Current Pulmonology Reports (2019) 8:86–95
https://doi.org/10.1007/s13665-019-00229-8

PULMONARY RADIOLOGY (M STEPHENS AND S KAPUR, SECTION EDITORS)

Lung Nodule: Imaging Features and Evaluation in the Age of Machine Learning

M. Sayyouh [1] · L. M. Hadjiiyski [1] · H-P. Chan [1] · P. Agarwal [1]

Published online: 22 July 2019

Abstract

Purpose of Review With the unprecedented increase in chest CT studies, especially due to implementation of lung cancer screening, evaluation of lung nodules by radiologists can be exhausting and time-consuming. Machine learning promises to be a useful tool for detection and characterization of nodules. The purpose of this review is to evaluate the recent literature pertaining to machine learning in lung nodule detection and evaluation.

Recent Findings There has been a recent surge of publications pertaining to machine learning and its applications in chest imaging. Many studies have shown promising results for automatic detection and characterization of lung nodules. Other studies have shown combined performance of a radiologist and computer-assisted detection (CAD) out performed a single radiologist, CAD alone, and double readers. Although these recent advances heighten expectations, it is important for developers and users to be mindful of challenges such as training, validation, independent testing, and proper user training.

Summary Computer-aided technology can help radiologists in evaluating lung nodules especially with the large number of scans performed. Recent advances in machine learning are replacing traditional methods and could significantly change the way radiology is practiced.

Keywords Machine learning · Lung · Nodules · Detection · Evaluation · Characterization

Introduction

Lung cancer is the leading cause of cancer-related deaths in both men and women worldwide with a 5-year survival of only 10–15% [1]. National Lung Screening Trial (NLST) demonstrated that early detection of lung cancer leads to a 20% reduction in mortality [2]. Computed tomography (CT) plays a pivotal role in the detection and characterization of lung nodules guiding management. However, detection of nodules can be an arduous task and prone to errors with performance of radiologists in detecting lung nodules variable [3, 4]. Increasing workloads can also lead to treatment delays [5].

New advances in artificial intelligence (AI) have the potential to serve as an aid to radiologists by improving workflow and reducing variability in reporting [6•]. These advances in AI involve a subset of algorithms in machine learning that use computers to make predictions based on learning from examples [7]. These computer algorithms can learn complex relationships from large amounts of data and apply that to make accurate decisions [8]. Facial recognition on social media websites, voice recognition, and gauging customer preferences based on online shopping sites are just a few examples of machine learning in day to day life.

In medicine, recent advances in machine learning algorithms have made it easier to apply them to radiological images. In particular, deep learning algorithms have shown increased accuracy in image recognition tasks [9, 10]. Deep learning analyzes the input data for a task incrementally through its multilayer neurons following a hierarchy of simple to more complex representations. Each artificial neuron or groups of neurons in the multilayer architecture represents certain aspect of the task and together provide a complete representation of the output. The weight of each neuron is continuously adjusted as the model learns from the training examples. A deep neural network contains millions of adjustable weights and generally requires thousands of samples to

This article is part of the Topical Collection on *Pulmonary Radiology*

✉ M. Sayyouh
msayyouh@med.umich.edu

[1] Department of Radiology, Michigan Medicine, University of Michigan, 1500 E. Medical Center Drive, Ann Arbor, MI 48109, USA

properly learn the representations from an iterative training process.

Due to increased accuracy from these deep learning models machine learning is becoming an attractive aid to radiologists and pulmonary physicians, enabling a more objective and consistent evaluation, and better performance compared with traditional computer-aided diagnosis (CAD) systems that were limited by low sensitivity and high number of false positives.

Machine Learning Systems (Traditional and Deep Learning Methods)

Traditional CAD systems use a sequence of algorithms for lung detection, candidate selection, feature extraction, and false positive reduction with the output showing the likelihood of each remaining candidate being a nodule [11–15]. False positive reduction is often based on segmentation of nodule candidates from which handcrafted features are extracted. The features are then input to a machine learning algorithm that are trained to classify false and true positive nodules by supervised learning [16]. Most studies of traditional machine learning report sensitivity of lung nodule detection that ranges from 70 to 85% and false positives of 3 to 5 nodules per scan on average [17, 18]. A recent study by Firmino et al. reported a nodule detection sensitivity of 94% but with a higher false positive rate of 7 nodules per scan [19].

Deep learning algorithms use multiple processing layers and automatically build a sequence for detection of a pulmonary nodule. The primary benefit of deep learning is that hand-tuned features previously defined by computer vision experts are no longer needed [16]. It provides an automatic way to generalize knowledge learned from training data to future unknown test data in a more generalizable way [8]. It is currently the preferred method of machine learning with a very low number of false positives when compared with traditional techniques [16].

Convolutional neural networks (CNNs) are a subtype of deep neural networks used primarily in image analysis. A CNN is a type of artificial neural networks that has the capability of discovering useful features from the input data using convolutions eliminating the need of manually designed features. CNN was developed in the early 1980s in computer vision field for pattern recognition such as handwritten numerals [20, 21]. It was first introduced into medical imaging for computer-aided detection of lung nodules in 1993 [22]. Early CNNs had very few convolutional layers and few kernels in each layer due to limited computational power of computers and small training data sets. This, besides the high cost, limited the early adoption of CNN. However, with recent development of low-cost graphical processing units and memory for large data collection, more complex CNNs have been developed. Also, new techniques and training network strategies such as layer-wise unsupervised pre-training followed by supervised fine-tuning reduced the risk of overfitting and also increased training speed [23]. These advances allow millions of weights to be adjusted enabling CNNs to have more layers and potential pathways for feature identification. In 2012, Krizhevsky et al. showed that a deep convolutional neural network (DCNN) with 5 convolutional layers and 3 fully connected layers (AlexNet) could outperform other methods in ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [24•]. Subsequent studies show that deeper DCNNs lead to less errors for complex classification tasks.

