuring organ volume or shape can offer its own important biomarkers. In addition, accurate delineation is often a prerequisite for any downstream disease analysis, so that the area of focus can be accurately determined. Within pulmonary analysis, AI-based segmentation primarily focuses on three structures: the lungs, lung lobes, and pulmonary airways. Below, we discuss each in turn.

For normal lungs the delineation is relatively straightforward, and classic techniques such as region growing or anatomical shape models can operate well as long as their strict assumptions on Hounsfield unit intensity and shape, respectively, are maintained. However, the problem becomes much more challenging once pathological patterns are present, such as consolidations, pleural effusions, or lung nodules, or if lung shapes do not follow expected distributions. Prior to the dominance of deep learning, effective pathological lung segmentation techniques relied on sophisticated, but handcrafted workflows,[30] that can struggle to generalize without significant calibration efforts. To address this, Harrison et al.[29] proposed the first deep model for pathological lung segmentation, called progressive holistically-nested networks (PHNNs), which classified each CT voxel individually in a bottom-up manner. Tested on 929 pathological CT studies, where disease patterns associated with infections, chronic obstructive pulmonary disease (COPD) and/or ILD were present, PHNN achieved an extremely high mean Dice score, or Sørensen–Dice coefficient score of 98.5%. After Harrison et al.,[29] many subsequent works reported their own deep segmentation approaches that followed similar strategies. While the PHNN

results are impressive, the model can still struggle on scenarios it did not see enough of in training, such as lung nodules or consolidations touching the lung border. Thus further work is still required to harden CNN models, like PHNN, to such unseen variations. Jin et al.[28] proposed one such interesting strategy, using GANs to simulate lung nodules to fine-tune the PHNN model so that it can successfully handle such cases (see Fig. 14.3). The continued development of strategies along this vein will be necessary to address outlier cases as much as possible.

Delineating the five lobes of the lung is another important task, particularly as infections are often limited to one or a few lobes. While lobe segmentation shares similarities with lung segmentation, successful solutions must incorporate much more top-down structural guidance. This is because lobe fissures are often incomplete, share the same appearance with accessory fissures, and can be obscured when pathologies are present. This challenges the bottom-up voxel-by-voxel strategies used in lung segmentation.[29] Because airways do not cross lobar boundaries, one structural approach to segmentation is to first segment the airways,[32,33] using the resulting airway trees as constraints or initializations. But airway segmentation is a challenging problem in its own right, which means these approaches require complex and brittle multicomponent workflows to segment lobes. Taking a different approach, George et al.[31] reported the first deep solution to this problem. The authors trained a bottom-up PHNN model to noisily segment lung fissures and then used the random walker (RW) algorithm to impose top-down structural constraints. To keep it simple and generalizable the RW algorithm's only assumption is that there are five lobes. Their method achieves a 88.8% mean Dice score, or Sørensen–Dice coefficient score under the presence of highly challenging interstitial lung pathologies, which outperformed a leading nondeep approach[32] by 5%. Fig. 14.4 provides some qualitative examples demonstrating the power of combining bottom-up CNN predictions with straightforward top-down constraints.

Airway segmentation is uniquely challenging due to its topological complexity. The extreme thin airway wall separating the lumen and lung parenchyma adds further difficulty since its resolution is even lower than that of the CT scanner at many middle or small airway branches. This often causes large segmentation leakage into the adjacent lung parenchyma. Many

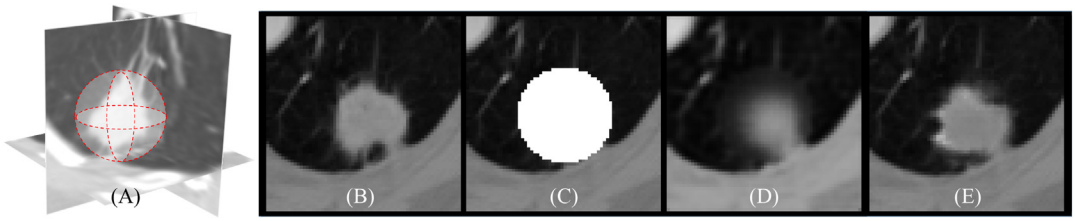


FIGURE 14.3 Jin et al.'s[28] simulated lung nodules. (A) A volume of interest centered at a lung nodule; (B) 2D axial view of (A), (C) same as (B), but with central sphere region erased; (D–E) simulated lung nodule using a competitor method and Jin et al.'s[28] method, respectively. These simulated lung nodules were used to fine-tune and enhance Harrison et al.'s[29] lung segmentation model. *Source: Credit: Jin D, Xu Z, Tang Y, Harrison AP, Mollura DJ. CT-realistic lung nodule simulation from 3D conditional generative adversarial networks for robust lung segmentation. In:* Medical image computing and computer-assisted intervention – MICCAI 2018, Lecture notes in computer science. *Cham: Springer International Publishing; 2018. p. 732–40. Available from: https://doi.org/10.1007/978-3-030-00934-2_81.*

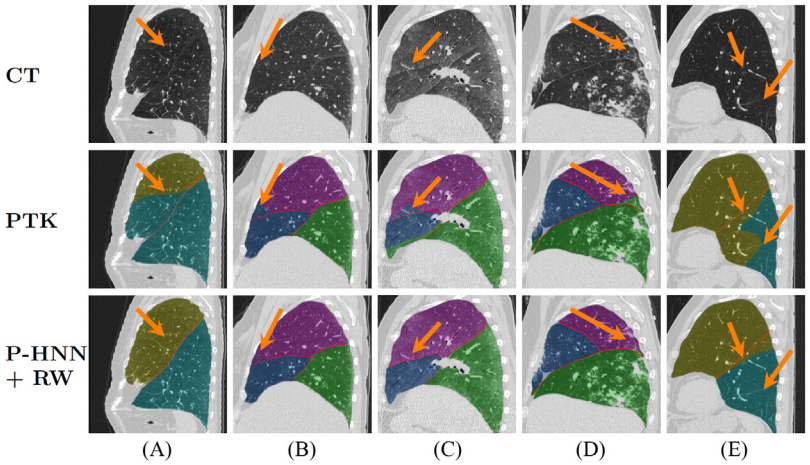


FIGURE 14.4 Lung lobe segmentation using George et al.'s[31] technique (P-HNN + RW). Here pulmonary toolkit (PTK) denotes Doel et al.'s[32] approach. Ground truth lobar boundaries are rendered in red. Despite its simplicity the P-HNN + RW technique can provide reliable lobe segmentations in challenging scenarios: (A) PTK follows an erroneous boundary, (B) P-HNN + RW handles an incomplete fissure, (C) PTK over segments one lobe, (D) P-HNN + RW does not get confounded by a fibrosis pattern that looks like a fissure, and (E) P-HNN + RW infers a reasonable lobar boundary even though there is no visual fissure. *Source: Credit: George K, Harrison AP, Jin D, Xu Z, Mollura DJ. Pathological pulmonary lobe segmentation from CT images using progressive holistically-nested neural networks and random walker. In: Cardoso MJ, Arbel T, Carneiro G, Syeda-Mahmood T, Tavares JMRS, Moradi M, et al. editors.* Deep learning in medical image analysis and multimodal learning for clinical decision support, lecture notes in computer science. *Cham: Springer International Publishing; 2017. p. 195–203. Available from: https://doi.org/10.1007/978-3-319-67558-9_23.*

automated methods have been developed to tackle this task including intensity-based[34]; morphology-based[35,36]; graph-based[37,38] and 2D learning-based.[39,40] Among these, different variations of region growing are often used. In contrast, 2D learning-based methods[39,40] can add potential robustness. However, their inability to consider the entire 3D volume greatly limits their learning capacities, since 3D information is crucial to detect small highly anisotropic tubular structures of airways. Another crucial limitation with learning-based approaches is that they rely on labeled training data to train their algorithms. However, the labor costs to fully annotate airways are much too high for large-scale datasets. To fill these gaps, Jin et al.[41] proposed the first 3D CNN-based method to fully leverage 3D airway tree features. A further graph-based refinement step addresses local discontinuities of the coarse 3D CNN output, which is then further refined by a curve skeletonization approach[42] to remove the blob-like segmentation leakages. It significantly improves over previous methods by extracting more than 30 airway branches per patient while maintaining similar false positive rates as compared to a prior art.[38] Importantly, their training process does not require perfect airway labels, as the 3D CNN is trained using the incomplete labels generated by Xu et al.[38] that have high specificity and moderate sensitivity. By learning from these incomplete labels, Jin et al.'s[41] approach can boost the sensitivity while maintaining high specificity. Fig. 14.5 provides some qualitative examples demonstrating the power of the 3D CNN for airway tree segmentation. After Jin et al.,[41] several subsequent works reported their own deep segmentation approaches that followed similar strategies.[43,44] Given the impossibility of obtaining large-scale and manually

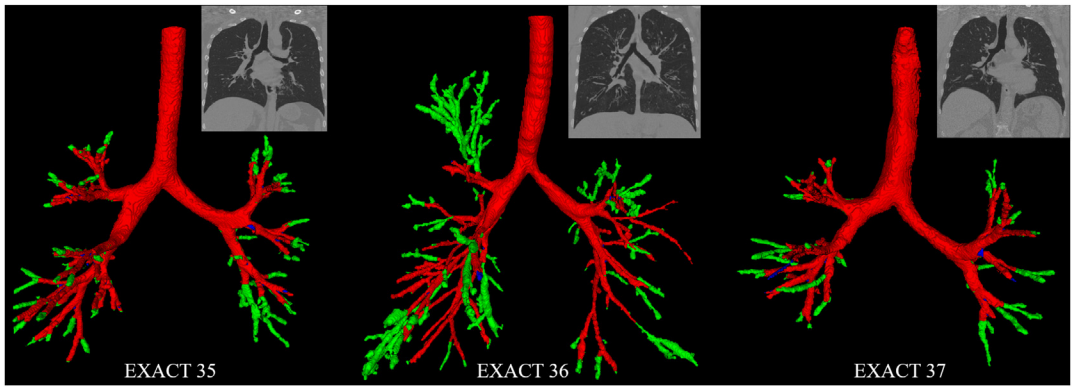


FIGURE 14.5 Examples of 3D rendering of airway segmentations using Jin et al.'s[41] 3D CNN technique as compared against a nonlearning-based prior work[38] on the EXACT09 dataset. Overlap regions are colored in red. Green and blue indicates additional extracted or missed branches, respectively, compared to the results from Xu et al.[38] *Source: Credit: Jin D, Xu Z, Harrison AP, George K, Mollura DJ. 3D convolutional neural networks with graph refinement for airway segmentation using incomplete data labels. In:* International workshop on machine learning in medical imaging. *Cham: Springer; 2017 September. p. 141–9.*

labeled airway datasets, continued work on approaches able to learn from weakly or incompletely labeled data will be vital to continue pushing progress.

