BJR

AI and Lung Cancer

Research papers from *BJR*

Contents

BJR

BJR is the international research journal of the British Institute of Radiology and is the oldest scientific journal in the field of radiology and related sciences.

An international, multi-disciplinary journal, *BJR* covers the clinical and technical aspects of medical imaging, radiotherapy, oncology, medical physics, radiobiology and the underpinning sciences. *BJR* is essential reading for radiologists, medical physicists, radiation oncologists, radiotherapists, radiographers and radiobiologists.

Aidence was founded in November 2015 by Mark-Jan Harte (CEO) and Jeroen van Duffelen (CBO). Based in the Netherlands and the UK, Aidence rallies over 60 data scientists, software engineers, medical, regulatory, and business professionals to provide intelligent software for the lung cancer pathway.

Aidence's first clinical application is Veye Lung Nodules, an AI-based solution for pulmonary nodule management on chest CTs. The solution is certified under the EU Medical Device Regulation and can be used as a second or concurrent reader. Veye Lung Nodules is currently running in routine practice and lung cancer screening across Europe, analysing thousands of patient scans each week.

Aidence is the preferred AI vendor for the NHSE Targeted Lung Health Checks. Most sites involved in the programme use Aidence's solutions to detect, assess the evolution of, and report on lung nodules.

Aidence is part of the AI division of RadNet, a US-based leading provider of diagnostic imaging services.
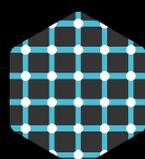
aidence
human sense in artificial intelligence

Do radiologists make better decisions with AI?
The INPACT study is set to find out

By David King, Project & Delivery Manager at Aidence

Project INPACT is go!, reads a recent tweet from clinical consultancy Hardian Health. We have officially launched novel research into a fascinating yet understudied area: the human-machine relationship. Project INPACT (Investigating Nodule Protocol Adherence using CADe/x Technology) will give us insights into the impact of an artificial intelligence (AI) medical solution on physicians' decision-making.

Most scientific literature on AI clinical applications focuses on their performance, for instance showing a device is safe, effective, and works as intended. INPACT, on the other hand, is one of the few studies examining AI in actual clinical practice: the reporting of chest CT scans in radiology departments across UK hospitals. The project is a collaboration between Aidence, Hardian Health, and the University of Edinburgh, with funding from the UK's National Health System (NHS) through the AI in Health and Care Award.

In this article, we explain the reasoning behind and the design of INPACT.

AI for early lung cancer detection

Early detection is the best chance of improving outcomes for lung cancer patients. There are two main opportunities to ensure it: targeted screening programmes inviting individuals at-risk for a low-dose chest scan, and the reporting and follow-up of indeterminate lung nodules found incidentally in clinical practice.

The INPACT study will focus on the latter. Most lung nodules detected in routine practice are benign, but some are cancerous. Their identification and follow-up are essential to an early lung cancer diagnosis, when a cure may still be possible. However, these tasks are challenging for physicians, requiring them to spot millimetric lesions with the naked eye, characterise, and measure or segment them (semi-)manually. For a workforce already under strain, it is additional, demanding work.

AI tools have the potential to identify incidental pulmonary nodules on chest CTs, classify them, and provide accurate measurements to assess growth. As part of the INPACT project, radiologists will report on these findings using AI-based solution Veye Lung Nodule, our CE certified second or concurrent reader, fully integrated into the hospital IT infrastructure. The aim is to investigate this solution's effect on physicians' follow-up decisions.

The reasoning behind INPACT

The Venn diagram below showcases the position of AI medical solutions in healthcare. On the left, clinical validation research is the topic of many publications aiming to demonstrate the performance of an application. For commercially available medical devices, it is also a regulatory requirement. (You can find out more about the validation study which confirmed Veye Lung Nodules' performance on this page.)

On the right, clinical outcomes represent the endpoints we are trying to influence by introducing AI devices in care pathways. This area is largely unexplored because AI technology is still in its early stages. We have not had the time or resources to complete extensive clinical trials, often spanning over many years, to assess if patients whose doctors use AI systems receive a more precise or timely diagnosis or care.

With the INPACT project, we zoom in on the middle: the real-world use of AI in clinical practice.
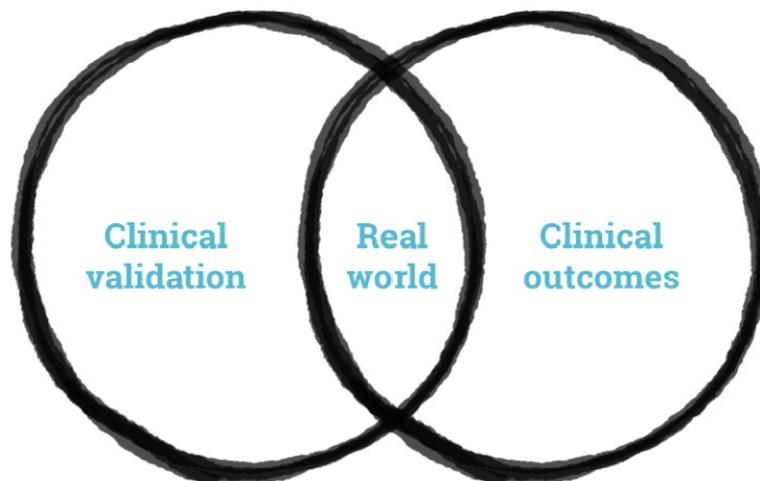
Our assumption is that radiologists using an AI-based solution for lung nodule management and reporting may come to a different follow-up recommendation than they would have without AI. This different follow-up recommendation may be more in line with what an experienced radiologist specialising in chest imaging would have given. It is possible that, in the absence of the supporting tool, radiologists would have missed small nodules or come up with inaccurate measurements.

Participating hospitals

The study will be performed in the radiology departments of six hospitals in the UK: NHS Lothian / Royal Infirmary of Edinburgh, University Hospital Southampton NHS Foundation Trust (FT), Royal Cornwall Hospitals NHS Trust, North West Anglia NHS FT, Liverpool Heart and Chest Hospital NHS FT, Royal United Hospitals Bath NHS FT.

Dr Rishi Ramaesh, consultant radiologist at NHS Lothian and principal investigator, explained his motivation for participating in the study:

> "The INPACT study is a highly novel and exciting evaluation that will demonstrate tangible and real world benefits of using artificial intelligence in clinical radiology practice. The study aims to show that radiologists using AI in their clinical practice are safer and more efficient, along with clear benefits to patients and healthcare systems.
>
> The results of the INPACT study are sure to strengthen the case for widespread adoption of AI tools into the NHS."

The study design

The sample of cases for this study will consist of up to 750 unique chest CT scans per centre, adding up to approximately 4,500 cases in total over six months. A primary consideration in the design of this study has been ensuring that radiologists spend the least amount of time possible on the study while collecting enough data to make it feasible. INPACT will be a prospective and comparative study, with the following design:

Aided and unaided readings

The radiologists at the participating centres will assess the eligible CT scans for lung nodules and provide their recommendations according to the British Thoracic Society (BTS) guidelines, as they would normally in routine clinical practice.

First, they will analyse the image unaided by Veye Lung Nodules and decide on the next steps. Then, they will provide a consecutive reading with access to Veye's results: detected nodules, their type

(solid or sub-solid), size (diameter and volume) and growth (growth percentage and volume doubling time from the prior scan, if available). Here is an example of a report generated by Veye Lung Nodules:

After the aided reading, the radiologist may adjust their initial follow-up or management recommendation and provide an explanation for the change and feedback on Veye's results. To streamline this step, we created a simple case report form which should only take a couple of minutes to fill in.

A radiology expert at each centre – i.e. a consultant with around ten years of experience in thoracic imaging – will independently evaluate all cases where one or more nodules were identified by either the radiologist or Veye. This expert may also consult with a pulmonologist on complicated cases. With readings aided by Veye Lung Nodules, it is likely to reach a higher agreement between the expert and the less specialised radiologists reading the scans.

We hypothesise that the AI-based automated detection, classification and quantification of lung nodules will reduce the influence of human measurement errors and allow radiologists to make a better recommendation for managing the nodules.

A mixed-method approach

Once collected, we will look at the results from two different perspectives:

- *Quantitative*

The radiologist's potential performance boost when using Veye Lung Nodules. We will gather this data by analysing the filled-in forms.

- *Qualitative*

The level of confidence that the radiologist has in understanding and accepting the outputs of Veye and using these outputs in their clinical decision-making. To obtain this data, researchers from the School of Social and Political Science at the University of Edinburgh will conduct interviews with participating radiologists before and after using Veye Lung Nodules.

Health economics

If Veye Lung Nodules has a beneficial impact on a radiologists' decision, we can follow this decision through the rest of the clinical pathway and into the patient outcome. This would allow us to pinpoint the benefit of using the device for early detection and possibly life-saving treatment.

For this purpose, our partners at Hardian Health will perform health economics and cost-effectiveness modelling. Cost-effectiveness analyses (CEA) are recognised as the gold standard in health economic evaluations. The National Institute for Health and Care Excellence (NICE), for example, is considering it for the upcoming Evidence Standards Framework for digital health technologies.