To develop a good deep learning system, one must first define the input data and the desired output. High-quality data must then be collected with enough variability to generally cover the types of examples you might see. In general, the more variable the data is for a desired output, the more data that will be required. The next step is to divide the data into training, validation, and test set. The training set is much larger that the validation and test set. The training set is then run through the deep neural network with each example generating an output. This output is then compared with the desired output and a cost function comparing the two is used to update the weights through a process called back propagation. At the end of each training run, the performance of the training set is compared with the validation set to evaluate for overfitting of the model. Once a model that performs well on both the training and validation set is created, it is finally tested on the independent test set for evaluation of the final performance [8].

For robust DCNN system, a large sample of data from the population of interest is necessary. Many studies have shown the importance to train, validate, and test DCNNs with internal and external data as well as to understand what information a system has learned for a given task in developing a dependable system [25]. The collection of a large, well-curated data set is the main challenge to develop DCNNs for lung cancer detection in CT scans or any other medical imaging tasks.

Comparing CAD results in different studies is difficult due to the variety of algorithms used, different evaluation methods, database sizes, and nodule characteristics [26]. Additionally, both sensitivity and specificity and their robustness against variabilities in CT scan protocols are important factors in evaluation of algorithm performance as these variabilities are common in clinical settings [26]. In an effort to evaluate machine learning algorithms in detection and classification of lung nodules, objective evaluation frameworks, called "Challenges," have been developed (such as Data Science Bowl, LUNGx, and LUNA16) [27–29]. These challenges are expected to provide a valuable source for medical imaging research community in the near future.

Evolution of Machine Learning in Evaluation of Lung Nodules

Algorithms for analysis of radiographic images first surfaced in the mid-1960s [30]; however, they did not attract much interest at that time due to low computational power and lack of high-quality digitized images. In the mid-1980s to mid-1990s, CAD algorithms for detection and diagnosis of cancer in chest radiographs were developed [31] which then expanded to include other modalities such as computed tomography (CT) [32]. Most of the early CAD technology depended on traditional machine leaning methods and algorithms [6].

In the early-1990s to 2000, CNNs, a key feature of deep learning systems, were introduced into the CAD field in medical imaging [22, 33–35]. However, CNNs were difficult to train and started to lose popularity in favor of traditional approaches that seemed to be outperforming CNNs at that time most likely due to limited data sets available for training at that time. Later in the 2000s, trials continued using traditional machine learning methods aiming to further increase sensitivity and reduce false positives which were the major constraints for these systems [12, 14, 36, 37].

In the last decade, there has been a revival of CNN due to a combination of factors such as availability of large data sets, development of low-cost and powerful graphical processing units, and technical advancements leading to faster training of network [6, 38]. In 2012, Krizhevsky introduced DCNNs which was the start of a new age in machine learning [24].

Applications and Performance of Machine Learning in Evaluation of Lung Nodules

I. Detection of lung nodules:

Detection of lung nodules is one of the major applications of CAD systems. Several studies have shown that CAD systems detect lung nodules in a different way compared with radiologists [39, 40]. Studies have also shown that the ability of a machine learning system to detect nodules can be affected by factors such as nodule size, location, composition, and other lung abnormalities that may be mistakenly detected as nodules (false positives) as well as technical factors that affect image quality such as reconstruction filter and section thickness. Features that might cause false positives in CAD systems include scars, bronchial wall thickening, vessel bifurcations, and sometimes motion artifacts [26].

Sensitivity of CAD decreases with decrease in nodule size. For example, in a study by Brown et al., detection sensitivity of CAD for nodules > 3 mm was 100% and dropped to 70% for nodules < 3 mm [41]. However, another study that compared 6 different CAD algorithms in detection of lung nodules showed that five out of the six tested systems had better sensitivity in small lung nodules compared with the larger ones [42]. This was explained by smaller nodules being more frequently isolated and more abundant in training data sets [42]. Hence, apart from size, nodule location plays a key role. The highest sensitivity is seen for isolated nodules and decreases for juxtavascular and juxtapleural nodules [42, 43]. Bae et al. found the sensitivities for isolated, juxtavascular, and juxtapleural nodules to be 97.4%, 94.1% and 92.3% respectively [43]. It has been demonstrated that CAD sensitivity in detecting small isolated nodules (< 5 mm) was higher than that of a radiologist [40], although radiologists outperform CAD in larger nodules (> 5 mm) and those that are adjacent to other structures [40]. This observation makes CAD systems a viable complementary tool to radiologists. Sahiner et al. studied the effect of CAD on radiologists' performance in nodule detection for nodule sizes greater than 3, 4, 5, and 6 mm, and showed that CAD could improve the radiologists' detection for all four thresholds but the improvement achieved statistical significance for thresholds of 3 and 4 mm [44].

Most of the CAD systems that have been approved for clinical use have generally been directed towards solid nodules which have well-defined spherical shape, uniform density, and high contrast to the surrounding lung parenchyma [45]. Sub-solid lung nodules are difficult to detect by CAD due to low contrast compared with surrounding lung parenchyma and poorly defined margins [26]. While solid nodules are more common, sub-solid nodules are more likely to be malignant [46]. In a study by Henscke et al., 63% of part-solid nodules and 18% of ground-glass nodules were malignant compared with only 7% of the solid nodules [46]. Relatively small number of studies have evaluated sensitivity of CAD systems in detection of sub-solid lung nodules including pure ground glass and part-solid nodules [45], but have consistently shown inferior performance compared with solid nodules. It has been shown that sub-solid nodules are more likely to be missed [47], with detection rate for part-solid nodules ranging from 72 to 85% [48, 49] and for pure ground-glass nodules being approximately 49% [49]. These results are consistent with the fact that most algorithms are based on attenuation differences which are larger between normal lung parenchyma and solid nodules than ground-glass nodules.