14.2.2.2 Interstitial lung disease pattern recognition

ILD comprises more than 150 lung disorders affecting the lung parenchyma, which may eventually lead to breathing dysfunction. For the diagnosis of an ILD, besides the patient's clinical history and physical examination, CT scan is often ordered to provide a visual assessment of the lung tissues. This is a less risky procedure compared to biopsies. However, reading and interpreting large amounts of 3D chest CT scans requires significant time, effort, and experience from physicians. Yet, inter- and intra- observer agreement is frequently low because of the subjectivity and difficulty in interpreting ILD patterns.[45,46] Hence, many computerized and AI-based systems have been developed to automatically identify these abnormal patterns for increasing accuracy and consistency. Note that the new coronavirus 2019 (COVID-2019) causes severe pneumonia in certain patients. The corresponding CT scans include quite a few patterns that match patterns found in ILD, such as ground glass opacity, consolidation, reticulation, and crazy paving. Two COVID-2019 CT examples are shown in Fig. 14.6. AI-based lung pattern classification methods can be categorized into conventional image analysis and deep learning–based approaches, which are detailed in the following two paragraphs. We end this subsection by discussing the limitation of the current works and point out the possible directions for solving this important problem.

ILD comprises more than 150 lung disorders affecting the lung parenchyma, which may eventually lead to breathing dysfunction. For the diagnosis of an ILD, besides the patient's clinical history and physical examination, CT scan is often ordered to provide a visual assessment of the lung tissues. This is a less risky procedure compared to biopsies. However, reading and interpreting large amounts of 3D chest CT scans requires significant time, effort, and experience from physicians. Yet, inter- and intraobserver agreement

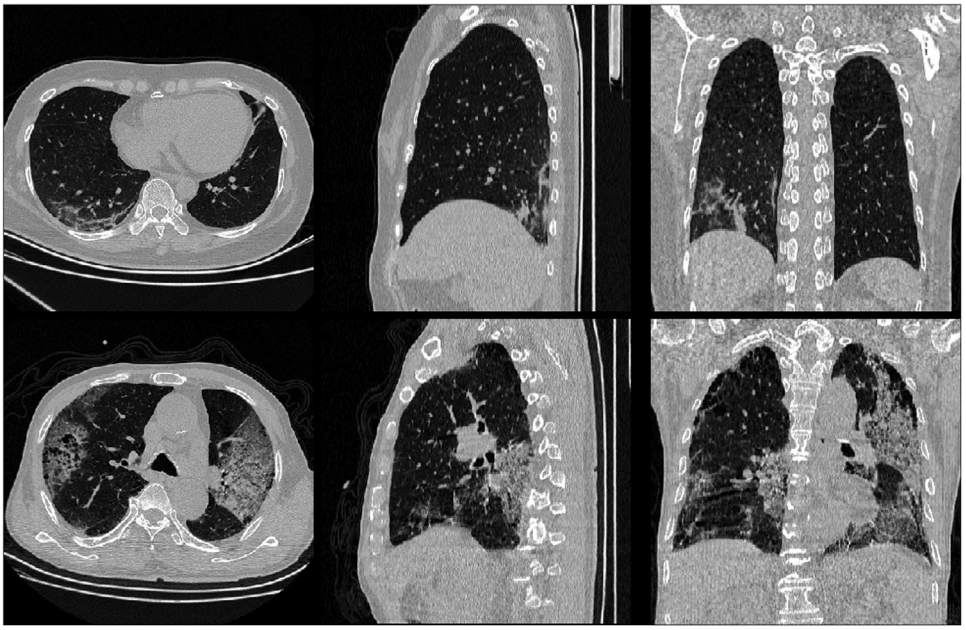


FIGURE 14.6 Examples of CT findings in two coronavirus 2019 (COVID-19) patients. The transverse, sagittal, and coronal views are shown for each case. The first row presents a patient with mild ground glass opacity in the right lower lobe. The bottom row shows a patient with severe radiologic progression with bilateral patchy shadowing. *CT*, Computerized tomography.

is frequently low because of the subjectivity and difficulty in interpreting ILD patterns.[45,46] Hence, many computerized and AI-based systems have been developed to automatically identify these abnormal patterns for increasing accuracy and consistency. Note that the new COVID-2019 causes severe pneumonia in certain patients. The corresponding CT scans include quite a few patterns that match patterns found in ILD, such as ground glass opacity, consolidation, reticulation, and crazy paving. Two COVID-2019 CT examples are shown in Fig. 14.6. AI-based lung pattern classification methods can be categorized into conventional image analysis and deep learning–based approaches, which are detailed in the following two paragraphs. We end this subsection by discussing the limitation of the current works and point out the possible directions for solving this important problem.

Early computerized lung pattern recognition works can trace back to the 1980s,[47,48] which used simple lung density analysis, such as mean or histogram percentile, to recognize emphysematous subjects. Later on, using local image patches, learning-based classification methods have been actively explored to identify various abnormal patterns, such as emphysema, honeycombing, ground glass opacity, consolidation, reticulation, nodular, or their combinations.[49–53] Various features have been designed for characterizing the distinct properties of the abnormal lung patterns, for example, basic statistical texture features, geometric features, features extracted by multiscale filter banks, and more complex features such as near-affine-invariant texture, rotation-invariant Gabor-local binary patterns, and the multicoordinate histogram of oriented gradients. Different classifiers

have been examined for their performance such as Bayesian classifier, linear discriminant classifier, and support vector machine with feature selections. These methods achieved quite divergent results due to different evaluation metrics distinct datasets.

Recently, deep learning–based AI solutions have shown promise. Anthimopoulos et al.[54] designed a customized CNN to conduct patch-based lung-pattern classification and gained markedly improved performance as compared to the non–deep-learning methods. This suggests that features automatically learned in a CNN network are more effective than previous handcrafted approaches. Gao et al.[55,56] further confirmed this, by introducing holistic slice-based classification for ILD diseases, where the CNN directly predicts if an axial slice contains any ILD disease patterns. This avoids needing to sample local image patches from manual Regions of Interest (ROIs) and can be used to prescreen a large amount of radiology data, which might be more clinically useful. Also of interest, Shin et al.[57] conducted a comprehensive evaluation of both patch- and holistic slice–based ILD pattern classification using different CNN structures and transfer learning.

Although deep-learning methods have shown promising results in recognizing abnormal patterns for ILD, current approaches face a bottleneck that there is no large-scale labeled data for training and evaluation. Note that there are two public datasets relevant to the ILD patterns: (1) the lung tissue research consortium (LTRC) contributed by the National Heart, Lung and Blood Institute[58,59] and (2) the specialized ILD dataset developed by University Hospitals of Geneva.[52] Although the LTRC includes more than 1000 (and counting) CT scans, from four centers, with COPDs and fibrotic ILD patterns, no manually annotated regions of interest are made available. In contrast, the ILD dataset contains manually annotated regions of 11 types of lung patterns. But these are only partially annotated. Moreover, only 108 CT scans with thick-slice spacing (10–15 mm) are made available, and they all originate from the same hospital and the partial labeling. Another limitation comes from the fact that these CT scans are from a single hospital and fail to cover sufficient variance of larger population with different scanners, which is crucial for enhancing the generalizability of the AI recognition systems. Thus limitations in labeled data is a major issue. There has been already published works addressing this issue, for example, Gao et al.[60] have explored deep-learning label propagation approaches to fully label the ILD dataset.[52] Nonetheless, further work is needed. Potentially, using techniques to mine unlabeled instances from multiple heterogeneous and incompletely labeled datasets, as explored in lesion detection,[61] might be a useful research direction.

14.3 Abdominal applications

AI systems have played a critical role for various cancer diagnostics in the abdomen, such as pancreatic ductal adenocarcinoma (PDAC), hepatocellular carcinoma, colorectal adenocarcinomas, etc. For instance, early computer-aided detection systems had been developed for polyps[62] and hepatic lesions.[63] In this section, we take pancreatic cancer as an example to show the importance of AI-based systems in cancer detection, segmentation, and the tumor growth prediction.