Using the outputs of the quantitative research, the researchers will extrapolate their impact using health economics modelling techniques. Dr Hugh Harvey, Managing Director at Hardian Health, emphasises the value of this part of the study design:

> *"Clinical impact is vitally important to demonstrate in order to gain trust from end-users, but health economic evaluations are equally as important in helping gain trust from payers, especially in state-funded and evidence-based healthcare systems such as the NHS."*

The below image is a summary of the INPACT approach:

Clinical cost-effectiveness
What is the potential gain in health outcomes and associated costs?

End user perceptions
What is the real world experience of working with an AI solution?

Patient & public views
Do patients and the public endorse the use of AI in radiology?

Hospital cost impact
Is the technical cost-incurring or cost-saving for an NHS site?

Clinical impact study
What difference does the software make to clinical decision making?

■ Clinical study   ■ Health economics   ■ Qualitative research

A blueprint for future research

The pace at which AI medical devices are brought to market exceeds the time and budget for clinical trials assessing their impact. Since the NHS is at the forefront of adopting new technologies, it needs more agile ways of evaluating AI solutions in the real world. NICE, in fact, is also consulting on a real-world evidence framework.

INPACT might serve as an example of such a framework. The design of the study and the lessons we will learn may be generalisable into a blueprint for assessing the application of AI in radiology. Its strength is that it does not require a large infrastructure or extended period. However, it will likely yield valuable results, allowing us to understand how AI makes a difference in patient outcomes.

The results of INPACT are planned for the end of 2022.

COMMENTARY

European lung cancer screening: valuable trial evidence for optimal practice implementation

[1]MARIO SILVA, MD, PhD, [1]GIANLUCA MILANESE, [1]ROBERTA E LEDDA, [2]SUNDEEP M NAYAK, [3]UGO PASTORINO and [1]NICOLA SVERZELLATI

[1]Scienze Radiologiche, Department of Medicine and Surgery (DiMeC), University of Parma, Parma, Italy
[2]Department of Radiology, Kaiser Permanente Northern California, San Leandro, California, USA
[3]Section of Thoracic Surgery, IRCCS Istituto Nazionale Tumori, Milano, Italy

Address correspondence to: Dr Mario Silva
E-mail: *mario.silva@unipr.it*

ABSTRACT

Lung cancer screening (LCS) by low-dose computed tomography is a strategy for secondary prevention of lung cancer. In the last two decades, LCS trials showed several options to practice secondary prevention in association with primary prevention, however, the translation from trial to practice is everything but simple. In 2020, the European Society of Radiology and European Respiratory Society published their joint statement paper on LCS. This commentary aims to provide the readership with detailed description about hurdles and potential solutions that could be encountered in the practice of LCS.

Lung cancer screening (LCS) trials have showed the potential of diagnosing early stage lung cancer (LC) with reduction of mortality in high risk populations. In 2011, the National Lung Screening Trial (NLST) demonstrated that CT outperformed chest radiography (CXR) in the detection of early stage LC.[1] In the following decade, several European trials reported analogous results, albeit with smaller populations and variable eligibility criteria and LCS algorithm.[2] To date, European LCS trials show that (i) high-risk profiling is critical,[3] (ii) volume measurement of lung nodule provides high yield,[4,5] (iii) prolonged LCS offers incremental efficacy,[6] (iv) biennial rounds may be considered after negative low-dose computed tomography (LDCT) to save costs and radiation burden,[4,5,7] and (v) females experience a higher reduction of LC mortality as compared to males.[8,9] These conclusions support more widespread promotion of LCS to European populations.

In 2020, the European Society of Radiology (ESR) and European Respiratory Society (ERS) issued a joint statement paper highlighting the pivotal requirements for LCS implementation in Europe.[10,11] This document expands former scientific and governmental recommendations[12–26] (Table 1), with the main objective of practice standardisation in different European countries, and across the various LCS stakeholders (*i.e.* candidate participant, general practitioner, specialized LCS centre).[10,27] In this commentary, we

discuss critical topics for efficient LCS: (i) engagement of the target population, (ii) optimized protocol for management of LDCT findings, and a (iii) reference standard for quality assurance (QA).

ENGAGEMENT OF THE TARGET POPULATION

The integration between LCS stakeholders is a substantive challenge for LCS initiation, as extensively documented in the United States since 2015.[28] The initial engagement of potential LCS participants is uniquely challenging.[29] LCS is the first major targeted large-scale screening programme: unlike breast or colon screening, the recruitment of LCS participants is highly predicated upon self-awareness of personal risk and willingness to participate over successive rounds.[30] The UK Lung Screen tackled the issue of LCS recruitment by testing mail letters: agreeable respondents were mostly below the threshold of established minimum risk of LC, whereas adherence was lacking among those at higher risk of LC.[31] Remarkably, current smokers and those of lower socioeconomic status have shown lower participation rates across both the EU and USA.[32,33] The perceived stigmatization of smoking-related risk might steer one away from LCS.[34] Quaife et al investigated the patients' views on LCS by interviewing a diverse cohort of current smokers, former smokers, and never smokers.[35] They reported that recruitment by general practitioners (GPs) recommendation was associated with the highest LCS intention. It was

Table 1.

	Eligibility criteria: age	Eligibility criteria: smoking history	LCS management: LDCT interval	LCS management: participants' selection - LCS design
US Preventive Services Task Force[16]	55–80	30 pack-year smoking history and currently smoke or have quit within the past 15 years	Annual	LCS should be discontinued once a person has not smoked for 15 years or develops a health problem that substantially limits life expectancy or the ability or willingness to have curative lung surgery.
American Cancer Society Recommendations for the Early Detection of Cancer in Average-Risk Asymptomatic Adults[17]	55–74	30 pack-year smoking history and currently smoke or have quit within the past 15 years	Annual	Eligible subjects have undergone a process of informed/shared decision making that included information about the potential benefits, limitations and harms of LDCT screening Having access to high-volume, high quality LCS and treatment center and - if current smokers - to smoking cessation counselling
National Comprehensive Cancer Network[18]	**Group 1**: 55–75 **Group 2**:≥50	**Group 1**:≥30 pack years, quit <within the past 14 years, current smoker **Group 2**:≥20 pack years, other risk factors (other than second-hand smoke)	Annual	Additional risk factors include cancer history, lung disease history, family history of LC, radon exposure, occupational exposure and history of COPD or pulmonary fibrosis
Centers for Medicare & Medicaid Services (CMS)[19]	55–77	≥30 pack years and either continue to smoke or have quit within the past 15 years	Annual	Asymptomatic subjects who receive written order for LDCT LCS
American College of Chest Physicians[20]	55–77	≥30 pack years and either continue to smoke or have quit within the past 15 years	Annual	Asymptomatic smokers and former smokers
NHS[21]	55–74			The Targeted Lung Health Check Programme will use the PLCOm2012 risk prediction model and the LLPver2 to select participants to be offered LCS PLCOm2012 risk of ≥1.51% over 6 years or LLPver2 5 year risk of ≥2.5%
Poland[22]	**Group 1**: 55–74 **Group 2**: 50–74	**Group 1**: 20 pack years, maximum 15 years since quitting **Group 2**: additional risk factor, 20 pack years, maximum 15 years since quitting	Annual	
Germany - Deutsche Krebsgesellschaft, Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (Association of the Scientific Medical Societies in Germany)[23]	**Group 1**: 55–74 **Group 2**: 50–74	**Group 1**:≥30 pack years and either continue to smoke or have quit within the past 15 years **Group 2**:≥20 pack years and other risk factors	Annual	Associated with smoking cessation counselling, for a duration of at least 2 years
Austria - Österreichischen Röntgengesellschaft und der Österreichischen Gesellschaft für Pneumologie (Austrian Society of Roentgenology and Austrian Pulmonological Society)[24]	>55	≥30 pack years and either continue to smoke or have quit within the past 15 years	Annual	
Canadian Task Force on Preventive Health Care[25]	55–74	≥30 pack years and either continue to smoke or have quit within the past 15 years	Annual	LCS - up to three consecutive years - should only be done in health care settings with access to expertise in early diagnosis and treatment of lung cancer

(Continued)

Table 1. (Continued)

	Eligibility criteria: age	Eligibility criteria: smoking history	LCS management: LDCT interval	LCS management: participants' selection - LCS design
Japan Radiological Society[26]	≥50	Brinkman Index ≥ 600		Brinkman Index is calculated as the number of cigarettes smoked per day multiplied by the number of years of smoking

COPD, chronic obstructive pulmonary disease; LC, lung cancer; LCS, lung cancer screening; LDCT, low-dose computed tomography.

also observed that smokers had little confidence in early diagnosis for improvement of quality of life and in surgical resection of early-stage LC, despite they were aware of the risk of LC and its poor prognosis when diagnosed at symptomatic stage.[35] Optimized recruitment strategy by active involvement of GP might improve willingness of high risk subjects to agree to multiple rounds of LCS.[36] This approach would tighten connections throughout the healthcare network, potentially improving the recruitment process by limiting the risk of stigma attached to tobacco and lung cancer.[35] The involvement of GP must be seen as a golden opportunity to introduce personalized secondary and primary prevention, as it also includes actively encouraging smoking cessation.[37] The Lung Screen Uptake Trial (LSUT) showed an excellent uptake above 50% by a targeted stepped invitation by letters from primary care practice.[38] Of note, the LSUT also reported that social disparities can be smoothen by targeting psychological barriers with low-burden information.[38] Furthermore, beyond smoking habit, the evolving epidemiology of LC demands continuous adaptation of inclusion criteria (and risk models) to emerging risk factors such as air pollution, former history of cancer, and genetics.[39–41]