In addition to nodule characteristics, CT parameters (such as section thickness) also affect CAD performance [45]. Thin-section CT has been shown to enhance nodule detection compared with larger section thickness. Nodule detection improves with decreasing section thickness and reconstruction intervals, however, at the cost of the larger volume of data that can cause significantly longer reading time [50, 51]. Narayanan et al. showed that 2.5 mm is the most effective in terms of accuracy, dosage level, computation, and memory consumption [52]. CAD is not recommended for section thickness of 4 mm or greater [39, 50].

Studies have shown that radiation dose can be significantly reduced by lowering tube current without compromising detection rates by CAD systems. Hein et al. compared two CAD systems for scans done with two different tube currents of 75 mAs and 5 mAs and found no significant difference in nodule detection rates [53]. In another study, Lee et al. compared sensitivities for detection of lung nodules with 32, 16, 8, and 4 mAs. They showed that there was no significant difference in nodule detectability between scans performed at 16 mAs, 8 mAs, and 32 mAs. However, scans done at 4 mAs had significantly lower nodule detection sensitivity compared with 32 mAs. They concluded that within certain range, nodule detection accuracy does not deteriorate. This was explained by the inherent high contrast within the aerated lungs and the less effect of image noise, produced at lower does, compared with solid organs [54].

Selected studies for the performance of CAD in detection of lung nodules are shown in Table 1.

II. Characterization/classification of lung nodules:

Many studies have been conducted to evaluate if machine learning algorithms using extracted nodule features such as textural and geometric features, density, shape, surface curvature, margin, and lung parenchyma surrounding the nodule can help radiologists differentiate between benign and malignant nodules [58]. In a study by Song et al., researchers tested 3 deep learning techniques for classifying benign and malignant nodules using size and textural features and they concluded that CNN had the best performance [59]. Ferreira Jr. et al. used nodule margin sharpness beside nodule texture as descriptors for classification of lung nodules. They reported statistically significant improvements on sensitivity of the CAD system in classification of benign and malignant nodules [58]. Tu and colleagues evaluated the use of CNN for automatic categorization of solid, part-solid, and non-solid nodules. In their study, no image segmentation processing was needed, avoiding potential errors caused by inaccurate image processing. They concluded that adoption of CNN-based CAD systems can improve the performance of CAD, reduce inter-observer variation, and provide reference for further nodule analysis [60]. More recently, Ciompi et al. presented a deep learning system based on multistream multiscale convolutional networks which can automatically classify all nodule types relevant for nodule work-up. They categorized nodules into 4 main categories: solid, non-solid, part-solid, and calcified nodules. Two subcategories of solid nodules were perifissural nodules and spiculated nodules. They showed that the deep learning system exceeds the performance of classical machine learning approaches and is within the inter-observer variability of four experienced radiologists [61]. In another study by Nishio et al., they developed a computer-aided diagnosis method for differentiation of benign nodules, primary lung cancer, and metastatic cancer and compared DCNN with and without transfer learning with a conventional handcrafted method. They concluded that classification was better using a DCNN compared with the conventional method and transfer learning improved image recognition [62].

Selected studies for the performance of CAD in characterization of lung nodules are shown in Table 2.

Nodule Volumetry Computer-aided volumetry can provide an accurate and more reproducible measurement for nodules enabling accurate assessment of nodule growth rate and also response to treatment. Growth rate as indicated by “doubling time” is used as a predictor of malignancy [26]. In the Dutch-Belgian Lung Cancer Screening Trial (NELSON), nodule management was based on volumetric nodule measurement. Nodule growth was defined as increase of volume by at least 25%. In case of part-solid nodules, only the solid part was used for volumetry [69].

Lung nodule volumetry can also be used in the assessment of response to treatment. It has been suggested that change in

Table 1 Selected CAD studies in detection of lung nodules (for additional studies before 2009, please see reference [15])

Study	Year	Database	Number of nodules	Section thickness (mm)	Nodule size (mm)	Sensitivity (%)	False positive rate
Sahiner et al. [44]	2009	Private	118	1.5–3.0	3.1–19.6	78	5.5/scan
Riccardi et al. [55]	2011	LIDC-IDRI	154	0.5–3	≥3	71	6.5/scan
Guo and Li [13]	2012	LIDC-IDRI	111	1.25–3	≥3	80	2.8/scan
Cascio et al. [11]	2012	LIDC-IDRI	148	1.25–3	≥3	97	2.5/scan
van Ginneken et al. [42]	2015	LIDC-IDRI	1147	≤2.5	≥3	76	1.0/scan
Setio et al. [56]	2016	LIDC-IDRI	1186	≤2.5	≥3	85.4	1.0/scan
Dou et al. [57]	2017	LIDC-IDRI	1186	≤2.5	≥3	90.7	1.0/scan

Table 2 Selected CAD studies in classification of lung nodules (for additional studies before 2009, please see reference [15])

Study	Year	Database	Number of nodules	Accuracy (%)	AUC
Way et al. [63]	2009	Private	256	N/A	0.863
El Baz [64]	2011	LIDC	327	93.6	N/A
Kumar et al. [65]	2015	LIDC-IDRI	4323	75.01	N/A
Song et al. [59]	2017	LIDC-IDRI	5204	84.15	N/A
Wei et al. [66]	2017	LIDC-IDRI	366	91.8	0.986
Zhao et al. [67]	2018	LDIC-IDRI	1018	82.2	0.877
Xie et al. [68]	2019	LDIC-IDRI	1945	91.6	0.957

tumor volume is more sensitive when compared with conventional unidimensional measurement after targeted therapy for lung cancer [70].

Volumetric measurement of lung nodules can be affected by various factors including nodule size, characteristics, technical parameters of the CT examination, and patient-related factors [26]. Small lung nodules, due to partial volume effect, are subject to greater measurement errors [71]. Nodules that come in contact with surrounding vessels or pleura are more difficult to segment from the background. Also, changes in lung volumes during inspiration and expiration can affect lung nodule volume measurement, leading to 23% difference in nodule volume measurement [72]. Technical scan factors such as section thickness significantly impacts volumetry. Thin CT sections reduce partial volume effect and therefore enable more accurate assessment [71, 73–77]. On the contrary, tube current (20 mAs versus 100 mAs) was not shown to have a significant impact [74].