14.3.1 Pancreatic cancer analysis in computerized tomography and magnetic resonance imaging

Pancreatic cancer mainly includes two types: PDAC (85% of cases) and pancreatic neuroendocrine tumor (PanNET, less than 5% cases). PDAC is a major cause of cancer-related death in Western countries and is anticipated to emerge as the second leading cause of cancer-related death in the United States by 2030.[64] The prognosis of patients with PDAC is extremely poor, marked by a dismal 9% survival rate at 5 years. Medical imaging, for example, CT, is now routinely performed for depiction, quantification, staging, resectability evaluation, vascular invasion, and metastasis diagnosis of pancreatic cancers.

Automated analysis of pancreas images is a challenging task compared to other organs in CT, such as the heart, liver, and kidney, as the pancreas has a variable shape, size, and location in the abdomen. Pancreatic tumors are even more challenging to identify: they are quite variable in their shape, size, location, and have complex enhanced patterns, such as hypo-, iso-, or even hyperenhancement in different CT phases; moreover, the heterogeneity of pancreas regions (i.e., pancreas tissue, duct, veins, and arteries) and the ill-defined tumor boundary make pancreatic tumor segmentation highly difficult even for radiologists. Recent advances in machine learning and especially deep learning have led to substantial improvements in automated pancreas cancer analysis and have enabled the prediction and prognosis studies, such as tumor growth prediction and patient survival prediction. In this section, we cover the representative works of the pancreas and pancreatic tumor segmentation/detection, as well as the prediction and prognosis of pancreatic cancer.

14.3.1.1 Pancreas segmentation in computerized tomography and magnetic resonance imaging

Segmentation of the pancreas from 3D scans can provide quantitative features, such as the volume and shape statistics. Before deep learning, conventional methods report only 46.6%–69.1% dice score in the automatic pancreas segmentation. The performance has been significantly improved after adopting the deep-learning techniques.[65–68] Starting from 2D image patch-based CNN[68] to multiscale coarse-to-fine 3D fully convolutional network,[66] the Dice score is improved from 71.8% to 86.9% for healthy pancreas segmentation (example shown in Fig. 14.7), and computational time is reduced from 3 hours to 3 minutes. For abnormal pancreas segmentation, researchers recently achieve a comparably high Dice score of 86.7%[70] by the fusion of the arterial and venous enhanced CT phases in a hyperparing 3D UNet framework, achieving a similar level as the interobserver variability between radiologists.

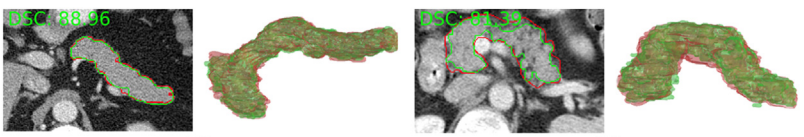
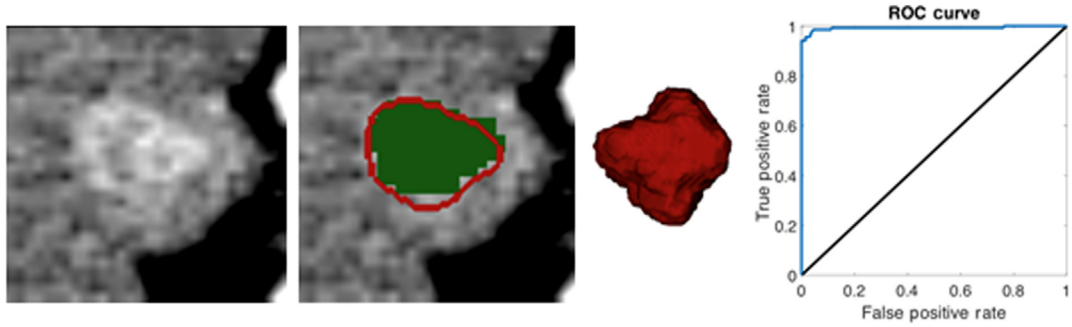


FIGURE 14.7 Example of pancreas segmentation results (green) (A) comparing with the ground-truth annotation (red) (B).[65] *Source: Credit: Roth HR, Lu L, Lay N, Harrison AP, Farag A, Sohn A. et al. Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation.* Med Image Anal *2018;**45**:94–107.*



(A) (B)

FIGURE 14.8 (A) Example of PanNET segmentation.[69] Red: algorithm segmentation; Green: ground truth. (B) ROC curve of pancreatic ductal adenocarcinoma screening.[66] *PanNET*, Pancreatic neuroendocrine tumor; *ROC*, Receiver operating characteristic. *Source: Credit: Zhu Z, Xia Y, Xie L, Fishman EK, Yuille AL. Multi-scale coarse-to-fine segmentation for screening pancreatic ductal adenocarcinoma. In:* International conference on medical image computing and computer-assisted intervention. *Cham: Springer; 2019, October. p. 3–12; Guo Z, Zhang L, Lu L, Bagheri M, Summers RM, Sonka M. et al. Deep LOGISMOS: deep learning graph-based 3D segmentation of pancreatic tumors on CT scans. In:* 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). *IEEE; 2018, April. p. 1230–3.*

14.3.1.2 Pancreatic tumor segmentation and detection in computerized tomography and magnetic resonance imaging

Precise tumor detection and segmentation are key elements in cancer imaging. For PDAC a multiscale coarse-to-fine 3D CNN method can automatically segment the tumors from venous phase CT with a Dice score of 57.3%.[66] With the identified suspicious regions of PDAC the pancreatic cancer screening/detection can be achieved. As such Zhu et al.[66] reports a sensitivity of 94.1% and a specificity of 98.5% for PDAC screening (Fig. 14.8B). To enhance the PDAC segmentation performance a hyperpairing framework[70] with the same network backbone as[66] is designed, which fuses venous and arterial phases at layer level. A much higher Dice score of 63.9% is reported. For PanNET a semiautomated method which combines UNet and 3D graph-based segmentation can segment tumor from arterial phase CT images with a Dice score of 83.2%[69] (Fig. 14.8A). This approach requires a manual click roughly at the tumor centroid for initialization. More generally, researchers attempt to segment the universal pancreatic tumors, that is, with a mix of PDAC and PanNET. Using the venous phase CT images, a cascade UNet approach produces a Dice score of 0.52 in a fully automated way.[67] Using dynamic contrast-enhanced MRI images, a patch-based semi-automated classification approach identifies tumor voxels in the pancreatic head region achieving a Dice score of 0.73, comparable to the interobserver variability.[71]

14.3.1.3 Prediction and prognosis with pancreatic cancer imaging

The prediction of patient-specific progression of pancreatic tumors at an earlier stage, such as PanNETs, will assist physicians in making decisions of the treatment plans. Such a prediction problem has long been tackled using principles of mathematical modeling. A few pieces of recent work[72,73] using deep-learning approaches can handle more complex

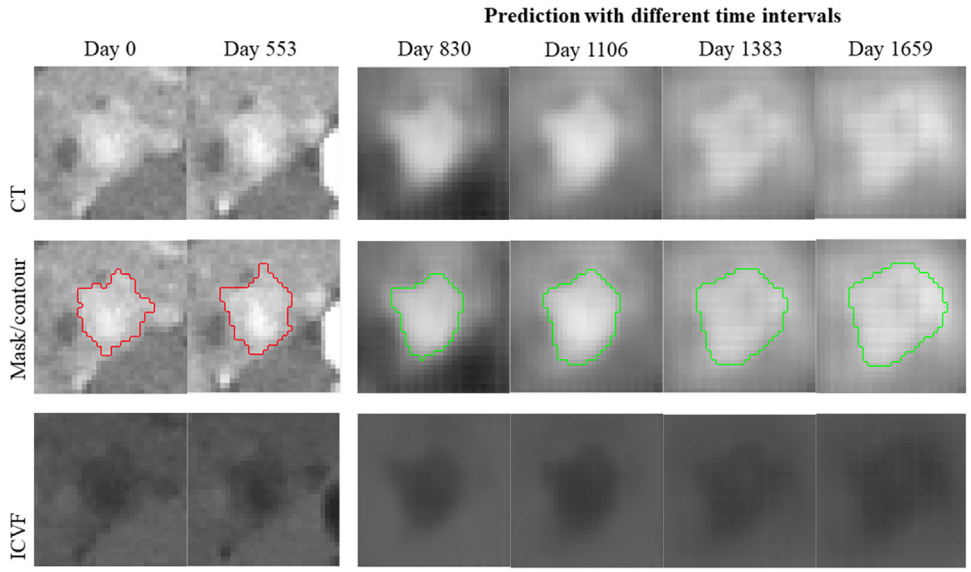


FIGURE 14.9 Example of deep-learning prediction of PanNET growth at different later time points.[72] *PanNET*, Pancreatic neuroendocrine tumor. *Source: Credit: Zhang L, Lu L, Wang X, Zhu RM, Bagheri M, Summers RM et al. Spatio-temporal convolutional LSTMs for tumor growth prediction by learning 4D longitudinal patient data. In:* IEEE transactions on medical imaging*; 2019.*

distributions from a larger patient population and provide more precise pixel-level prediction results. As demonstrated in Zhang et al.,[73] the two-stream CNN model achieves an average volume prediction error of 6.6% compared to a 13.9% error of a state-of-the-art mathematical modeling method using the same PanNET longitudinal dataset. The most recent work further enables the prediction of cell density and CT intensity,[72] and at the arbitrary future time point (shown in Fig. 14.9). There are also great interests in developing effective imaging-based biomarkers to stratify the patients with PDAC[74] and predict gene mutation status from CT imaging,[75] etc. Radiomics is still the mainstream approach in this direction. Making these biomarkers to reach the clinical practices, a highly automatic model and standardized radiomic features are desirable, as they can improve the objectiveness and enable the multicenter validation on large-scale patient cohorts.