Geographical accessibility to CT equipment is another important roadblock[27,34] and was initially acknowledged in the UK, where mobile CT scanners were deployed to obtain equity in access in remote areas.[31] However, this solution has not proven to confer any substantive advantage for engagement of screenees in both the EU and USA, and applications may differ between densely populated urban centers and more rural areas.[42,43] Moreover, recent evidence demonstrated that LCS adherence is higher among participants screened with a centralized approach as compared to those screened within a decentralized program.[44]

Complex psychosocial effects of multiple rounds of screening must be acknowledged and addressed pre-emptively. Anxiety might be generated by radiological reporting of abnormalities, hence the lexicon in LDCT reporting should be adapted to minimize the psychological impact. One example comes from the definition of positive nodules: the NLST reported "positive" finding for any case with solid nodule >4 mm (about 40% of the population in three rounds), whereas more recently the European literature sustained milder and more nuanced terminology such as "indeterminate" finding for solid nodules below 300 mm[3] or 8 mm in size.[15,45] The latter approach is targeted to reduce the psychological burden in a large proportion of screening participants for whom detection of intermediate size nodule is expected and will likely represent a benign finding. Brain et

al reported anxiety for participants with abnormal screening results requiring repeat imaging, with progressive reduction in self-reported anxiety after 2 years of follow-up.[46] Other authors reported no evidence of distress beyond 6 months.[47] Despite general agreement on the psychological distress from LCS, Kummer et al reported that the overall increase in anxiety is unlikely to represent clinically relevant difference.[48]

Engagement and adherence to LCS program might be favored in subgroups of population already undergoing regular screening rounds (*e.g.* mammography). LCS trials reported a more favorable outcome in females compared to males,[8,9] probably because of the relatively higher representation of slow-growing adenocarcinoma in females.[8] LCS screening in female gender will assume greater importance with the reported increase in tobacco use among females.[49] Such habit anticipates a long perspective of increased LC risk, which is expected to fluctuate with a 30–40 years delay since change in population smoking habit.[50]

Opportunistic screening is undergoing screening outside of the screening environment that might occur when a LCS program does not exist or, otherwise, when LCS program exists but information about its applicability is lacking.[51,52] It is characterized by high rate of inappropriate referral to CT (or even worse to radiography) and by high risk of inadequate management of radiological findings.[51] Thus, one further reason to promote LCS practice is for tackling the growing phenomenon of voluntary opportunistic screening. The variable response between screening reluctance and opportunistic conduct warrants improving trust relationships with participants by a clear explanation and understanding of both advantages and potential harms of LCS. Major topics for discussion with LCS candidates include the possibility of false positive (invasive procedures/surgery for non-malignant findings), interval cancers, and psychological impact.[27,34,53] For these reasons, personalized stratification of risk should be prioritized and pursued by proactive collaboration between peripheral healthcare professionals (*e.g.* GP) and specialized personnel operating in LCS facilities (*e.g.* specialized nurse, radiologist, pulmonologist, surgeon).[54–56] Refined information strategies for both participants and peripheral facilities might indeed represent one reason of the recent growth in LCS uptake in the USA.[43,57]

OPTIMIZED PROTOCOL FOR MANAGEMENT OF LDCT FINDINGS

As LCS is performed by LDCT,[1,58] the spatial resolution of this imaging technique provides high anatomical detail and leads to the detection of a wide range of findings. The management of any

finding should be weighted on the individual risk that takes into consideration personal comorbidities and survival rates. This is a main determinant of the risk-benefit balance in LCS, which includes elderly smokers with comorbidities and, therefore, requires evidence-based protocols to mitigate overinvestigation, overdiagnosis, overtreatment, and associated psychological impact, costs, and risks.

Each pulmonary nodule on LDCT is individually assessed for its cancer risk by using density, size, and further morphological descriptors.[59,60] These features subsequently drive the protocols for lung nodule management in major guidelines.[13,15,61] Enrolment or baseline (initial) LCS round shows a higher frequency of indeterminate findings compared with the following rounds.[62,63] The high prevalence of indeterminate nodules may be addressed with the Brock model, a multifactorial score for cancer risk on baseline LDCT.[64] The Brock model underscores the high likelihood of cancer in subsolid nodules, however, the risk of overdiagnosis and overtreatment is high in this type of nodules.[65] Subsolid nodule represents a phenotype of adenocarcinoma with slow growth and relatively low risk of distant disease.[66] This scenario introduces the major concern related to LC overtreatment and its associated risks, notably in elderly smokers with comorbidities. First, most overinvestigation and overtreatment is expected from nodules with intermediate size (about 10–20% at baseline round) and/or subsolid in density.[56] Second, most nodules with these features are extremely unlikely to evolve in stage shift within the following 6–12 months.[67] Most guidelines suggest 3 month follow up for indeterminate lung nodules, whereas the Lung Reporting And Data System (LungRADS) recommends 6 month follow-up (category 3). The LungRADS approach might be regarded as pragmatic adaptation of follow-up interval to warrant an optimized time range for evolution of the finding and more conclusive longitudinal measurement, compared with 3 month follow-up.[61] Furthermore, the literature showed that a conservative approach by annual LDCT until signs of growth is a

safe strategy for controlling work-up and intervention of subsolid nodules.[68–70] Appropriate and rigorous application of evidence-based management of nodules is key for optimized risk–benefit ratio of LCS. The extended analysis of NLST beyond 10 years showed similar frequency of LC between LDCT and CXR arm, suggesting that early diagnosis by CT does not imply overdiagnosis.[71] One large randomized controlled trial, the SUMMIT study (50,000 participants; ClinicalTrials.gov NCT03934866), was designed with parsimonious use of LDCT information (both for nodule and incidental findings)[72] and their conclusions should be ready within a decade.

Compared to nodule management, the handling of incidental LDCT finding seems to be more difficult. The potential costs of frequent incidental LDCT findings in the NLST were a matter of concern for both Medicare Evidence Development and Coverage Advisory Committee and the American College of Physicians.[73] The NELSON group reported neglectable benefit from seeking incidental findings on LDCT.[74] However, incidental findings and comorbidities may represent extrapulmonary malignancies[75] or non-oncologic actionable findings: potentially significant abnormalities were reported in almost 20% of screenees.[76] Gareen et al demonstrated that the medical costs were similar across the two arms of the NLST, despite the higher rates of incidental findings detected by LDCT as compared to CXR.[77] The management of incidental findings still represents a gray area of LCS. This matter should be approached with extreme caution when considering deployment of LCS at the population level: excess reporting might both increase cost–risks of LCS and discourage participants, who feel overwhelmed by the burden of information. The frequency of incidental findings might approach 100% in selected high-risk population (*e.g.* coronary calcifications, signs of emphysema, etc.), and may depend on the experience of the reading radiologist. Therefore, clinically oriented reporting of incidental findings is a fundamental of LCS with LDCT. Incidental LDCT findings with moderate risk of cancer