Most studies for lung nodule volumetric measurements have been done on solid lung nodules. Only few studies regarding volumetry of sub-solid lung nodules have been reported [26] with higher error values for ground-glass nodules than solid nodules [78].

III. Quantitative analysis of image features:

Radiomics, defined as automatic extraction of quantitative features from medical images, turns image voxels into set of numbers that characterize the biological property of interest such as malignancy, tumor grade, or therapy response [79]. Recently, CNNs were used for quantitative analysis using feature learning, unlike the engineered pre-selected radiomics/texture analysis approach [80]. Detailed discussion of radiomics is beyond the scope of this review.

IV. Evaluation of treatment response:

CAD systems can be an important tool in evaluation of treatment response in lung cancer patients. As discussed above, volumetric nodule measurements can be valuable in this setting. Besides the volumetric change, other radiomics features may also be useful for assessing tumor response to treatment [81]. An additional benefit of machine learning is image registration and automatic nodule matching which significantly reduces the time involved with manual matching and measurement of nodules on follow-up studies [26]. Lee et al. evaluated the performance of automated matching software using two serial CT scans (at 5-mm section thickness). The matching rate was highly influenced by the difference in the lung configurations between the two scans ranging from as high as 82% (in patients with relatively unchanged configuration) to 29% (in patients with substantial change in configuration) [82]. In screening studies, however, the interval configuration change is usually less with reported matching rates of 91–93% [83, 84].

V. Lung nodule biopsy:

A recent study by Sumathipala et al. suggested that machine learning tools can predict whether a lung nodule is best biopsied surgically or by a minimally invasive procedure. In their study, an algorithm incorporating semantic (provided by 4 expert radiologists) and computational imaging features using the public domain imaging data from the Lung Image Database Consortium Image Collection (LIDC-IDRI) was developed. These features included nodule size (3D volume), shape (sphericity, spiculation, lobulation), accessibility (distance from trachea to the nodule and distance to outer skin), and composition (calcification, texture, and internal structure). They found that the most informative features were nodule spiculation, volume, and maximum distance to outer skin. This was suggested to be helpful in deciding between surgical biopsy and minimally invasive procedure [85].

Machine Learning and Lung Cancer Screening

Implementation of lung cancer screening programs has led to an unprecedented increase in the number of chest CT scans. These screening scans are reported based on a standardized Lung CT Reporting And Data System (Lung-RADS) [86]. The detection of nodules, evaluation, and reporting of management recommendations can be tedious and time-consuming [16].

If a CAD system is designed and validated properly to be a concurrent reader, a first reader, or even a pre-screener that can reliably exclude negative cases from radiologists' reading, it can improve workflow and make reading more efficient [61], although the performance criteria for CAD to do so still needs to be determined. Applications of CAD systems in lung cancer screening may include lung nodule detection, risk assessment, calculation of their 2D dimensions and 3D volume, tracking size over time to assess growth, and providing appropriate follow-up recommendations. CAD systems can also provide a consistent and reliable way to evaluate nodules, avoiding inter-reader variability by radiologists. Many studies have shown that the diagnostic image quality does not suffer at low doses [87–89]. However, the effect of low-dose CT techniques on accuracy of CAD systems is still under investigation.

Machine Learning and Radiologists

Various factors affect the performance of radiologists in detection of lung nodules such as nodule characteristics (size, location, attenuation) as well as observer experience. Besides, these challenges, fatigue, emotional state, distractions during reading time, satisfaction of search, and environment can have an important impact [8].

Although double reading by 2 radiologists was shown to increase sensitivity of lung nodule detection [90, 91], it is considered time-consuming and comes at a high cost which might not be practical. Machine learning has the potential to assist radiologists and reduce reading time at a lower cost. Observer studies demonstrated that lung nodule detection sensitivity in CT by radiologists improved significantly when reading with CAD [15]. Also, CAD systems have been proven to have higher sensitivity in detection of lung nodules than double reading by radiologists [92, 93]. In a study by Zhao et al., 22% of nodules (randomly selected from the NELSON study) were detected only by CAD and missed by 2 radiologist readers. Three percent of these nodules were diagnosed to be cancer in the following year [92]. Lee et al. showed that CAD sensitivity as a standalone tool (81%) is not significantly different from that of radiologists alone (85%) [40].

Beyer et al. suggested that integration of CAD into clinical practice can be realized in 3 different ways:

– First reader (where CAD functions as a screener and only CAD detected imaging slices are presented to the reader).
– Second reader (where CAD findings are reviewed by the radiologist as a second step after the initial read).
– Concurrent reader (where CT scan is read by radiologist and CAD findings are displayed concurrently).

In their study, they evaluated the performance of CAD as second reader and concurrent reader and found that sensitivity of CAD as a second reader was higher than without CAD or with CAD as a concurrent reader. However, in the same study, the reading time was significantly shorter with CAD as a concurrent reader compared with that without CAD or with CAD as a second reader [94]. So CAD can be envisioned as a tool that can improve sensitivity when used as a second reader (at the cost of increasing reading time) or as a tool that improves efficiency when used as a concurrent reader (by reducing the reading time and without losing sensitivity).

Christe et al. investigated the best pairing of first and second reader for detecting lung nodules with CT at various dose levels. They paired 2 radiologists and 3 different CAD software to find the highest sensitivity. They found that the highest sensitivity (between 97 and 99%) was achieved by combining any radiologist with any CAD at any dose level. Combining any two CADs, sensitivity was significantly lower (85% and 88%) [95]. The value of CAD as a second reader has also been shown by several other observer studies [44].

Most of the data available, however, concerns only solid nodules [48]. A study by Yanagawa et al. showed that only 21% of 102 ground-glass nodules were detected by CAD compared with 60–80% detected by radiologists [49].