14.3.2 AI in other abdominal imaging

Multiorgan segmentation in CT and MRI has attracted lots of research interest. Researchers have built several datasets with voxel-level annotations of the major abdominal organs and vessels.[65,76,77] The recent deep-learning approaches (either 2D or 3D based) have already achieved high accuracies for some larger organs, for example, Dice score of 98%, 97%, and 98% for liver, spleen, and kidney in CT images.[77,80] Small object segmentation is still challenging: Dice score of the duodenum is only 75% and that of the esophagus is only 76%. Segmentation of other abdominal tumors is also

important. Investigators build several public datasets with annotations of abdominal tumors (e.g., liver, kidney, and colon)[81], providing an opportunity for the whole community to develop the algorithms and helping to accelerate the development in this field.

14.4 Pelvic applications

While bone fracture detection is not the only AI application in the pelvic region, it is one of the most important and promising. Hip and pelvic fractures are among the most frequent fracture types worldwide.[82] Due to its low-cost, high-efficiency, and wide availability, pelvic X-ray imaging is the standard imaging tool for diagnosing pelvic and hip fractures. However, anatomical complexities and perspective projection distortions in the X-ray image contribute to a high rate of diagnostic errors[83] that may delay treatment and increase patient care cost, morbidity, and mortality.[84] As such, an effective AI system for both pelvic and hip fractures is of high clinical interest, with the aim of reducing diagnostic errors and improving patient outcomes. In this section, we will cover recent advances in AI-based fracture detection in pelvic X-rays.

The medical reports in PACSs and/or radiology information systems (RISs) provide natural sources of image labels for training a deep-learning-based AI system. These labels typically indicate positive finding of abnormalities (e.g., fracture) in the image, without specifying the exact location. The convenience of obtaining massive image-level labeled data from PACSs and/or RISs without manual annotation has driven the development of weakly supervised learning for the AI models in X-ray images, especially CXR applications.[4,22,85,86] In this formulation an image-level classification CNN is trained, and localizations of the detected abnormalities are provided via attention methods, for example, class activation mapping[15] or gradient-weighted class activation mapping.[87]

Hip fractures are the most common type of fracture visible in pelvic X-rays. Due to their high incidence, hip fractures are also the most well studied fracture type by the AI systems in pelvic X-rays. Cheng et al.[88] pretrained a popular CNN model on 25,505 limb radiographs and fine-tuned it on 3605 pelvic X-rays with hip fracture labels. The trained model reports an area under curve (AUC) of 0.980. Gale et al.[89] collected a training set of 45,492 pelvic X-rays with hip fractures labeled using a combination of orthopedics unit records and radiology reports. Training their AI model using manually extracted hip ROIs, they reported an impressive AUC of 0.994 on hip fracture identification, which matches radiologist-level performance. Their findings suggest that due to the localized nature of fractures and the complexity of the surrounding anatomical regions in the pelvis, concentrating on an ROI around the target anatomy (i.e., hip) is an effective strategy for detecting fractures. The effectiveness of employing ROI for hip fracture detection has also been demonstrated by Jiménez-Sánchez et al.,[90] who reported significant improvements in F1 scores using a ROI-base approach compared to a global approach. Jiménez-Sánchez et al.[90] further demonstrated that a curriculum learning scheme that starts from learning "easy" subtypes of hip fractures and gradually moves toward "hard" subtypes leads to a better performance with fewer training data.

Beside hip fractures, detecting the more complex pelvic fractures (fractures in three pelvic bones: the ilium, ischium, and pubis) is also of utmost importance, due to the potential critical

TABLE 14.1 Results of the Computer-aided detection system[78] and physicians performances on fracture detection in a reader study.

		Hip fracture		Pelvic fracture	
	Accuracy (%)	**Sensitivity (%)**	**Specificity (%)**	**Sensitivity (%)**	**Specificity (%)**
Emergency	88.1	98.3	93.7	81.3	95.5
Surgeon	85.5	93.1	92.8	82.9	93.2
Orthedpics	93.2	100	95.3	90.5	99.0
Radiology	93.0	99.0	96.5	87.0	99.5
Physician average	88.2	96.2	93.8	84.2	95.3
Wang et al.[78]	90.7	96.0	98.0	84.0	96.0

Credit: Wang Y, Lu L, Cheng CT, Jin D, Harrison AP, Xiao J, et al. Weakly supervised universal fracture detection in pelvic X-rays. In: International conference on medical image computing and computer-assisted intervention. *Cham: Springer; 2019d, October. p. 459–67.*

complications associated with pelvic fractures. The makeup of pelvic fractures is much more complex, as there are a large variety types with very different visual patterns at various locations. The overlap of pelvic bones with the lower abdomen anatomies further confounds image patterns. In addition, unlike hip fractures, which occur at the femoral neck/head, pelvic fractures can occur anywhere on the large pelvis, which precludes the use of anatomical ROIs to concentrate on local fracture patterns. To address the previously mentioned challenges in universal fracture detection in pelvic X-rays, Wang et al.[78] proposed a global-to-local two-stage gradient-weighted class activation mapping approach and reported radiologist-level performance. In the first stage a CNN is trained using a multiinstance learning formulation to generate proposals of potential fracture sites. ROIs of the generated proposals are collected and used to train the second stage local fracture identification network. During inference the two-stage models are chained together to provide a complete solution. This two-stage solution has the ability to concentrate on local fracture patterns despite the large field of view of pelvic X-rays. This method reports a high AUC of 0.975 on detecting both hip and pelvic fractures. A reader study involving 23 physicians from 4 departments (i.e., surgical, orthopedics, emergency, and radiology) on 150 pelvic X-rays demonstrates that the method outperforms emergency physicians and surgeons. Table 14.1 depicts the performances of physicians as well as the model on diagnosing hip and pelvic fractures. The model is also shown to be able to detect ambiguous fracture sites that are missed by physicians in the reader study. Fig. 14.10 shows a few examples of frequently missed fracture sites and their corresponding model detection results.

In summary the recent advances in the AI system for pelvic X-ray fracture detection has shown a trend of shifting from detecting a single fracture type toward universal fracture detection, which is often required to be deployed in real-world clinical scenarios such as emergency rooms or trauma centers. We also observe a paradigm shift from global classifier to local fracture pattern identification, represented by Gale et al.[89] and Wang et al.,[78] which significantly improves fracture-detection performance to reach radiologist-level.

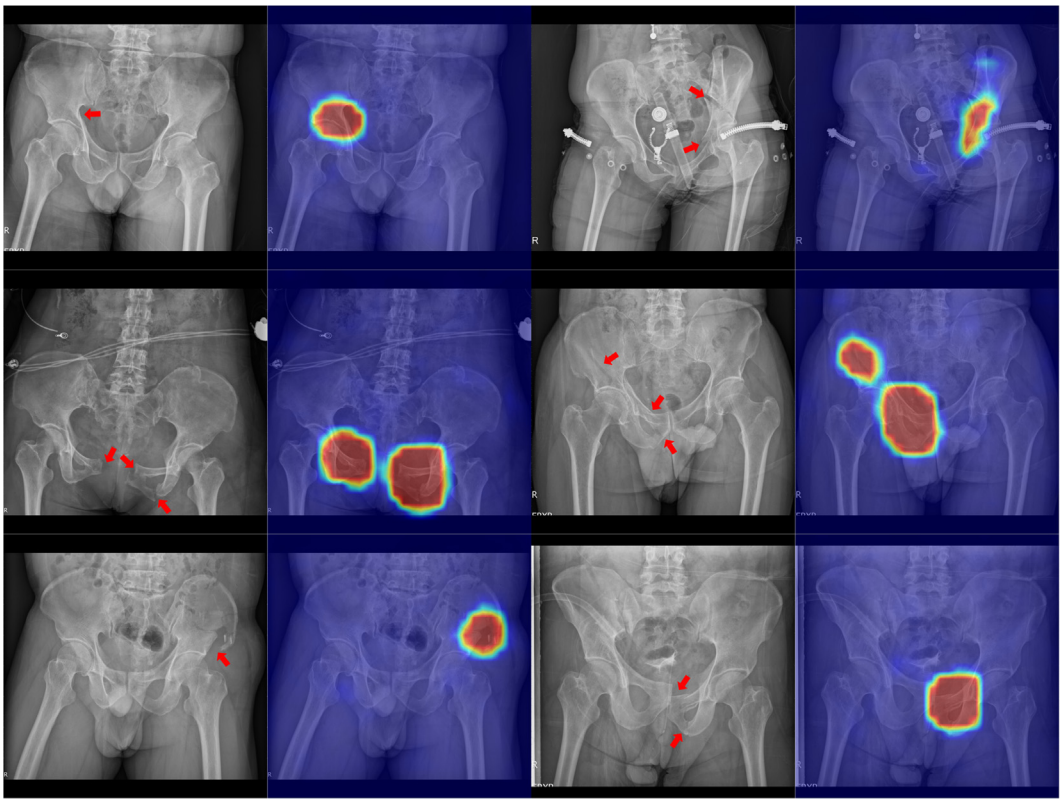


FIGURE 14.10 Examples of frequently missed fracture sites and their corresponding model detection results.[78]

14.5 Universal lesion analysis

When reading medical images, such as CT scans, radiologists generally search across the entire image to find lesions, characterize and measure them, and then describe them in the radiological report. This routine process is tedious and time-consuming. More importantly, human readers may miss some critical abnormal findings. This spurs research on automated lesion analysis algorithms (detection, classification, and segmentation) to decrease reading time and improve accuracy. However, most existing works focus on lesions of specific types and organs, such as lung nodules,[91] breast lesions,[92] and liver lesions.[93] Yet, in clinical scenarios, a CT scan may contain multiple types of lesions in different organs. For instance, metastasis can spread from a primary site to regional lymph nodes and other body parts or organs. Designing a model for each organ/lesion type is inefficient and less scalable. In addition, given the wide range of lesion types, a group of single-type models will still miss some infrequent types. To help radiologists find and characterize all of them, a universal lesion analysis (ULA) algorithm is ideal. While AI algorithms for specific lesions will always be valuable, ULA addresses an important part of radiologists' daily workflows and needs.