Figure 1. A. LDCT of a 73-year-old female patient, showing a 28-mm solid nodule (white arrow) located in the right upper lobe. The nodule is attached to vessels and was missed by a computer-aided detection system: false negative of CADe. B. Same patient of A, Maximum Intensity Projection image. CADe, computer-assisted detection; LDCT, low-dose computed tomography
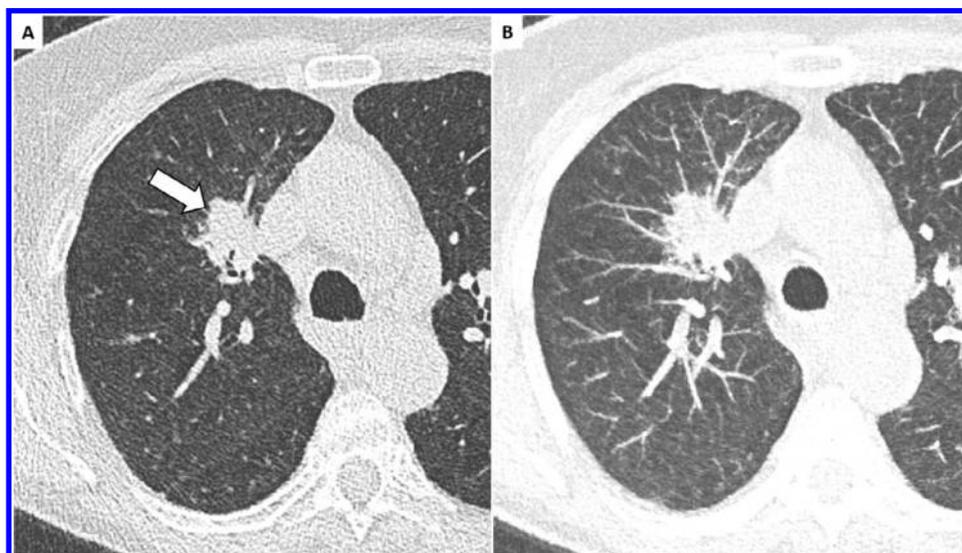
Figure 2. LDCT images showing a solid nodule abutting the pleura (A, manual caliper; B, semi-automated volume segmentation) compared to a solid nodule without solid interface (C, manual caliper; D, semi-automated volume segmentation). Despite same average diameter (6 mm), the volumetry resulted in substantially different values due to the inaccurate volumetric segmentation of the subpleural nodule (B), which partially included the pleura. Manual measurement is suggested in case of nodules with at least 1/3 solid interface. LDCT, low-dose computed tomography
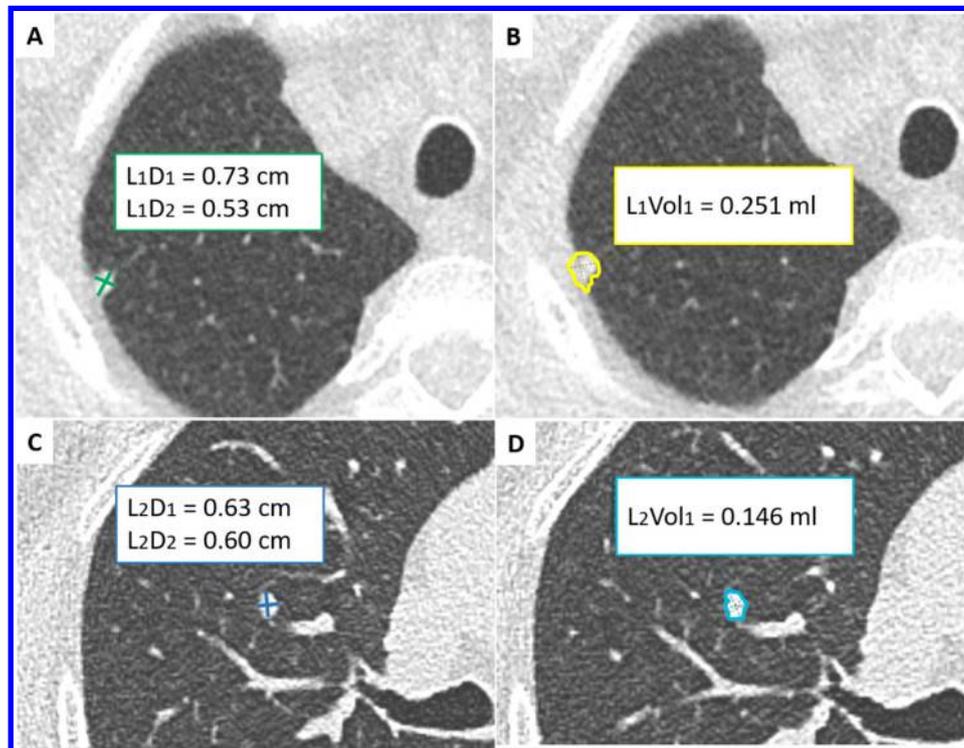
might deserve work-up or, if possible, follow-up during annual surveillance. The current practical experience of most LCS multidisciplinary teams encourages a standardized management for incidentaloma, similar to that in the general population, as advocated by several white papers from the American College of Radiology[78] and the recommendations from the National Health System England.[21]

REFERENCE STANDARD FOR QUALITY ASSURANCE

QA is backbone to all medical procedures, especially for those practices that involve heterogeneous level of skills and expertise. For this reason, several scientific societies and public authorities put substantial effort to provide guidance for standardized QA in LCS. In the USA, the American Cancer Society proposed standard indicators of QA with major emphasis on compliance to follow-up recommendations.[79] In Europe, the European Society of Thoracic Imaging promotes educational initiatives to ensure uniform interpretation and limit the variation in responding to the detected findings and their management by using predefined protocols.[80] Semi-automated software is encouraged for nodule detection and measurement, in alignment with both interim analysis and results of the NELSON trial.[9,81–83] The use of computer-aided detection (CADe) for LCS meets two major needs: (a) resource optimization by reduction of reading time,[84,85] (b) a guarantee of measurement reproducibility.[86]

CADe systems are typically set for detection of "nodule candidates" within a limited size range, to minimize the rate of false candidates: for this reason, CADe systems are extremely sensitive for nodules <10 mm, but also prone to overlooking larger nodules. Therefore, false negatives for large nodules should be expected while using CADe (Figure 1). Large findings, however, are easily detected by visual reading, especially by maximum-intensity projection (MIP) reconstructions.[87–89] The most accurate reading of LDCT is granted by coupled CADe and MIP read-out.

The use of semi-automated nodule volumetry is associated with variability in nodule volume.[90,91] On the other hand, manual electronic caliper tolerance is reported 1.5–2 mm, reflecting a significant volume variability (even above 100%) in small-intermediate nodule (diameter 6–8 mm).[61,92] Therefore, semi-automated volumetry is strongly recommended to minimize the number of positive or indeterminate tests. Furthermore, this is an excellent mean for fast and accurate calculation of longitudinal dynamic metrics, such as the volume doubling time.[82] Nonetheless, pitfalls of semi-automated volumetry must be acknowledged to avoid excessive confidence. Major concerns of semi-automated volumetry are represented by variability between software vendors (also between different versions of the same software) and high variability in case of nodules attached to solid structures (Figure 2).[93–95] Finally, beyond "technological" parameters, clinical and morphological characterization

are still of utmost importance for optimal management of relatively large nodules with clearly benign characteristics, such as fissure attached lymph nodes and benign calcifications.[96–98] Potential future perspective for the characterization of the biological behavior of pulmonary nodules are seen in "advanced quantitative" descriptors.[99–103] Artificial intelligence (AI) with "multiomic" approach (*i.e.* by integrating imaging, functional and biological parameters) is now under development with the aim of further stratifying the risk of LC.[100,103] Nonetheless, for AI to be fully implemented within the forthcoming LCS setting, future studies are needed to test and confirm its consistency and accuracy in lung nodule characterization. Integration of imaging and circulating biomarkers (*e.g.* plasma microRNA) proposed

for further refinement of individualized biological characterization,[104] will likely be object of debate in coming years. For the best integration of the multifaceted aspects of LCS, dedicated facilities and skilled human resources (including trained radiologists) are critical to limit the risk of excessive work-up.[105]

In conclusion, LCS as part of our population health management is advancing in several countries and it is expected to improve LC survival. The success of LCS will depend on optimal engagement of target population, optimized use of resources, and continuous development of procedures at local, national, and international levels.

REFERENCES

1. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 4, 2011; **365**: 395–409. https://doi.org/10.1056/NEJMoa1102873

2. Silva M, Pastorino U, Sverzellati N. Lung cancer screening with low-dose ct in europe: strength and weakness of diverse independent screening trials. *Clin Radiol* 2017; **72**: 389–400. https://doi.org/10.1016/j.crad.2016.12.021

3. Wille MMW, Dirksen A, Ashraf H, Saghir Z, Bach KS, Brodersen J, et al. Results of the randomized danish lung cancer screening trial with focus on high-risk profiling. *Am J Respir Crit Care Med* 1, 2016; **193**: 542–51. https://doi.org/10.1164/rccm.201505-1040OC

4. Sverzellati N, Silva M, Calareso G, Galeone C, Marchianò A, Sestini S, et al. Low-dose computed tomography for lung cancer screening: comparison of performance between annual and biennial screen. *Eur Radiol* 2016; **26**: 3821–29. https://doi.org/10.1007/s00330-016-4228-3

5. Pastorino U, Sverzellati N, Sestini S, Silva M, Sabia F, Boeri M, et al. Ten-year results of the multicentric italian lung detection trial demonstrate the safety and efficacy of biennial lung cancer screening. *Eur J Cancer* 2019; **118**: 142–48. https://doi.org/10.1016/j.ejca.2019.06.009

6. Pastorino U, Silva M, Sestini S, Sabia F, Boeri M, Cantarutti A, et al. Prolonged lung cancer screening reduced 10-year mortality in the mild trial. *Annals of Oncology: Official Journal of the European Society for Medical Oncology / ESMO* 2019; **30**: 1162–69. https://doi.org/10.1093/annonc/mdz117

7. Yousaf-Khan U, van der Aalst C, de Jong PA, Heuvelmans M, Scholten E, Lammers J-W, et al. Final screening round of the nelson lung cancer screening trial: the effect of a 2.5-year screening interval. *Thorax* 2017; **72**: 48–56. https://doi.org/10.1136/thoraxjnl-2016-208655

8. Becker N, Motsch E, Trotter A, Heussel CP, Dienemann H, Schnabel PA, et al. Lung cancer mortality reduction by ldct screening-results from the randomized german lusi trial. *Int J Cancer* 15, 2020; **146**: 1503–13. https://doi.org/10.1002/ijc.32486

9. de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, et al. Reduced lung-cancer mortality with volume ct screening in a randomized trial. *N Engl J Med* 6, 2020; **382**: 503–13. https://doi.org/10.1056/NEJMoa1911793