Limitations and Future of Machine Learning in Evaluation of Lung Nodules

Despite the advantages machine learning has shown, there are still barriers for widespread application in clinical practice [96]. One of the major challenges in developing an accurate CNN for machine learning is the requirement of a large data set for training and validation. Also, the generalizability of the developed machine learning model should be evaluated with independent testing data sets [6]. These data sets should be representative of the population, imaging equipment, and acquisition protocols in the clinical setting for which it is going to be used. Collecting such data sets, however, can be very expensive.

A machine learning algorithm trained on a small data set may not perform well on a large data set as distribution of features may differ [8]. Overfitting (overtraining) is a known problem with small training sets [6, 38]. With overfitting, a classifier models the small training set very well so that it fails to generalize on new unseen data. Several ways have been suggested to reduce overfitting including regularization, early stopping, and dropout [97].

To alleviate this problem in medical imaging, "transfer learning" has been used. In transfer learning, a DCNN already trained with large data from a different task (pre-trained model) can be adapted to a new target task by further fine-tuning it with data from the target domain [98]. Although this might help to a certain extent, the performance of a pre-trained DCNN still depends on the size of the training data set [99]. Another way to reduce the limited data set problem is "data augmentation." Data augmentation generates multiple slightly different variations of images from the original data set. Although this has been shown to reduce overfitting [24], it is not as effective as increasing the real training sample size. Moreover, if the original training set lacks the representation of certain imaging features, data augmentation will not fill the gaps. Digitally generated artificial lesions to represent certain characteristics have also been explored [100].

Conclusion

In the current clinical practice, efficiency and costs are major considerations. Machine learning in detection and characterization of lung nodules, especially with the recent success of deep learning methods, has sparked the interest to develop more advanced CAD systems that can potentially save time and improve accuracy. With the availability of big data sets, advances in deep learning algorithms, and processing power machine learning models, will continue to improve. Furthermore, systems that can integrate clinical data and molecular biomarkers in addition to the imaging data can be great tools to support clinical decision-making in the near future. However, it is very important for developers and users to understand the importance of large training data sets, independent testing, and validation of its generalizability both retrospectively and prospectively [101]. It is also important that pulmonologists become aware of these systems and if they are being used by the radiologists as this can have an effect on performance depending on the system used. Proper quality assurance and monitoring after clinical implementation, user training, and mindfulness of machine learning limitations are also essential to increase the overall performance and efficiency in the clinical practice [102].

Compliance with Ethical Standards

Conflict of Interest M. Sayyouh, L. M. Hadjiiyski, C-H. Chan, and P. Agarwal declare no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance

1. World Cancer Report, World Health Organization (WHO) 2014. http://publications.iarc.fr/Non-Series-Publications/World-Cancer-Reports/World-Cancer-Report-2014. Access 17 July 2019.
2. National Lung Screening Trial Research, T, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;**365**(5):395–409.
3. Naidich DP, Bankier AA, MacMahon H, Schaefer-Prokop CM, Pistolesi M, Goo JM, et al. Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the Fleischner Society. Radiology. 2013;266(1):304–17.
4. Balekian AA, Silvestri GA, Simkovich SM, Mestaz PJ, Sanders GD, Daniel J, et al. Accuracy of clinicians and models for estimating the probability that a pulmonary nodule is malignant. Ann Am Thorac Soc. 2013;10(6):629–35.
5. Awai K, Murao K, Ozawa A, Nakayama Y, Nakaura T, Liu D, et al. Pulmonary nodules: estimation of malignancy at thin-section helical CT–effect of computer-aided diagnosis on performance of radiologists. Radiology. 2006;239(1):276–84.
6.• Sahiner B, et al. Deep learning in medical imaging and radiation therapy. Med Phys. 2019;**46**(1):e1–e36. **This paper provide up to date information about applications of deep learning in medical imaging in general including lungs**.
7. Bastanlar Y, Ozuysal M. Introduction to machine learning. Methods Mol Biol. 2014;1107:105–28.
8. Wang S, Summers RM. Machine learning and radiology. Med Image Anal. 2012;16(5):933–51.
9. McCarville MB, Lederman HM, Santana VM, Daw NC, Shochat SJ, Li CS, et al. Distinguishing benign from malignant pulmonary nodules with helical chest CT in children with malignant solid tumors. Radiology. 2006;239(2):514–20.
10. Cheng JZ, Ni D, Chou YH, Qin J, Tiu CM, Chang YC, et al. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. Sci Rep. 2016;6:24454.
11. Cascio D, Magro R, Fauci F, Iacomi M, Raso G. Automatic detection of lung nodules in CT datasets based on stable 3D mass-spring models. Comput Biol Med. 2012;42(11):1098–109.
12. Ge Z, Sahiner B, Chan HP, Hadjiiski LM, Cascade PN, Bogot N, et al. Computer-aided detection of lung nodules: false positive reduction using a 3D gradient field method and 3D ellipsoid fitting. Med Phys. 2005;32(8):2443–54.
13. Guo W, Li Q. High performance lung nodule detection schemes in CT using local and global information. Med Phys. 2012;39(8):5157–68.
14. Gurcan MN, Sahiner B, Petrick N, Chan HP, Kazerooni EA, Cascade PN, et al. Lung nodule detection on thoracic computed tomography images: preliminary evaluation of a computer-aided diagnosis system. Med Phys. 2002;29(11):2552–8.
15. Chan HP, Hadjiiski L, Zhou C, Sahiner B. Computer-aided diagnosis of lung cancer and pulmonary embolism in computed tomography-a review. Acad Radiol. 2008;15(5):535–55.
16. Gruetzemacher R, Gupta A. Using deep learning for pulmonary nodule detection & diagnosis. Conference Proceedings of the 22nd Americas Conference on Information Systems (AMCIS). San Diego, California, United States; 2016.
17. Messay T, Hardie RC, Rogers SK. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. Med Image Anal. 2010;14(3):390–406.