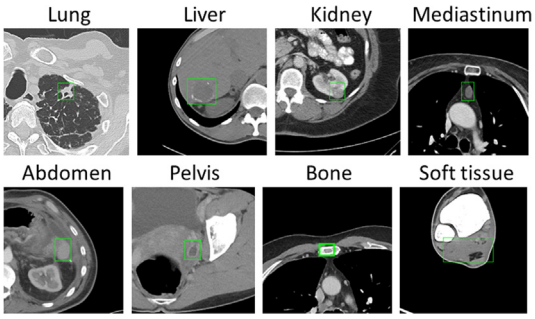


FIGURE 14.11 Exemplar lesions in the DeepLesion dataset.[94,95] *Source: Credit: Yan K, Wang X, Lu L, Summers RM. DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning.* J Med Imaging *2018a;5:1. Available from: https://doi.org/10.1117/1.JMI.5.3.036501 95; Yan K, Wang X, Lu L, Zhang L, Harrison A, Bagheri M, et al. Deep lesion graphs in the wild: relationship learning and organization of significant radiology image findings in a diverse large-scale lesion database. In:* CVPR*; 2018b.*

In this section, we first introduce the large-scale DeepLesion dataset[94] serving the purpose of ULA for the CT modality. Then we describe representative works for specific lesion analysis tasks, including lesion detection, classification, quantification, and retrieval, and mining.

14.5.1 DeepLesion dataset

To achieve ULA the first step is to collect a large-scale and diverse lesion dataset with comprehensive labels. Conventional data collection efforts would recruit experienced radiologists to manually annotate all lesions in 3D scans, which is extremely costly to acquire. Taking a different approach, the DeepLesion dataset[94,95] was collected from the PACS of the NIH Clinical Center by mining the response evaluation criteria in solid tumors (RECIST)[96] marks already annotated by radiologists during their daily work. DeepLesion contains 32,735 lesions annotated on 32,120 axial CT slices from 10,594 studies of 4427 patients. A visualization of lesions in the dataset can be found in Fig. 14.11. This dataset greatly boosted research on ULA.[11,79,80,95,97–105] It can also be readily updated or extended as it was mined automatically with minimal manual effort. Nonetheless, as with PACS-mined data in other domains, for example, CXR datasets, there are limitations. One important limitation is that the data are incompletely labeled, as radiologists do not typically mark all found lesions with RECIST marks. As outlined below, active research is currently underway to address this.

14.5.2 Lesion detection and classification

Universal lesion detection (ULD) is one of the most important tasks in ULA. It aims at finding a variety of lesions in the whole body and thus is more challenging than traditional single-type lesion detection because of the large appearance variation in different lesions types and the sometimes subtle distinction between lesions and nonlesions. CNN-based object detection frameworks such as the Faster Region based CNN[106] and Mask Region based CNN[107] are often adopted for ULD. Its performance has been improved by through various enhancements in the analysis. For instance, 3D context information in neighboring slices is important for detection, as lesions may be less distinguishable in just one 2D axial slice. Yan et al.[98,102] and Wang et al.[79] exploited 3D information with multi-slice image inputs and a 2.5D network by fusing features of multiple slices. On the other hand, Zlocha et al.,[103] Wang et al.,[79,80] and Li et al.[104] used attention mechanisms to

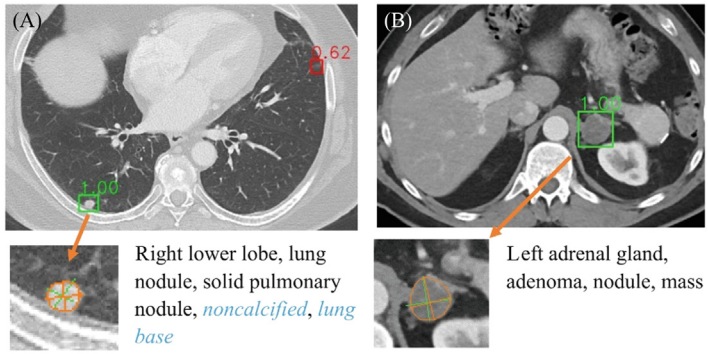


FIGURE 14.12 Examples of the lesion detection, tagging, and segmentation results of MULAN.[102] For detection, boxes in green and red are predicted TPs and FPs, respectively. The number above each box is the confidence score. For tagging, tags in black and blue are predicted TPs and FNs, respectively. For segmentation the green lines are ground-truth RECIST measurements; the orange contours and lines show predicted masks and RECIST measurements, respectively. *FN*, False negative; *FP*, false positive; *MULAN*, multitask universal lesion analysis network; *RECIST*, response evaluation criteria in solid tumors; *TP*, true positive. *Source: Reproduced from and Credit Yan K, Tang Y, Peng Y, Sandfort V, Bagheri M, Lu Z, et al. MULAN: multitask universal lesion analysis network for joint lesion detection, tagging, and segmentation. In:* MICCAI*; 2019b. p. 194–202. Available from: https://doi.org/10.1007/978-3-030-32226-7_22.*

emphasize important regions and features within the deep CNN. Wang et al.[80] went even further and proposed a domain attention module to learn from DeepLesion and 10 other object detection datasets simultaneously. ULDor[99] used a trained detector to mine hard negative proposals and then retrained the model. Finally, the multitask ULA network (MULAN)[102] jointly learned lesion detection, segmentation, and tagging, and used a score refinement layer to improve detection with tagging. It achieved the current state-of-the-art accuracy on DeepLesion, that is, 83.7% recall at one false positive per key slice. Fig. 14.12 illustrates exemplar results of MULAN.

Automatic lesion classification can assist diagnostic decision-making and structured report generation. Existing algorithms usually focus on certain body parts and attempt to distinguish between a limited set of labels.[91–93] In contrast, Yan et al. and Peng et al.[100,101] learned from the DeepLesion dataset to predict 171 comprehensive labels for a variety of lesions to describe their body part, type, and attributes. They first designed a natural language processing algorithm to extract relevant semantic labels from the radiology reports associated with the lesion images and then proposed a lesion annotation network (LesaNet) for multilabel classification, leveraging hierarchical, and mutually exclusive relations between the labels to improve the label prediction accuracy. LesaNet's average classification AUC of the 171 labels is 0.934.

14.5.3 Lesion segmentation and quantification

Lesion segmentation and measurement results are useful for clinicians to evaluate lesion sizes and treatment responses. In DeepLesion, lesions were annotated with two RECIST diameters including one long axis and the orthogonal short axis.[94,96] However,

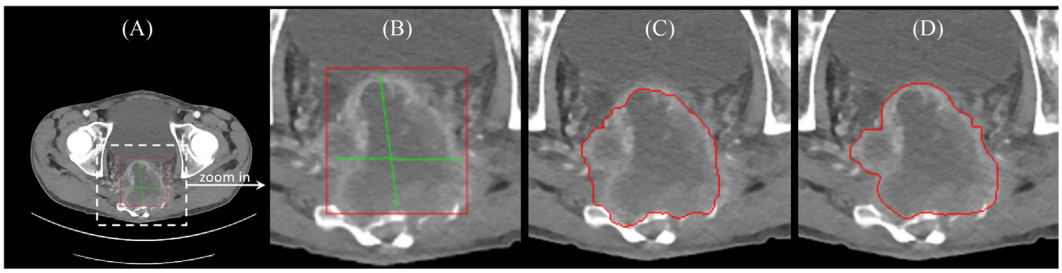


FIGURE 14.13 Example of automatic lesion segmentation with weakly supervised slice-propagated segmentation method.[97] We show an axial CT slice that contains a lesion measured by a RECIST mark in (A). The highlighted lesion and the RECIST mark is shown in (B) using green color. The red box is the region of interest that is conducted from the RECIST mark and used for initializing automatic segmentation. (C) and (D) show the result of automatic segmentation and manually delineated ground-truth segmentation, respectively. *CT*, Computerized tomography; *RECIST*, response evaluation criteria in solid tumors. *Source: Credit: Cai J, Tang Y, Lu L, Harrison AP, Yan K, Xiao J, et al. Accurate weakly-supervised deep lesion segmentation using large-scale clinical annotations: slice-propagated 3D mask generation from 2D RECIST. In:* MICCAI*; 2018b.*

RECIST marks are subjective and can be prone to inconsistency among different observers, especially when selecting the corresponding axial slices at different time-points where RECIST diameters are measured. To alleviate this problem, Tang et al.[11] designed a cascaded CNN to automatically predict the endpoints of the RECIST diameters, yielding reliable and reproducible lesion measurement results with an average error of $\sim$3 pixels.