10. Kauczor H-U, Baird A-M, Blum TG, Bonomo L, Bostantzoglou C, Burghuber O, et al. ESR/ers statement paper on lung cancer screening. *Eur Respir J* 2020; **55**(2): 1900506. https://doi.org/10.1183/13993003.00506-2019

11. Kauczor H-U, Baird A-M, Blum TG, Bonomo L, Bostantzoglou C, Burghuber O, et al. ESR/ers statement paper on lung cancer screening. *Eur Radiol* 2020; **30**: 3277–94. https://doi.org/10.1007/s00330-020-06727-7

12. Jaklitsch MT, Jacobson FL, Austin JHM, Field JK, Jett JR, Keshavjee S, et al. The american association for thoracic surgery guidelines for lung cancer screening using low-dose computed tomography scans for lung cancer survivors and other high-risk groups. *J Thorac Cardiovasc Surg* 2012; **144**: 33–38. https://doi.org/10.1016/j.jtcvs.2012.05.060

13. Callister MEJ, Baldwin DR, Akram AR, Barnard S, Cane P, Draffan J, et al. British thoracic society guidelines for the investigation and management of pulmonary nodules. *Thorax* 2015; **70 Suppl 2**: ii1-54. https://doi.org/10.1136/thoraxjnl-2015-207168

14. Kauczor H-U, Bonomo L, Gaga M, Nackaerts K, Peled N, Prokop M, et al. ESR/ers white paper on lung cancer screening. *Eur Radiol* 2015; **25**: 2519–31. https://doi.org/10.1007/s00330-015-3697-0

15. . Oudkerk M, Devaraj A, Vliegenthart R, Henzler T, Prosch H, Heussel CP, et al. European position statement on lung cancer screening. Lancet Oncol. 2017;18(12):e754-e66.

16. US Preventive Service Task Force. Lung Cancer: Screening. 2020. Available from: www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/lung-cancer-screening

17. Smith RA, Andrews KS, Brooks D, Fedewa SA, Manassaram-Baptiste D, Saslow D, et al. Cancer screening in the united states, 2018: a review of current american cancer society guidelines and current issues in cancer screening. *CA Cancer J Clin* 2018; **68**: 297–316. https://doi.org/10.3322/caac.21446

18. National Comprehensive Cancer Network. National Comprehensive Cancer Network Guidelines, Version 1.2020, Lung Cancer Screening. 2020. Available from: https://www.nccn.org/professionals/physician_gls/pdf/lung_screening.pdf [accessed 13 Oct 2019].

19. Centers for Medicare & Medicaid Services. Decision Memo for Screening for Lung Cancer with Low Dose Computed Tomography (LDCT). 2015. Available from: https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=274

20. Mazzone PJ, Silvestri GA, Patel S, Kanne JP, Kinsinger LS, Wiener RS, et al. Screening

for lung cancer. *Chest* 2018; **153**: 954–85. https://doi.org/10.1016/j.chest.2018.01.016

21. NHS England - National Cancer Programme. Targeted Screening for Lung Cancer with Low Radiation Dose Computed Tomography. 2019. Available from: https://www.england.nhs.uk/publication/targeted-screening-for-lung-cancer/ [accessed 4 Feb 2020].

22. Rzyman W, Szurowska E, Adamek M. Implementation of lung cancer screening at the national level: polish example. *Transl Lung Cancer Res* 2019; **8**: S95-105. https://doi.org/10.21037/tlcr.2019.03.09

23. Leitlinien Programm Onkologie. S3-Leitline Prävention, Diagnostik, Therapie und Nachsorge des Lungenkarzinoms. 2018. Available from: https://www.awmf.org/uploads/tx_szleitlinien/020-007OL_l_S3_Lungenkarzinom_2018-03.pdf

24. Prosch H, Studnicka M, Eisenhuber E, Olschewski H, Stiefsohn E, Hartl S, et al. Opinion of the austrian society of radiology and the austrian society of pneumology. *Wien Klin Wochenschr* 2013; **125**: 339–45. https://doi.org/10.1007/s00508-013-0356-9

25. Canadian Task Force on Preventive Health Care. Recommendations on screening for lung cancer. *CMAJ* 2016; **188**: 425–32. https://doi.org/10.1503/cmaj.151421

26. Japan Radiological Society. Lung Cancer Screening Guidelines from Japan Radiological Society (JRS) and Japanese College of Radiology (JCR. 2013. Available from: http://www.radiology.jp/content/files/diagnostic_imaging_guidelines_2013_e.pdf

27. Ruparel M, Quaife S, Baldwin D, Waller J, Janes S. Defining the information needs of lung cancer screening participants: a qualitative study. *BMJ Open Respir Res* 2019; **6**(1): e000448. https://doi.org/10.1136/bmjresp-2019-000448

28. Quaife SL, Marlow LAV, McEwen A, Janes SM, Wardle J. Attitudes towards lung cancer screening in socioeconomically deprived and heavy smoking communities: informing screening communication. *Health Expect* 2017; **20**: 563–73. https://doi.org/10.1111/hex.12481

29. Doubeni CA, Wilkinson JM, Korsen N, Midthun DE. Lung cancer screening guidelines implementation in primary care: a call to action. *Ann Fam Med* 2020; **18**: 196–201. https://doi.org/10.1370/afm.2541

30. Baldwin DR, Callister MEJ. An update on ct screening for lung cancer: the first major targeted cancer screening programme. *Br J Radiol* 1, 2020; **93**(1116): 20200636. https://doi.org/10.1259/bjr.20200636

31. Crosbie PA, Balata H, Evison M, Atack M, Bayliss-Brideaux V, Colligan D, et al. Implementing lung cancer screening: baseline results from a community-based "lung health check" pilot in deprived areas of manchester. *Thorax* 2019; **74**: 405–9. https://doi.org/10.1136/thoraxjnl-2017-211377

32. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Clapp JD, Clingan KL, et al. Baseline characteristics of participants in the randomized national lung screening trial. *J Natl Cancer Inst* 1, 2010; **102**: 1771–79. https://doi.org/10.1093/jnci/djq434

33. Hestbech MS, Siersma V, Dirksen A, Pedersen JH, Brodersen J. Participation bias in a randomised trial of screening for lung cancer. *Lung Cancer* 2011; **73**: 325–31. https://doi.org/10.1016/j.lungcan.2010.12.018

34. Quaife SL, Janes SM. Lung cancer screening: improving understanding of the psychological impact. *Thorax* 2016; **71**: 971–72. https://doi.org/10.1136/thoraxjnl-2016-208966

35. Quaife SL, Vrinten C, Ruparel M, Janes SM, Beeken RJ, Waller J, et al. Smokers' interest in a lung cancer screening programme: a national survey in england. *BMC Cancer* 2018; **18**(1): 497. https://doi.org/10.1186/s12885-018-4430-6

36. Ruparel M, Quaife SL, Dickson JL, Horst C, Tisi S, Hall H, et al. Lung screen uptake trial: results from a single lung cancer screening round. *Thorax* 2020; **75**: 908–12. https://doi.org/10.1136/thoraxjnl-2020-214703

37. Jiménez-Ruiz CA, Andreas S, Lewis KE, Tonnesen P, van Schayck CP, Hajek P, et al. Statement on smoking cessation in copd and other pulmonary diseases and in smokers with comorbidities who find it difficult to quit. *Eur Respir J* 2015; **46**: 61–79. https://doi.org/10.1183/09031936.00092614

38. Quaife SL, Ruparel M, Dickson JL, Beeken RJ, McEwen A, Baldwin DR, et al. Lung screen uptake trial (lsut): randomized controlled clinical trial testing targeted invitation materials. *Am J Respir Crit Care Med* 15, 2020; **201**: 965–75. https://doi.org/10.1164/rccm.201905-0946OC

39. Su S-Y, Liaw Y-P, Jhuang J-R, Hsu S-Y, Chiang C-J, Yang Y-W, et al. Associations between ambient air pollution and cancer incidence in taiwan: an ecological study of geographical variations. *BMC Public Health* 9, 2019; **19**(1): 1496. https://doi.org/10.1186/s12889-019-7849-z

40. Lebrett MB, Crosbie EJ, Smith MJ, Woodward ER, Evans DG, Crosbie PAJ. Targeting lung cancer screening to individuals at greatest risk: the role of genetic factors. *J Med Genet* 2021; **58**: 217–26. https://doi.org/10.1136/jmedgenet-2020-107399

41. Henderson LM, Durham DD, Tammemägi MC, Benefield T, Marsh MW, Rivera MP. Lung cancer screening with low dose computed tomography in patients with and without prior history of cancer in the national lung screening trial. *J Thorac Oncol* 2021; **16**: 980–89. https://doi.org/10.1016/j.jtho.2021.02.003

42. Bartlett EC, Kemp SV, Ridge CA, Desai SR, Mirsadraee S, Morjaria JB, et al. Baseline results of the west london lung cancer screening pilot study - impact of mobile scanners and dual risk model utilisation. *Lung Cancer* October 2020; **148**: 12–19. https://doi.org/10.1016/j.lungcan.2020.07.027