18. Armato SG 3rd, et al. Computerized detection of pulmonary nodules on CT scans. Radiographics. 1999;19(5):1303–11.
19. Firmino M, Angelo G, Morais H, Dantas MR, Valentim R. Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy. Biomed Eng Online. 2016;15:2.
20. Fukushima K, Miyake S. Neocognitron - a new algorithm for pattern-recognition tolerant of deformations and shifts in position. Pattern Recogn. 1982;15(6):455–69.
21. LeCun, Y., et al. Handwritten digit recognition with a back-projection network. Conference Proceedings of the 2nd international conference on Neural Information Processing Systems (NIPS). 1989; 396-404.
22. Lo SCB, et al. Computer-assisted diagnosis of lung nodule detection using artificial convolution neural-network. Medical Imaging 1993: Image Processing, 1993. 1898, p. 859–869.
23. Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neural Comput. 2006;18(7):1527–54.
24.• Krizhevsky, A., I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. in 25th International Conference on Neural Information Processing Systems. 2012. Lake Tahoe, Nevada. **This paper was a major paper that showed the value of deep learning and how it can outperform traditional methods of machine learning**.
25. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. PLoS Med. 2018;15(11):e1002683.
26. Goo JM. A computer-aided diagnosis for evaluating lung nodules on chest CT: the current status and perspective. Korean J Radiol. 2011;12(2):145–55.
27. Armato SG 3rd, et al. LUNGx challenge for computerized lung nodule classification. J Med Imaging (Bellingham). 2016;3(4):044506.
28. Setio AAA, Traverso A, de Bel T, Berens MSN, Bogaard C, Cerello P, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. Med Image Anal. 2017;42:1–13.
29. Data Science Bowl https://datasciencebowl.com/competitions/turning-machine-intelligence-against-lung-cancer/. 2016–2017. Access 1 April 2019.
30. Meyers PH, Nice CM Jr, Becker HC, Nettleton WJ Jr, Sweeney JW, Meckstroth GR. Automated computer analysis of radiographic images. Radiology. 1964;83:1029–34.
31. Giger ML, Doi K, MacMahon H. Image feature analysis and computer-aided diagnosis in digital radiography. 3. Automated detection of nodules in peripheral lung fields. Med Phys. 1988;15(2):158–66.
32. Kanazawa K, Kawata Y, Niki N, Satoh H, Ohmatsu H, Kakinuma R, et al. Computer-aided diagnosis for pulmonary nodules based on helical CT images. Comput Med Imaging Graph. 1998;22(2):157–67.
33. Lin, J.S., et al., Application of artificial neural networks for reduction of false-positive detections in digital chest radiographs. Conference Proceedings of the Annual Symposium on Computer Application in Medical Care. 1993;434–8.
34. Lo SB, et al. Artificial convolution neural network techniques and applications for lung nodule detection. IEEE Trans Med Imaging. 1995;14(4):711–8.
35. Lo SC, Freedman MT, Lin JS, Mun SK. Automatic lung nodule detection using profile matching and back-propagation neural network techniques. J Digit Imaging. 1993;6(1):48–54.
36. McCulloch CC, Kaucic RA, Mendonça PRS, Walter DJ, Avila RS. Model-based detection of lung nodules in computed tomography exams. Thoracic computer-aided diagnosis. Acad Radiol. 2004;11(3):258–66.
37. Way TW, Hadjiiski LM, Sahiner B, Chan HP, Cascade PN, Kazerooni EA, et al. Computer-aided diagnosis of pulmonary nodules on CT scans: segmentation and classification using 3D active contours. Med Phys. 2006;33(7):2323–37.
38. Chan HP, Sahiner B, Wagner RF, Petrick N. Classifier design for computer-aided diagnosis: effects of finite sample size on the mean performance of classical and neural network classifiers. Med Phys. 1999;26(12):2654–68.
39. Marten K, Engelke C, Seyfarth T, Grillhösl A, Obenauer S, Rummeny EJ. Computer-aided detection of pulmonary nodules: influence of nodule characteristics on detection performance. Clin Radiol. 2005;60(2):196–206.
40. Lee JW, Goo JM, Lee HJ, Kim JH, Kim S, Kim YT. The potential contribution of a computer-aided detection system for lung nodule detection in multidetector row computed tomography. Investig Radiol. 2004;39(11):649–55.
41. Brown MS, Goldin JG, Suh RD, McNitt-Gray MF, Sayre JW, Aberle DR. Lung micronodules: automated method for detection at thin-section CT–initial experience. Radiology. 2003;226(1):256–62.
42. van Ginneken B, Armato SG III, de Hoop B, van Amelsvoort-van de Vorst S, Duindam T, Niemeijer M, et al. Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the ANODE09 study. Med Image Anal. 2010;14(6):707–22.
43. Bae KT, Kim JS, Na YH, Kim KG, Kim JH. Pulmonary nodules: automated detection on CT images with morphologic matching algorithm–preliminary results. Radiology. 2005;236(1):286–93.
44. Sahiner B, Chan HP, Hadjiiski LM, Cascade PN, Kazerooni EA, Chughtai AR, et al. Effect of CAD on radiologists' detection of lung nodules on thoracic CT scans: analysis of an observer performance study by nodule size. Acad Radiol. 2009;16(12):1518–30.
45. Godoy MC, et al. Benefit of computer-aided detection analysis for the detection of subsolid and solid lung nodules on thin- and thick-section CT. AJR Am J Roentgenol. 2013;200(1):74–83.
46. Henschke CI, Yankelevitz DF, Mirtcheva R, McGuinness G, McCauley D, Miettinen OS. CT screening for lung cancer: frequency and significance of part-solid and nonsolid nodules. AJR Am J Roentgenol. 2002;178(5):1053–7.
47. Yuan R, Vos PM, Cooperberg PL. Computer-aided detection in screening CT for pulmonary nodules. AJR Am J Roentgenol. 2006;186(5):1280–7.
48. Benzakoun J, Bommart S, Coste J, Chassagnon G, Lederlin M, Boussouar S, et al. Computer-aided diagnosis (CAD) of subsolid nodules: evaluation of a commercial CAD system. Eur J Radiol. 2016;85(10):1728–34.
49. Yanagawa M, Honda O, Yoshida S, Ono Y, Inoue A, Daimon T, et al. Commercially available computer-aided detection system for pulmonary nodules on thin-section images using 64 detectors-row CT: preliminary study of 48 cases. Acad Radiol. 2009;16(8):924–33.
50. Kim JS, Kim JH, Cho G, Bae KT. Automated detection of pulmonary nodules on CT images: effect of section thickness and reconstruction interval–initial results. Radiology. 2005;236(1):295–9.
51. White CS, Pugatch R, Koonce T, Rust SW, Dharaiya E. Lung nodule CAD software as a second reader: a multicenter study. Acad Radiol. 2008;15(3):326–33.
52. Narayanan BN, Hardie RC, Kebede TM. Performance analysis of a computer-aided detection system for lung nodules in CT at different slice thicknesses. J Med Imaging (Bellingham). 2018;5(1):014504.
53. Hein PA, et al. Computer-aided pulmonary nodule detection - performance of two CAD systems at different CT dose levels. Rofo. 2009;181(11):1056–64.