Compared with RECIST diameters, volumetric lesion measurement can be a better metric for holistic and accurate quantitative assessment of lesion growth rates, avoiding the subjective selection of axial slice for RECIST measurement. Unfortunately, obtaining full volumetric lesion measurements with manual segmentations is labor-intensive and time-consuming. For this reason, RECIST is treated as the default, but imperfect, clinical surrogate of measuring lesion progression. To facilitate automatic segmentation of lesion volumes, Cai et al.[97] presented a weakly supervised slice-propagated segmentation method with DeepLesion to learn from the RECIST annotations and predict 3D lesion masks. They reported a patient-wise mean Dice score of 91.5% for lesion segmentation measured on the key slices (the axial slice containing the RECIST mark). Fig. 14.13 shows an example of automatic lesion segmentation on the RECIST-marked CT slice. With slice-wise propagation, Cai et al.'s[97] method can produce volumetric segmentations, achieving 76.4% Dice scores across the entire lesion.

14.5.4 Lesion retrieval and mining

The goal of lesion retrieval is to find similar lesions from a database to help the user understand the query lesion. DeepLesion also provides a valuable platform to explore the similarity relationship among a variety of lesions. For instance, Yan et al.[95] trained a triplet network to learn quantitative lesion embeddings that reflected lesion "similarity." Similarity was defined hierarchically based on the lesion type, anatomical location, and size. The embeddings can also be used to build a lesion graph for intra-patient lesion

matching.[95] The lesion labels mined from radiological reports can also be adopted to learn embeddings to encode more fine-grained semantic information.[100]

In terms of lesion mining, one limitation of DeepLesion is that not all lesions in the dataset were annotated. Cai et al.[105] exploited a small fully labeled subset of volumes and used it to intelligently mine annotations from the remainder of images in DeepLesion. They showed that lesion detectors trained on the harvested lesions and hard negatives can significantly outperform the same variants only trained on the original annotations, boosting average precision by 7%−10%.

Despite the progress of ULA in recent years, there is still room for improvement, for example, the detection accuracy for lesions in confusing or rare body parts[98] is still insufficient for practical use. One interesting research direction is to combine existing single-type lesion datasets with DeepLesion and leverage their synergy to further improve detection accuracy.

14.6 Conclusion

Significant advances in AI technology may greatly impact and eventually alter radiology workflows. In this chapter, several important medical imaging tasks in different anatomies are reviewed. Specifically, we overview AI applications in thoracic, abdominal, and pelvic regions as well as general lesion analysis. Different tasks, such as detection, segmentation, and classification, are discussed to highlight their strengths and limitations. These should provide radiologists with a better understanding of current AI technology and its potential going forward in improving efficiency, accuracy, and consistency of various radiology procedures.

References

1. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* 2018;**24**(9):1342−50.
2. Mettler FA, Bhargavan M, Faulkner K, Gilley DB, Gray JE, Ibbott GS, et al. Radiologic and nuclear medicine studies in the United States and worldwide: frequency, radiation dose, and comparison with other radiation sources—1950−2007. *Radiology* 2009;**253**:520−31. Available from: https://doi.org/10.1148/radiol.2532082010.
3. Raoof S, Feigin D, Sung A, Raoof S, Irugulpati L, Rosenow EC. Interpretation of plain chest roentgenogram. *Chest* 2012;**141**:545−58. Available from: https://doi.org/10.1378/chest.10-1302.
4. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-Ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: *2017 IEEE conference on computer vision and pattern recognition (CVPR). Presented at the 2017 IEEE conference on computer vision and pattern recognition (CVPR)*; 2017. pp. 3462−71. Available from: https://doi.org/10.1109/CVPR.2017.369.
5. Gohagan JK, Prorok PC, Hayes RB, Kramer B-S. The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: history, organization, and status. *Controlled Clin Trials* 2000;**21**:251S−72S. Available from: https://doi.org/10.1016/S0197-2456(00)00097-0.
6. Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, et al. CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison. *Proc AAAI Conf Artif Intell* 2019;**33**:590−7. Available from: https://doi.org/10.1609/aaai.v33i01.3301590.
7. Johnson AEW, Pollard TJ, Berkowitz SJ, Greenbaum NR, Lungren MP, Deng C, et al. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Sci Data* 2019;**6**:1−8. Available from: https://doi.org/10.1038/s41597-019-0322-0.

8. Bustos A, Pertusa A, Salinas J-M, de la Iglesia-Vayá M. *PadChest: a large chest x-ray image dataset with multi-label annotated reports*. arXiv:1901.07441 [cs, eess], 2019.
9. Oakden-Rayner L. Exploring large-scale public medical image datasets. *Acad Radiol* 2020;**27**:106–12. Available from: https://doi.org/10.1016/j.acra.2019.10.006.
10. Rajpurkar P, Joshi A, Pareek A, Chen P, Kiani A, Irvin J, et al. *CheXpedition: investigating generalization challenges for translation of chest x-ray algorithms to the clinical setting*. arXiv:2002.11379 [cs, eess]. 2020.
11. Tang Y, Harrison AP, Bagheri M, Xiao J, Summers RM. Semi-automatic RECIST labeling on CT scans with cascaded convolutional neural networks. In: *MICCAI*; 2018a. pp. 405–13. Available from: https://doi.org/10.1007/978-3-030-00937-3_47.
12. Tang Y, Wang X, Harrison AP, Lu L, Xiao J, Summers RM. Attention-guided curriculum learning for weakly supervised classification and localization of thoracic diseases on chest radiographs. In: Shi Y, Suk H-I, Liu M, editors. *Machine learning in medical imaging, lecture notes in computer science*. Cham: Springer International Publishing; 2018. pp. 249–58. Available from: https://doi.org/10.1007/978-3-030-00919-9_29.
13. Pesce E, Ypsilantis P-P, Withey S, Bakewell R, Goh V, Montana G. *Learning to detect chest radiographs containing lung nodules using visual attention networks*. arXiv:1712.00996 [cs, stat], 2017.
14. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med* 2018;**15**: e1002686. Available from: https://doi.org/10.1371/journal.pmed.1002686.
15. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. In: *2016 IEEE conference on computer vision and pattern recognition (CVPR). Presented at the 2016 IEEE conference on computer vision and pattern recognition (CVPR)*; 2016. pp. 2921–9. Available from: https://doi.org/10.1109/CVPR.2016.319.
16. Tang Y-X, Tang Y-B, Han M, Xiao J, Summers RM Abnormal chest x-ray identification with generative adversarial one-class classifier. In: *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019). Presented at the 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*; 2019c. pp. 1358–61. Available from: https://doi.org/10.1109/ISBI.2019.8759442.
17. Li Z, Wang C, Han M, Xue Y, Wei W, Li L-J, et al. Thoracic disease identification and localization with limited supervision. In: *2018 IEEE/CVF conference on computer vision and pattern recognition. Presented at the 2018 IEEE/CVF conference on computer vision and pattern recognition (CVPR), IEEE, Salt Lake City, UT*; 2018. pp. 8290–9. Available from: https://doi.org/10.1109/CVPR.2018.00865.
18. Cai J, Lu L, Harrison AP, Shi X, Chen P, Yang L. Iterative attention mining for weakly supervised thoracic disease pattern localization in chest X-rays. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *Medical image computing and computer assisted intervention – MICCAI 2018, lecture notes in computer science*. Cham: Springer International Publishing; 2018. pp. 589–98. Available from: https://doi.org/10.1007/978-3-030-00934-2_66.
19. Zhang Y, Miao S, Mansi T, Liao R. Task driven generative modeling for unsupervised domain adaptation: application to X-ray image segmentation. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *Medical image computing and computer assisted intervention – MICCAI 2018, lecture notes in computer science*. Cham: Springer International Publishing; 2018. pp. 599–607. Available from: https://doi.org/10.1007/978-3-030-00934-2_67.
20. Tang Y-B, Tang Y-X, Xiao J, Summers RM. XLSor: a robust and accurate lung segmentor on chest X-rays using criss-cross attention and customized radiorealistic abnormalities generation. In: *International conference on medical imaging with deep learning. Presented at the international conference on medical imaging with deep learning*; 2019a. pp. 457–67.
21. Tang Y, Tang Y, Sandfort V, Xiao J, Summers RM. TUNA-net: task-oriented UN supervised adversarial network for disease recognition in cross-domain chest X-rays. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, Yap P-T, Khan A, editors. *Medical image computing and computer assisted intervention – MICCAI 2019, lecture notes in computer science*. Cham: Springer International Publishing; 2019. pp. 431–40. Available from: https://doi.org/10.1007/978-3-030-32226-7_48.
22. Chen H, Miao S, Xu D, Hager GD, Harrison AP. Deep hierarchical multi-label classification of chest X-ray images. Un: *Proceedings of machine learning research. Presented at the MIDL 2019*; 2019. 12.
23. Armato III SG, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, 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**(2):915–31.