43. Wang GX, Baggett TP, Pandharipande PV, Park ER, Percac-Lima S, Shepard J-A, et al. Barriers to lung cancer screening engagement from the patient and provider perspective. *Radiology* 2019; **290**: 278–87. https://doi.org/10.1148/radiol.2018180212

44. Smith HB, Ward R, Frazier C, Angotti J, Tanner NT. Guideline-recommended lung cancer screening adherence is superior with a centralized approach. *Chest* 15, 2021. https://doi.org/10.1016/j.chest.2021.09.002

45. Baldwin DR, Duffy SW, Wald NJ, Page R, Hansell DM, Field JK. UK lung screen (ukls) nodule management protocol: modelling of a single screen randomised controlled trial of low-dose ct screening for lung cancer. *Thorax* 2011; **66**: 308–13. https://doi.org/10.1136/thx.2010.152066

46. Brain K, Lifford KJ, Carter B, Burke O, McRonald F, Devaraj A, et al. Long-term psychosocial outcomes of low-dose ct screening: results of the uk lung cancer screening randomised controlled trial. *Thorax* 2016; **71**: 996–1005. https://doi.org/10.1136/thoraxjnl-2016-208283

47. Slatore CG, Sullivan DR, Pappas M, Humphrey LL. Patient-centered outcomes among lung cancer screening recipients with computed tomography: a systematic review. *J Thorac Oncol* 2014; **9**: 927–34. https://doi.org/10.1097/JTO.0000000000000210

48. Kummer S, Waller J, Ruparel M, Duffy SW, Janes SM, Quaife SL. Psychological outcomes of low-dose ct lung cancer screening in a multisite demonstration screening pilot: the lung screen uptake trial (lsut). *Thorax* 2020; **75**: 1065–73. https://doi.org/10.1136/thoraxjnl-2020-215054

49. Corsi DJ, Lear SA, Chow CK, Subramanian SV, Boyle MH, Teo KK. Socioeconomic and geographic patterning of smoking behaviour in canada: a cross-sectional multilevel analysis. *PLoS One* 2013; **8**(2): e57646.

https://doi.org/10.1371/journal.pone.0057646

50. Ridge CA, McErlean AM, Ginsberg MS. Epidemiology of lung cancer. *Semin Intervent Radiol* 2013; **30**: 93–98. https://doi.org/10.1055/s-0033-1342949

51. Richards TB, Doria-Rose VP, Soman A, Klabunde CN, Caraballo RS, Gray SC, et al. Lung cancer screening inconsistent with u.s. *Preventive Services Task Force Recommendations Am J Prev Med* 2019; **56**: 66–73. https://doi.org/10.1016/j.amepre.2018.07.030

52. Wang GX, Neil JM, Fintelmann FJ, Little BP, Narayan AK, Flores EJ. Guideline-discordant lung cancer screening: emerging demand and provided indications. *J Am Coll Radiol* 2021; **18**: 395–405. https://doi.org/10.1016/j.jacr.2020.08.005

53. Bartlett EC, Silva M, Callister ME, Devaraj A. False-negative results in lung cancer screening-evidence and controversies. *J Thorac Oncol* 2021; **16**: 912–21. https://doi.org/10.1016/j.jtho.2021.01.1607

54. Mazzone PJ, Tenenbaum A, Seeley M, Petersen H, Lyon C, Han X, et al. Impact of a lung cancer screening counseling and shared decision-making visit. *Chest* 2017; **151**: 572–78. https://doi.org/10.1016/j.chest.2016.10.027

55. Cheung LC, Berg CD, Castle PE, Katki HA, Chaturvedi AK. Life-gained-based versus risk-based selection of smokers for lung cancer screening. *Ann Intern Med* 2019; **171**: 623–32. https://doi.org/10.7326/M19-1263

56. Sands J, Tammemägi MC, Couraud S, Baldwin DR, Borondy-Kitts A, Yankelevitz D, et al. Lung screening benefits and challenges: a review of the data and outline for implementation. *J Thorac Oncol* 2021; **16**: 37–53. https://doi.org/10.1016/j.jtho.2020.10.127

57. Liu B, Dharmarajan K, Henschke CI, Taioli E. State-level variations in the utilization of lung cancer screening among medicare fee-for-service beneficiaries: an analysis of the 2015 to 2017 physician and other supplier data. *Chest* 2020; **157**: 1012–20. https://doi.org/10.1016/j.chest.2019.11.005

58. Nawa T, Fukui K, Nakayama T, Sagawa M, Nakagawa T, Ichimura H, et al. A population-based cohort study to evaluate the effectiveness of lung cancer screening using low-dose ct in hitachi city, japan. *Jpn J Clin Oncol* 1, 2019; **49**: 130–36. https://doi.org/10.1093/jjco/hyy185

59. Snoeckx A, Reyntiens P, Desbuquoit D, Spinhoven MJ, Van Schil PE, van Meerbeeck JP, et al. Evaluation of the solitary pulmonary nodule: size matters, but do not ignore the power of morphology. *Insights Imaging* 2018; **9**: 73–86. https://doi.org/10.1007/s13244-017-0581-2

60. Han D, Heuvelmans MA, van der Aalst CM, van Smoorenburg LH, Dorrius MD, Rook M, et al. New fissure-attached nodules in lung cancer screening: a brief report from the nelson study. *J Thorac Oncol* January 2020; **15**: 125–29. https://doi.org/10.1016/j.jtho.2019.09.193

61. ACoR. Lung-screening reporting and data system (lungrads). 2019. Available from: https://www.acr.org/-/media/ACR/Files/RADS/Lung-RADS/LungRADSAssessmentCategoriesv1-1.pdf?la=en

62. Walter JE, Heuvelmans MA, de Jong PA, Vliegenthart R, van Ooijen PMA, Peters RB, et al. Occurrence and lung cancer probability of new solid nodules at incidence screening with low-dose ct: analysis of data from the randomised, controlled nelson trial. *Lancet Oncol* 2016; **17**: 907–16. https://doi.org/10.1016/S1470-2045(16)30069-9

63. Walter JE, Heuvelmans MA, Bock G de, Yousaf-Khan U, Groen HJM, Aalst C van der, et al. Characteristics of new solid nodules detected in incidence screening rounds of low-dose ct lung cancer screening: the nelson study. *Thorax* 2018; **73**: 741–47. https://doi.org/10.1136/thoraxjnl-2017-211376

64. McWilliams A, Tammemagi MC, Mayo JR, Roberts H, Liu G, Soghrati K, et al. Probability of cancer in pulmonary nodules detected on first screening ct. *N Engl J Med* 5, 2013; **369**: 910–19. https://doi.org/10.1056/NEJMoa1214726

65. Yip R, Li K, Liu L, Xu D, Tam K, Yankelevitz DF, et al. Controversies on lung cancers manifesting as part-solid nodules. *Eur Radiol* 2018; **28**: 747–59. https://doi.org/10.1007/s00330-017-4975-9

66. Yip R, Yankelevitz DF, Hu M, Li K, Xu DM, Jirapatnakul A, et al. Lung cancer deaths in the national lung screening trial attributed to nonsolid nodules. *Radiology* 2016; **281**: 589–96. https://doi.org/10.1148/radiol.2016152333

67. Silva M, Milanese G, Sestini S, Sabia F, Jacobs C, van Ginneken B, et al. Lung cancer screening by nodule volume in lung-rads v1.1: negative baseline ct yields potential for increased screening interval. *Eur Radiol* 2021; **31**: 1956–68. https://doi.org/10.1007/s00330-020-07275-w

68. Silva M, Prokop M, Jacobs C, Capretti G, Sverzellati N, Ciompi F, et al. Long-term active surveillance of screening detected subsolid nodules is a safe strategy to reduce overtreatment: official publication of the international association for the study of lung cancer. *J Thorac Oncol* 2018; **13**: 1454–63. https://doi.org/10.1016/j.jtho.2018.06.013

69. Scholten ET, de Jong PA, de Hoop B, van Klaveren R, van Amelsvoort-van de Vorst S, Oudkerk M, et al. Towards a close computed tomography monitoring approach for screen detected subsolid pulmonary nodules? *Eur Respir J* 2015; **45**: 765–73. https://doi.org/10.1183/09031936.00005914

70. Mets OM, Chung K, Zanen P, Scholten ET, Veldhuis WB, van Ginneken B, et al. *in vivo* growth of 60 non-screening detected lung cancers: a computed tomography study. *Eur Respir J* 2018; **51**(4): 1702183. https://doi.org/10.1183/13993003.02183-2017

71. National Lung Screening Trial Research T. Lung cancer incidence and mortality with extended follow-up in the national lung screening trial. *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer* 2019; **14**: 1732–42.