54. Lee JY, Chung MJ, Yi CA, Lee KS. Ultra-low-dose MDCT of the chest: influence on automated lung nodule detection. Korean J Radiol. 2008;9(2):95–101.
55. Riccardi A, Petkov TS, Ferri G, Masotti M, Campanini R. Computer-aided detection of lung nodules via 3D fast radial transform, scale space representation, and Zernike MIP classification. Med Phys. 2011;38(4):1962–71.
56. Setio AA, et al. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Trans Med Imaging. 2016;35(5):1160–9.
57. Dou Q, Chen H, Yu L, Qin J, Heng PA. Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection. IEEE Trans Biomed Eng. 2017;64(7):1558–67.
58. Ferreira JR Jr, Oliveira MC, de Azevedo-Marques PM. Characterization of pulmonary nodules based on features of margin sharpness and texture. J Digit Imaging. 2018;31(4):451–63.
59. Song Q, et al. Using deep learning for classification of lung nodules on computed tomography images. J Healthc Eng. 2017;2017:8314740.
60. Tu X, Xie M, Gao J, Ma Z, Chen D, Wang Q, et al. Automatic categorization and scoring of solid, part-solid and non-solid pulmonary nodules in CT images with convolutional neural network. Sci Rep. 2017;7(1):8533.
61. Ciompi F, Chung K, van Riel SJ, Setio AAA, Gerke PK, Jacobs C, et al. Towards automatic pulmonary nodule management in lung cancer screening with deep learning. Sci Rep. 2017;7:46479.
62. Nishio M, Sugiyama O, Yakami M, Ueno S, Kubo T, Kuroda T, et al. Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning. PLoS One. 2018;13(7):e0200721.
63. Way TW, Sahiner B, Chan HP, Hadjiiski L, Cascade PN, Chughtai A, et al. Computer-aided diagnosis of pulmonary nodules on CT scans: improvement of classification performance with nodule surface features. Med Phys. 2009;36(7):3086–98.
64. El-Baz A, et al. 3D shape analysis for early diagnosis of malignant lung nodules. Med Image Comput Comput Assist Interv. 2011;14(Pt 3):175–82.
65. Kumar, D., A. Wong, and D.A. Clausi, Lung Nodule Classification Using Deep Features in CT Images. 2015 12th Conference on Computer and Robot Vision Crv 2015, 2015: p. 133–138.
66. Wei G, et al. Content-based image retrieval for lung nodule classification using texture features and learned distance metric. J Med Syst. 2017;42(1):13.
67. Zhao X, Liu L, Qi S, Teng Y, Li J, Qian W. Agile convolutional neural network for pulmonary nodule classification using CT images. Int J Comput Assist Radiol Surg. 2018;13(4):585–95.
68. Xie Y, Xia Y, Zhang J, Song Y, Feng D, Fulham M, et al. Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT. IEEE Trans Med Imaging. 2019;38(4):991–1004.
69. van Klaveren RJ, Oudkerk M, Prokop M, Scholten ET, Nackaerts K, Vernhout R, et al. Management of lung nodules detected by volume CT scanning. N Engl J Med. 2009;361(23):2221–9.
70. Zhao B, Schwartz LH, Moskowitz CS, Ginsberg MS, Rizvi NA, Kris MG. Lung cancer: computerized quantification of tumor response–initial results. Radiology. 2006;241(3):892–8.
71. Goo JM, Tongdee T, Tongdee R, Yeo K, Hildebolt CF, Bae KT. Volumetric measurement of synthetic lung nodules with multi-detector row CT: effect of various image reconstruction parameters and segmentation thresholds on measurement accuracy. Radiology. 2005;235(3):850–6.
72. Goo JM, Kim KG, Gierada DS, Castro M, Bae KT. Volumetric measurements of lung nodules with multi-detector row CT: effect of changes in lung volume. Korean J Radiol. 2006;7(4):243–8.
73. Zhao B, Schwartz LH, Moskowitz CS, Wang L, Ginsberg MS, Cooper CA, et al. Pulmonary metastases: effect of CT section thickness on measurement–initial experience. Radiology. 2005;234(3):934–9.
74. Das M, Ley-Zaporozhan J, Gietema HA, Czech A, Mühlenbruch G, Mahnken AH, et al. Accuracy of automated volumetry of pulmonary nodules across different multislice CT scanners. Eur Radiol. 2007;17(8):1979–84.
75. Petrou M, Quint LE, Nan B, Baker LH. Pulmonary nodule volumetric measurement variability as a function of CT slice thickness and nodule morphology. AJR Am J Roentgenol. 2007;188(2):306–12.
76. Kuhnigk JM, Dicken V, Bornemann L, Bakai A, Wormanns D, Krass S, et al. Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans. IEEE Trans Med Imaging. 2006;25(4):417–34.
77. Ravenel JG, Leue WM, Nietert PJ, Miller JV, Taylor KK, Silvestri GA. Pulmonary nodule volume: effects of reconstruction parameters on automated measurements–a phantom study. Radiology. 2008;247(2):400–8.
78. Ko JP, Rusinek H, Jacobs EL, Babb JS, Betke M, McGuinness G, et al. Small pulmonary nodules: volume measurement at chest CT–phantom study. Radiology. 2003;228(3):864–70.
79. Aerts HJ, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.
80. Kadir T, Gleeson F. Lung cancer prediction using machine learning and advanced imaging techniques. Transl Lung Cancer Res. 2018;7(3):304–12.
81. Khorrami, M., et al. A combination of intra- and peritumoral features on baseline CT scans is associated with overall survival in non-small cell lung cancer patients treated with immune check point inhibitors: a multi-agent multi-site study. Proc. SPIE 10950, Medical Imaging 2019: Computer-Aided Diagnosis, 109500R. 2019 https://doi.org/10.1117/12.2513001.
82. Lee KW, Kim M, Gierada DS, Bae KT. Performance of a computer-aided program for automated matching of metastatic pulmonary nodules detected on follow-up chest CT. AJR Am J Roentgenol. 2007;189(5):1077–81.
83. Beigelman-Aubry C, Raffy P, Yang W, Castellino RA, Grenier PA. Computer-aided detection of solid lung nodules on follow-up MDCT screening: evaluation of detection, tracking, and reading time. AJR Am J Roentgenol. 2007;189(4):948–55.
84. Tao C, Gierada DS, Zhu F, Pilgram TK, Wang JH, Bae KT. Automated matching of pulmonary nodules: evaluation in serial screening chest CT. AJR Am J Roentgenol. 2009;192(3):624–8.
85. Sumathipala Y, Shafiq M, Bongen E, Brinton C, Paik D. Machine learning to predict lung nodule biopsy method using CT image features: a pilot study. Comput Med Imaging Graph. 2019;71:1–8.
86. McKee BJ, Regis SM, McKee AB, Flacke S, Wald C. Performance of ACR lung-RADS in a clinical CT lung screening program. J Am Coll Radiol. 2015;12(3):273–6.
87. Christe A, Torrente JC, Lin M, Yen A, Hallett R, Roychoudhury K, et al. CT screening and follow-up of lung nodules: effects of tube current-time setting and nodule size and density on detectability and of tube current-time setting on apparent size. AJR Am J Roentgenol. 2011;197(3):623–30.
88. Christe A, Charimo-Torrente J, Roychoudhury K, Vock P, Roos JE. Accuracy of low-dose computed tomography (CT) for detecting and characterizing the most common CT-patterns of pulmonary disease. Eur J Radiol. 2013;82(3):e142–50.
89. Rusinek H, Naidich DP, McGuinness G, Leitman BS, McCauley DI, Krinsky GA, et al. Pulmonary nodule detection: low-dose versus conventional CT. Radiology. 1998;209(1):243–9.
90. Wormanns D, Ludwig K, Beyer F, Heindel W, Diederich S. Detection of pulmonary nodules at multirow-detector CT:

effectiveness of double reading to improve sensitivity at standard-dose and low-dose chest CT. Eur Radiol. 2005;15(1):14–22.
91. Quekel LG, et al. Detection of lung cancer on the chest radiograph: impact of previous films, clinical information, double reading, and dual reading. J Clin Epidemiol. 2001;54(11):1146–50.
92. Zhao Y, de Bock GH, Vliegenthart R, van Klaveren RJ, Wang Y, Bogoni L, et al. Performance of computer-aided detection of pulmonary nodules in low-dose CT: comparison with double reading by nodule volume. Eur Radiol. 2012;22(10):2076–84.
93. Fraioli F, Bertoletti L, Napoli A, Pediconi F, Calabrese FA, Masciangelo R, et al. Computer-aided detection (CAD) in lung cancer screening at chest MDCT: ROC analysis of CAD versus radiologist performance. J Thorac Imaging. 2007;22(3):241–6.
94. Beyer F, Zierott L, Fallenberg EM, Juergens KU, Stoeckel J, Heindel W, et al. Comparison of sensitivity and reading time for the use of computer-aided detection (CAD) of pulmonary nodules at MDCT as concurrent or second reader. Eur Radiol. 2007;17(11):2941–7.
95. Christe A, Leidolt L, Huber A, Steiger P, Szucs-Farkas Z, Roos JE, et al. Lung cancer screening with CT: evaluation of radiologists and different computer assisted detection software (CAD) as first and second readers for lung nodule detection at different dose levels. Eur J Radiol. 2013;82(12):e873–8.
96. Chan HP, Hadjiiski L, Samala RK. Computer-aided diagnosis in the era of deep learning. Med Phys (accepted), 2019.
97. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44.
98. 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.
99. Samala RK, Chan HP, Hadjiiski L, Helvie MA, Richter CD, Cha KH. Breast cancer diagnosis in digital breast tomosynthesis: effects of training sample size on multi-stage transfer learning using deep neural nets. IEEE Trans Med Imaging. 2019;38(3):686–96.
100. Gao C, et al. Augmenting LIDC dataset using 3D generative adversarial networks to improve lung nodule detection. Proc. SPIE 10950, *Medical Imaging* 2019: Computer-Aided Diagnosis, 109501K.2019https://doi.org/10.1117/12.2513011.
101. Petrick N, Sahiner B, Armato SG III, Bert A, Correale L, Delsanto S, et al. Evaluation of computer-aided detection and diagnosis systems. Med Phys. 2013;40(8):087001.
102. Huo Z, Summers RM, Paquerault S, Lo J, Hoffmeister J, Armato SG III, et al. Quality assurance and training procedures for computer-aided detection and diagnosis systems in clinical use. Med Phys. 2013;40(7):077001.