24. Grenier P, Valeyre D, Cluzel P, Brauner MW, Lenoir S, Chastang C. Chronic diffuse interstitial lung disease: diagnostic value of chest radiography and high-resolution CT. *Radiology* 1991;**179**(1):123–32.
25. Lee Y, Hara T, Fujita H, Itoh S, Ishigaki T. Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique. *IEEE Trans Med Imaging* 2001;**20**(7):595–604.
26. McNitt-Gray MF, Hart EM, Wyckoff N, Sayre JW, Goldin JG, Aberle DR. A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: preliminary results. *Med Phys* 1999;**26**(6):880–8.
27. Blechschmidt RA, Werthschutzky R, Lorcher U. Automated CT image evaluation of the lung: a morphology-based concept. *IEEE Trans Med Imaging* 2001;**20**(5):434–42.
28. Jin D, Xu Z, Tang Y, Harrison AP, Mollura DJ. *CT-realistic lung nodule simulation from 3D conditional generative adversarial networks for robust lung segmentation. Medical image computing and computer assisted intervention – MICCAI 2018, Lecture notes in computer science*. Cham: Springer International Publishing; 2018. pp. 732–40. Available from: https://doi.org/10.1007/978-3-030-00934-2_81.
29. Harrison AP, Xu Z, George K, Lu L, Summers RM, Mollura DJ. Progressive and multi-path holistically nested neural networks for pathological lung segmentation from CT images. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S, editors. *Medical image computing and computer assisted intervention – MICCAI 2017, Lecture notes in computer science*. Springer International Publishing; 2017. pp. 621–9.
30. Mansoor A, Bagci U, Xu Z, Foster B, Olivier KN, Elinoff JM, et al. A generic approach to pathological lung segmentation. *IEEE Trans Med Imaging* 2014;**33**:2293–310. Available from: https://doi.org/10.1109/TMI.2014.2337057.
31. George K, Harrison AP, Jin D, Xu Z, Mollura DJ. Pathological pulmonary lobe segmentation from CT images using progressive holistically nested neural networks and random walker. In: Cardoso MJ, Arbel T, Carneiro G, Syeda-Mahmood T, Tavares JMRS, Moradi M, et al., editors. *Deep learning in medical image analysis and multimodal learning for clinical decision support, lecture notes in computer Science*. Cham: Springer International Publishing; 2017. pp. 195–203. Available from: https://doi.org/10.1007/978-3-319-67558-9_23.
32. Doel T, Matin TN, Gleeson FV, Gavaghan DJ, Grau V. Pulmonary lobe segmentation from CT images using fissureness, airways, vessels and multilevel B-splines. In: *2012 9th IEEE international symposium on biomedical imaging (ISBI). Presented at the 2012 9th IEEE international symposium on biomedical imaging (ISBI)*; 2012. pp. 1491–4. Available from: https://doi.org/10.1109/ISBI.2012.6235854.
33. Bragman FJS, McClelland JR, Jacob J, Hurst JR, Hawkes DJ. Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior. *IEEE Trans Med Imaging* 2017;**36**:1650–63. Available from: https://doi.org/10.1109/TMI.2017.2688377.
34. Van Rikxoort EM, Baggerman W, Van Ginneken B. Automatic segmentation of the airway tree from thoracic CT scans using a multi-threshold approach. In: *Proc of second international workshop on pulmonary image analysis*; 2009. pp. 341–9.
35. Aykac D, Hoffman EA, McLennan G, Reinhardt JM. Segmentation and analysis of the human airway tree from three-dimensional X-ray CT images. *IEEE Trans Med Imaging* 2003;**22**(8):940–50.
36. Nadeem SA, Jin D, Hoffman EA, Saha PK. *An iterative method for airway segmentation using multiscale leakage detection. Med imaging 2017: image process*, 10133. International Society for Optics and Photonics; 2017. p. 1013308.
37. Tschirren J, Hoffman EA, McLennan G, Sonka M. Intrathoracic airway trees: segmentation and airway morphology analysis from low-dose CT scans. *IEEE Trans Med Imaging* 2005;**24**(12):1529–39.
38. Xu Z, Bagci U, Foster B, Mansoor A, Udupa JK, Mollura DJ. A hybrid method for airway segmentation and automated measurement of bronchial wall thickness on CT. *Med Image Anal* 2015;**24**(1):1–17.
39. Charbonnier JP, Van Rikxoort EM, Setio AA, Schaefer-Prokop CM, van Ginneken B, Ciompi F. Improving airway segmentation in computed tomography using leak detection with convolutional networks. *Med Image Anal* 2017;**36**:52–60.
40. Lo P, Sporring J, Ashraf H, Pedersen JJ, de Bruijne M. Vessel-guided airway tree segmentation: a voxel classification approach. *Med Image Anal* 2010;**14**(4):527–38.
41. Jin D, Xu Z, Harrison AP, George K, Mollura DJ. *3D convolutional neural networks with graph refinement for airway segmentation using incomplete data labels. International workshop on machine learning in medical imaging*. Cham: Springer; 2017. pp. 141–9.
42. Jin D, Iyer KS, Chen C, Hoffman EA, Saha PK. A robust and efficient curve skeletonization algorithm for tree-like objects using minimum cost paths. *Pattern Recognit Lett* 2016;**76**:32–40.

43. Yun J, Park J, Yu D, Yi J, Lee M, Park HJ, et al. Improvement of fully automated airway segmentation on volumetric computed tomographic images using a 2.5 dimensional convolutional neural net. *Med Image Anal* 2019;**51**:13–20.
44. Qin Y, Chen M, Zheng H, Gu Y, Shen M, Yang J, et al. *AirwayNet: a voxel-connectivity aware approach for accurate airway segmentation using convolutional neural networks. International conference on medical image computing and computer-assisted intervention*. Cham: Springer; 2019. pp. 212–20.
45. Nishimura K, Izumi T, Kitaichi M, Nagai S, Itoh H. The diagnostic accuracy of high-resolution computed tomography in diffuse infiltrative lung diseases. *Chest* 1993;**104**(4):1149–55.
46. Padley SPG, Hansell DM, Flower CDR, Jennings P. Comparative accuracy of high resolution computed tomography and chest radiography in the diagnosis of chronic diffuse infiltrative lung disease. *Clin Radiol* 1991;**44**(4):222–6.
47. Goddard PA, Nicholson EM, Laszlo G, Watt I. Computed tomography in pulmonary emphysema. *Clin Radiol* 1982;**33**:379–87.
48. Hayhurst MD, MacNee W, Flenley DC, Wright D, McLean A, Lamb D, et al. Diagnosis of pulmonary emphysema by computed tomography. *Lancet* 1984;**2**:320–2.
49. Uppaluri R, Hoffman EA, Sonka M, Hartley PG, Hunninghake GW, McLennan G. Computer recognition of regional lung disease patterns. *Am J Respiratory Crit Care Med* 1999;**160**(2):648–54.
50. Sluimer IC, van Waes PF, Viergever MA, van Ginneken B. Computer-aided diagnosis in high resolution CT of the lungs. *Med Phys* 2003;**30**(12):3081–90.
51. Depeursinge A, Van de Ville, Platon D, Geissbuhler A, Poletti PA, Muller H. Near-affine-invariant texture learning for lung tissue analysis using isotropic wavelet frames. *IEEE Trans Inf Technol Biomed* 2012;**16**(4):665–75.
52. Depeursinge A, Vargas A, Platon A, Geissbuhler A, Poletti PA, Müller H. Building a reference multimedia database for interstitial lung diseases. *Computerized Med Imaging Graph* 2012;**36**(3):227–38.
53. Song Y, Cai W, Zhou Y, Feng DD. Feature-based image patch approximation for lung tissue classification. *IEEE Trans Med Imaging* 2013;**32**(4):797–808.
54. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans Med imaging* 2016;**35**(5):1207–16.
55. Gao M, Bagci U, Lu L, Wu A, Buty M, Shin HC, et al. Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks. *Comput Methods Biomech Biomed Eng: Imaging Vis* 2016;**6**(1):1–6.
56. Gao M, Xu Z, Lu L, Harrison AP, Summers RM, Mollura DJ. *Multi-label deep regression and unordered pooling for holistic interstitial lung disease pattern detection. International workshop on machine learning in medical imaging*. Cham: Springer; 2016. pp. 147–55.
57. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 2016;**35**(5):1285–98.
58. Karwoski RA, Bartholmai B, Zavaletta VA, Holmes D, Robb RA. *Processing of CT images for analysis of diffuse lung disease in the lung tissue research consortium. Medical imaging 2008: physiology, function, and structure from medical images*, 6916. International Society for Optics and Photonics; 2008. pp. 691614.
59. Bartholmai B, Karwoski R, Zavaletta V, Robb R, Holmes DRI. The Lung Tissue Research Consortium: an extensive open database containing histological, clinical, and radiological data to study chronic lung disease. *Insight J* 2006.
60. Gao M, Xu Z, Lu L, Wu A, Nogues I, Summers RM, et al. *Segmentation label propagation using deep convolutional neural networks and dense conditional random field. 2016 IEEE 13th international symposium on biomedical imaging (ISBI)*. IEEE; 2016. pp. 1265–8.
61. Yan K, Cai J, Harisson AP, Jin D, Xiao J, Lu L. *Universal lesion detection by learning from multiple heterogeneously labeled datasets. Under review*. 2020.
62. Summers RM, Jerebko AK, Franaszek M, Malley JD, Johnson CD. Colonic polyps: complementary role of computer-aided detection in CT colonography. *Radiology* 2002;**225**(2):391–9.
63. Bilello M, Gokturk SB, Desser T, Napel S, Jeffrey Jr RB, Beaulieu CF. Automatic detection and classification of hypodense hepatic lesions on contrast-enhanced venous-phase CT. *Med Phys* 2004;**31**(9):2584–93.
64. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res* 2014;**74**(11):2913–21.