72. Horst C, Dickson JL, Tisi S, Ruparel M, Nair A, Devaraj A, et al. Delivering low-dose ct screening for lung cancer: a pragmatic approach. *Thorax* 2020; **75**: 831–32. https://doi.org/10.1136/thoraxjnl-2020-215131

73. Owens DK, Qaseem A, Chou R, Shekelle P, Clinical Guidelines Committee of the American College of Physicians. High-value, cost-conscious health care: concepts for clinicians to evaluate the benefits, harms, and costs of medical interventions. *Ann Intern Med* 2011; **154**: 174–80. https://doi.org/10.7326/0003-4819-154-3-201102010-00007

74. van de Wiel JCM, Wang Y, Xu DM, van der Zaag-Loonen HJ, van der Jagt EJ, van Klaveren RJ, et al. Neglectable benefit of searching for incidental findings in the dutch-belgian lung cancer screening trial (nelson) using low-dose multidetector ct. *Eur Radiol* 2007; **17**: 1474–82. https://doi.org/10.1007/s00330-006-0532-7

75. Rampinelli C, Preda L, Maniglio M, Sirica L, Travaini LL, Veronesi G, et al. Extrapulmonary malignancies detected at lung cancer screening. *Radiology* 2011; **261**: 293–99. https://doi.org/10.1148/radiol.11102231

76. Nguyen XV, Davies L, Eastwood JD, Hoang JK. Extrapulmonary findings and malignancies in participants screened with chest ct in the national lung screening trial. *J Am Coll Radiol* 2017; **14**: 324–30. https://doi.org/10.1016/j.jacr.2016.09.044

77. Gareen IF, Black WC, Tosteson TD, Wang Q, Sicks JD, Tosteson ANA. Medical Care Costs Were Similar Across the Low-dose Computed Tomography and Chest X-Ray Arms of the National Lung

Screening Trial Despite Different Rates of Significant Incidental Findings. Med Care. 2018;56(5):403-9.

78. Incidental Findings Committee. Incidental Findings White Papers from the American College of Radiology. 2021. Available from: https://publish.smartsheet.com/42d18e87 4a164318a0f702481f2fbb70 [accessed 25 Feb 2021].

79 . Mazzone PJ, Thomson CC, Kazerooni EA, Smith RA, White CS. Proposed Quality Metrics for Lung Cancer Screening Programs: A National Lung Cancer Roundtable Project. Chest. 2021.

80. European Society of Thoracic Imaging (ESTI. ESTI Lung Cancer Screening Certification Project. 2019. Available from: https://www.myesti.org/lungcancerscreen ingcertificationproject

81. Heuvelmans MA, Oudkerk M, de Bock GH, de Koning HJ, Xie X, van Ooijen PMA, et al. Optimisation of volume-doubling time cutoff for fast-growing lung nodules in ct lung cancer screening reduces false-positive referrals. *Eur Radiol* 2013; **23**: 1836–45. https://doi.org/10.1007/s00330-013-2799-9

82. Horeweg N, van Rosmalen J, Heuvelmans MA, van der Aalst CM, Vliegenthart R, Scholten ET, et al. Lung cancer probability in patients with ct-detected pulmonary nodules: a prespecified analysis of data from the nelson trial of low-dose ct screening. *Lancet Oncol* 2014; **15**: 1332–41. https://doi. org/10.1016/S1470-2045(14)70389-4

83. Horeweg N, Scholten ET, de Jong PA, van der Aalst CM, Weenink C, Lammers J-W, et al. Detection of lung cancer through low-dose ct screening (nelson): a prespecified analysis of screening test performance and interval cancers. *Lancet Oncol* 2014; **15**: 1342–50. https://doi.org/10.1016/S1470-2045(14)70387-0

84. 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**: 2076–84. https://doi.org/10.1007/s00330-012-2437-y

85. Silva M, Schaefer-Prokop CM, Jacobs C, Capretti G, Ciompi F, van Ginneken B, et al. Detection of subsolid nodules in lung cancer screening: complementary sensitivity of visual reading and computer-aided diagnosis. *Invest Radiol* 2018; **53**: 441–49. https://doi.org/10.1097/RLI. 0000000000000464

86. Nair A, Bartlett EC, Walsh SLF, Wells AU, Navani N, Hardavella G, et al. Variable radiological lung nodule evaluation leads to divergent management recommendations.

Eur Respir J 2018; **52**(6): 1801359. https:// doi.org/10.1183/13993003.01359-2018

87. 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**: e873-8. https:// doi.org/10.1016/j.ejrad.2013.08.026

88. Ebner L, Roos JE, Christensen JD, Dobrocky T, Leidolt L, Brela B, et al. Maximum-intensity-projection and computer-aided-detection algorithms as stand-alone reader devices in lung cancer screening using different dose levels and reconstruction kernels. *AJR Am J Roentgenol* 2016; **207**: 282–88. https://doi.org/10.2214/AJR.15. 15588

89. Milanese G, Eberhard M, Martini K, Vittoria De Martini I, Frauenfelder T. Vessel suppressed chest computed tomography for semi-automated volumetric measurements of solid pulmonary nodules. *Eur J Radiol* 2018; **101**: 97–102. https://doi.org/10.1016/j. ejrad.2018.02.020

90. Gietema HA, Schaefer-Prokop CM, Mali W, Groenewegen G, Prokop M. Pulmonary nodules: interscan variability of semiautomated volume measurements with multisection ct-- influence of inspiration level, nodule size, and segmentation performance. *Radiology* 2007; **245**: 888–94. https://doi.org/10.1148/radiol.2452061054

91. Marchiano A, Calabro E, Civelli E, Di Tolla G, Frigerio LF, Morosi C, et al. Pulmonary nodules: volume repeatability at multidetector ct lung cancer screening. *Radiology* 2009; **251**: 919–25. https://doi. org/10.1148/radiol.2513081313

92. Revel MP, Bissery A, Bienvenu M, Aycard L, Lefort C, Frija G. Are two-dimensional ct measurements of small noncalcified pulmonary nodules reliable? *Radiology* 2004; **231**: 453–58. https://doi.org/10.1148/ radiol.2312030167

93. Soo E, Edey AJ, Mak SM, Moser J, Mohammadi S, Rodrigues T, et al. Impact of choice of volumetry software and nodule management guidelines on recall rates in lung cancer screening. *Eur J Radiol* 2019; **120**: 108646. https://doi.org/10.1016/j.ejrad. 2019.108646

94. QIBA. CT Volumetry Biomarker Ctte. 2021. Available from: http://qibawiki.rsna.org/ index.php/CT_Volumetry_Biomarker_Ctte

95. Devaraj A, van Ginneken B, Nair A, Baldwin D. Use of volumetry for lung nodule management: theory and practice. *Radiology* September 2017; **284**: 630–44. https://doi.org/10.1148/radiol.2017151022

96. de Hoop B, van Ginneken B, Gietema H, Prokop M. Pulmonary perifissural nodules on ct scans: rapid growth is not a predictor of malignancy. *Radiology* 2012; 265: 611–16. https://doi.org/10.1148/radiol.12112351

97. Schreuder A, van Ginneken B, Scholten ET, Jacobs C, Prokop M, Sverzellati N, et al. Classification of ct pulmonary opacities as perifissural nodules: reader variability. *Radiology* September 2018; 288: 867–75. https://doi.org/10.1148/radiol.2018172771

98. Barnett J, Pulzato I, Wilson R, Padley S, Nicholson AG, Devaraj A. Perinodular vascularity distinguishes benign intrapulmonary lymph nodes from lung cancer on computed tomography. *J Thorac Imaging* 2019; **34**: 326–28. https://doi.org/ 10.1097/RTI.0000000000000394

99. Ciompi F, de Hoop B, van Riel SJ, Chung K, Scholten ET, Oudkerk M, et al. Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2d views and a convolutional neural network out-of-the-box. *Med Image Anal* 2015; **26**: 195–202. https://doi.org/10. 1016/j.media.2015.08.001

100. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 2019; **25**: 954–61. https://doi.org/10.1038/s41591-019-0447-x

101. 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. https://doi.org/10. 1038/srep46479

102. Silva M, Milanese G, Seletti V, Ariani A, Sverzellati N. Pulmonary quantitative ct imaging in focal and diffuse disease: current research and clinical applications. *Br J Radiol* 2018; **91**: 20170644. https://doi.org/ 10.1259/bjr.20170644

103. Heuvelmans MA, van Ooijen PMA, Ather S, Silva CF, Han D, Heussel CP, et al. Lung cancer prediction by deep learning to identify benign lung nodules. *Lung Cancer* 2021; **154**: 1–4. https://doi.org/10.1016/j. lungcan.2021.01.027

104. Pastorino U, Boeri M, Sestini S, Sabia F, Silva M, Suatoni P, et al. Blood microrna and ldct reduce unnecessary ldct repeats in lung cancer sreening: results of prospective biomild trial. *World Congress of Lung Cancer* 2019; **9**: 2.

105. Nishi SPE, Zhou J, Okereke I, Kuo YF, Goodwin J. Use of imaging and diagnostic procedures after low-dose ct screening for lung cancer. *Chest* 2020; **157**: 427–34. https://doi.org/10.1016/j.chest.2019.08.2187

FUNCTIONAL IMAGING OF THE LUNG SPECIAL FEATURE: REVIEW ARTICLE

Artificial intelligence in functional imaging of the lung

RAÚL SAN JOSÉ ESTÉPAR

Applied Chest Imaging Laboratory, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, United States

Address correspondence to: Dr Raúl San José Estépar
E-mail: *rsanjose@bwh.harvard.edu*

ABSTRACT

Artificial intelligence (AI) is transforming the way we perform advanced imaging. From high-resolution image reconstruction to predicting functional response from clinically acquired data, AI is promising to revolutionize clinical evaluation of lung performance, pushing the boundary in pulmonary functional imaging for patients suffering from respiratory conditions. In this review, we overview the current developments and expound on some of the encouraging new frontiers. We focus on the recent advances in machine learning and deep learning that enable reconstructing images, quantitating, and predicting functional responses of the lung. Finally, we shed light on the potential opportunities and challenges ahead in adopting AI for functional lung imaging in clinical settings.