65. Roth HR, Lu L, Lay N, Harrison AP, Farag A, Sohn A, et al. Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. *Med Image Anal* 2018;**45**:94–107.
66. Zhu Z, Xia Y, Xie L, Fishman EK, Yuille AL. *Multi-scale coarse-to-fine segmentation for screening pancreatic ductal adenocarcinoma. International conference on medical image computing and computer-assisted intervention*. Cham: Springer; 2019. pp. 3–12.
67. Isensee F, Petersen J, Klein A, Zimmerer D, Jaeger PF, Kohl S, et al. *nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:1809.10486*. 2018.
68. Roth HR, Lu L, Farag A, Shin HC, Liu J, Turkbey EB, et al. *Deeporgan: multi-level deep convolutional networks for automated pancreas segmentation. International conference on medical image computing and computer-assisted intervention*. Cham: Springer; 2015. pp. 556–64.
69. Guo Z, Zhang L, Lu L, Bagheri M, Summers RM, Sonka M, et al. *Deep LOGISMOS: deep learning graph-based 3D segmentation of pancreatic tumors on CT scans. 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. IEEE; 2018. pp. 1230–3.
70. Zhou Y, Li Y, Zhang Z, Wang Y, Wang A, Fishman EK, et al. *Hyper-pairing network for multi-phase pancreatic ductal adenocarcinoma segmentation. International conference on medical image computing and computer-assisted intervention*. Cham: Springer; 2019. pp. 155–63.
71. Liang Y, Schott D, Zhang Y, Wang Z, Nasief H, Paulson E, et al. Auto-segmentation of pancreatic tumor in multi-parametric MRI using deep convolutional neural networks. *Radiother Oncol* 2020;**145**:193–200.
72. Zhang L, Lu L, Wang X, Zhu RM, Bagheri M, Summers RM et al. Spatio-temporal convolutional LSTMs for tumor growth prediction by learning 4D longitudinal patient data. In: *IEEE transactions on medical imaging*; 2019.
73. Zhang L, Lu L, Summers RM, Kebebew E, Yao J. Convolutional invasion and expansion networks for tumor growth prediction. *IEEE Trans Med Imaging* 2018;**37**(2):638–48.
74. Attiyeh MA, Chakraborty J, Doussot A, Langdon-Embry L, Mainarich S, Gönen M, et al. Survival prediction in pancreatic ductal adenocarcinoma by quantitative computed tomography image analysis. *Ann Surg Oncol* 2018;**25**(4):1034–42.
75. Attiyeh MA, Chakraborty J, McIntyre CA, Kappagantula R, Chou Y, Askan G, et al. CT radiomics associations with genotype and stromal content in pancreatic ductal adenocarcinoma. *Abdom Radiol* 2019;**44**(9):3148–57.
76. Gibson E, Giganti F, Hu Y, Bonmati E, Bandula S, Gurusamy K, et al. Automatic multi-organ segmentation on abdominal CT with dense v-networks. *IEEE Trans Med Imaging* 2018;**37**(8):1822–34.
77. Wang Y, Zhou Y, Shen W, Park S, Fishman EK, Yuille AL. Abdominal multi-organ segmentation with organ-attention networks and statistical fusion. *Med Image Anal* 2019;**55**:88–102.
78. Wang Y, Lu L, Cheng CT, Jin D, Harrison AP, Xiao J, et al. *Weakly supervised universal fracture detection in pelvic X-rays. International conference on medical image computing and computer-assisted intervention*. Cham: Springer; 2019. p. 459–67.
79. Wang X, Han S, Chen Y, Gao D, Vasconcelos N. Volumetric attention for 3D medical image segmentation and detection. In: *MICCAI*; 2019a. pp. 175–84. Available from: https://doi.org/10.1007/978-3-030-32226-7_20.
80. Wang X, Cai Z, Gao D, Vasconcelos N. Towards universal object detection by domain attention. In: *CVPR*; 2019b. pp. 7281–90. Available from: https://doi.org/10.1109/CVPR.2019.00746.
81. Simpson AL, Antonelli M, Bakas S, Bilello M, Farahani K, Van Ginneken B, et al. *A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063*; 2019.
82. Johnell O, Kanis JA. An estimate of the worldwide prevalence, mortality and disability associated with hip fracture. *Osteoporos Int* 2004;**15**(11):897–902.
83. Chellam WB. Missed subtle fractures on the trauma-meeting digital projector. *Injury* 2016;**47**(3):674–6.
84. Tarrant SM, Hardy BM, Byth PL, Brown TL, Attia J, Balogh ZJ. Preventable mortality in geriatric hip fracture inpatients. *Bone Jt. J* 2014;**96**(9):1178–84.
85. Badgeley MA, Zech JR, Oakden-Rayner L, Glicksberg BS, Liu M, Gale W, et al. Deep learning predicts hip fracture using confounding patient and healthcare variables. *NPJ Digital Med* 2019;**2**(1):1–10.
86. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. *Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225*; 2017.
87. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE international conference on computer vision*; 2017. pp. 618–26.
88. Cheng CT, Ho TY, Lee TY, Chang CC, Chou CC, Chen CC, et al. Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs. *Eur Radiol* 2019;**29**(10):5469–77.

89. Gale W, Oakden-Rayner L, Carneiro G, Bradley AP, Palmer LJ. *Detecting hip fractures with radiologist-level performance using deep neural networks. arXiv preprint arXiv:1711.06504*; 2017.
90. Jiménez-Sánchez A, Kazi A, Albarqouni S, Kirchhoff S, Sträter A, Biberthaler P, et al. *Weakly-supervised localization and classification of proximal femur fractures. arXiv preprint arXiv:1809.10692*; 2018.
91. Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, et al. Deep learning in medical imaging and radiation therapy. *Med Phys* 2019;**46**:e1–e36. Available from: https://doi.org/10.1002/mp.13264.
92. Ribli D, Horváth A, Unger Z, Pollner P, Csabai I. Detecting and classifying lesions in mammograms with Deep Learning. *Sci Rep* 2018;**8**. Available from: https://doi.org/10.1038/s41598-018-22437-z.
93. Diamant I, Hoogi A, Beaulieu CF, Safdari M, Klang E, Amitai M, et al. Improved patch-based automated liver lesion classification by separate analysis of the interior and boundary regions. *IEEE J Biomed Heal Inform* 2016;**20**:1585–94. Available from: https://doi.org/10.1109/JBHI.2015.2478255.
94. Yan K, Wang X, Lu L, Summers RM. DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *J Med Imaging* 2018;**5**:1. Available from: https://doi.org/10.1117/1.JMI.5.3.036501.
95. Yan K, Wang X, Lu L, Zhang L, Harrison A, Bagheri M, et al. Deep lesion graphs in the wild: relationship learning and organization of significant radiology image findings in a diverse large-scale lesion database. In: *CVPR*; 2018b.
96. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;**45**:228–47. Available from: https://doi.org/10.1016/j.ejca.2008.10.026.
97. Cai J, Tang Y, Lu L, Harrison AP, Yan K, Xiao J, et al. Accurate weakly-supervised deep lesion segmentation using large-scale clinical annotations: slice-propagated 3D mask generation from 2D RECIST. In: *MICCAI*; 2018b.
98. Yan K, Bagheri M, Summers RM. 3D context enhanced region-based convolutional neural network for end-to-end lesion detection. In: *MICCAI*; 2018c. pp. 511–9. Available from: https://doi.org/10.1007/978-3-030-00928-1_58.
99. Tang Y-B, Yan K, Tang Y-X, Liu J, Xiao J, Summers RM. Uldor: a universal lesion detector for CT scans with pseudo masks and hard negative example mining. In: *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019). Presented at the 2019a IEEE 16th international symposium on biomedical imaging (ISBI 2019)*; 2019b. pp. 833–6. Available from: https://doi.org/10.1109/ISBI.2019.8759478.
100. Yan K, Peng Y, Sandfort V, Bagheri M, Lu Z, Summers RM. Holistic and comprehensive annotation of clinically significant findings on diverse CT images: learning from radiology reports and label ontology. In: *CVPR*; 2019a. pp. 8515–24. Available from: https://doi.org/10.1109/CVPR.2019.00872.
101. Peng Y, Yan K, Sandfort V, Summers RM, Lu Z. A self-attention based deep learning method for lesion attribute detection from CT reports. In: *2019 IEEE international conference on healthcare informatics, ICHI 2019*; 2019. Available from: https://doi.org/10.1109/ICHI.2019.8904668.
102. Yan K, Tang Y, Peng Y, Sandfort V, Bagheri M, Lu Z, et al. MULAN: multitask universal lesion analysis network for joint lesion detection, tagging, and segmentation. In: *MICCAI*; 2019b. pp. 194–202. Available from: https://doi.org/10.1007/978-3-030-32226-7_22>.
103. Zlocha M, Dou Q, Glocker B. Improving RetinaNet for CT lesion detection with dense masks from weak RECIST labels. In: *MICCAI*; 2019. pp. 402–10. <https://doi.org/10.1007/978-3-030-32226-7_45.
104. Li Z, Zhang S, Zhang J, Huang K, Wang Y, Yu Y. MVP-Net: multi-view FPN with position-aware attention for deep universal lesion detection. In: *MICCAI*; 2019. pp. 13–21. Available from: https://doi.org/10.1007/978-3-030-32226-7_2.
105. Cai J, Harrison AP, Zheng Y, Yan K, Huo Y, Xiao J, et al. *Lesion harvester: iteratively mining unlabeled lesions and hard-negative examples at scale*. 2020. Available from: http://arxiv.org/abs/2001.07776.
106. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. In: *NIPS*; 2015. pp. 91–9. <https://doi.org/10.1109/TPAMI.2016.2577031>.
107. He K, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. In: *ICCV*; 2017. pp. 2980–8. Available from: https://doi.org/10.1109/ICCV.2017.322.