INTRODUCTION

The advent of artificial intelligence (AI) heralds a new era in digital data analysis and empowers us to interpret complex systems through unprecedented modeling capabilities. This power of AI has led to an explosion of applications across multiple disciplines including computer vision, and more recently, health care. Clinical care stands to benefit tremendously from AI to expose meaningful relationships in complex data sets obtained from clinical imaging to molecular medicine. Although AI still is a nascent field in many health-care domains, initial applications and proof-of-concept studies have shown promising and impactful results in diagnosing different disease conditions using only raw data sources like diagnostic imaging.[1,2] Thus, the immense analytic capacity of AI technology based on machine learning and deep learning will power human decision-making and complement human cognitive capabilities. Beyond equipping physicians with new abilities, data-driven modeling, as opposed to just model-based methods, is serving as a robust paradigm that can further improve the current cutting-edge algorithmic approaches in image formation, reconstruction, and post-processing.

The functional lung imaging community is recognizing the transformative power of AI. The data-driven approaches are well-positioned to invigorate established techniques in this field, improving robustness and often surpassing existing capabilities. Current functional lung imaging modalities utilize the underlying physics of the image properties related to different disease conditions of the lung.[3] The amount of data elements generated in functional imaging acquisitions, such as multiple MRI snapshots during free-breathing acquisitions or different CT energies, is amenable to applying data-driven approaches to discover novel relationships across different imaging phases, which otherwise would be difficult to identify. Various functional imaging modalities rely on advanced acquisitions and post-processing approaches, and hence AI is attractive as a primary modeling strategy.

Although AI applications in diagnostic imaging have increased rapidly in the last few years,[4,5] its clinical application to functional lung imaging is currently more of an evolving opportunity than a tested reality. Farhat et al[6] recently reviewed the application of deep learning in pulmonary medicine imaging and noticed that the use of AI in lung imaging is mostly circumscribed to chest CT and X-rays (CXR). In this review, we take a comprehensive look at the growing interest in applying AI technology specifically to pulmonary functional imaging and assess the underlying concepts of the proposed methodologies that utilize machine- and deep learning for state-of-the-art image reconstructions, functional assessment, and functional imaging synthesis. We evaluate the opportunities AI presents and weigh in on the challenges ahead for successfully implementing AI in pulmonary functional imaging.

Figure 1. Schematic of a CNN architecture. (A) Traditional CNN architecture is used for image classification or regression. An input image is decomposed into multiple globally aggregated features by a final-stage fully connected neural network. Convolutional layers are the main component in CNNs. Additional layers include data pooling to downsample the image domain, drop-out for model simplification, and batch normalization. (B) U-Net architecture is a type of fully convolutional network that is widely employed in medical imaging applications. U-Net contains two convolutional steps: an encoder and a decoder. The encoder reduces the input data to a latent space, and the decoder uses this information to recreate a new image. Adapted from Chartrand et al and Zha et al[10,11] with permission. CNN, convolutional neural network.
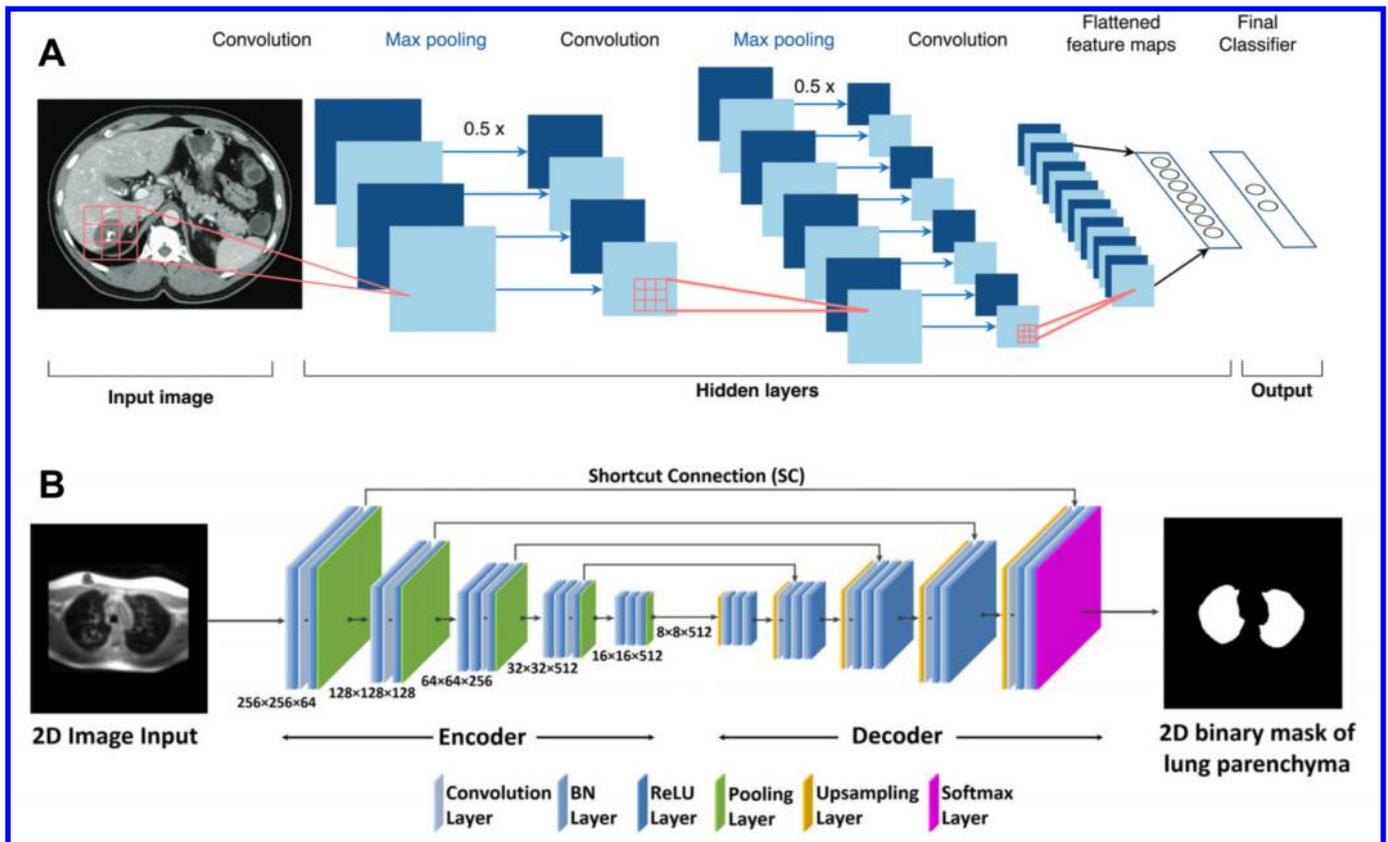
DEEP LEARNING IN MEDICAL IMAGING

The emergence of AI as a key component in medical imaging techniques is largely propelled by vast improvements in machine learning, specifically, deep learning. Deep learning performs a wide variety of challenging tasks, including classification, regression, clustering, image reconstruction artifact reduction, lesion detection, segmentation, and registration.[7] Deep learning is an extension of artificial neural networks[8] as a core building block. Deep learning gained importance in computer vision when neural networks outperformed other methods on several visual recognition tasks. Deep learning in medical imaging is primarily based on the convolutional neural network (CNN) paradigm. LeCun[9] introduced the CNNs to extend the use of neural networks from 1D signals to multi dimensional signals like 2D or 3D volumes and provide a powerful way to learn representations of images and solve recognition tasks. CNNs are constructed with units of a compact kernel of neurons that slides across an image to produce an output image map. Neurons act like logistic regressors that generate a response at each image location as a weighted sum of the image intensities. The kernels define the weight of each location, and these neural kernels are assembled in multiple channels to create a CNN convolutional

layer. Several such layers that function differently but complementary make up the CNN (Figure 1A). Information flows in a forward fashion, and deeper and deeper layers aggregate it in a non-linear manner. The success of CNNs in medical imaging inspired the development of other deep learning paradigms to exploit the various aspects of the information flowing through the network. A few examples of such advanced network methods are recurrent neural networks (RNN), autoencoders (AE), and its variations like U-Nets, generative adversarial networks (GANs), and more recently, transformers, among others.[12–14] Figure 1B illustrates the architecture of a U-Net used in medical applications to generate an output image from an input image after aggregating information at different scales. For more information, we refer the readers to the recent reviews of deep learning in radiology.[5,10,12,15]

Machine learning approaches can be classified into four major categories depending on the nature of the problem being solved and the data elements used as part of the training, *viz.* supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. A model maps a set of inputs to given outputs in supervised learning and requires annotated data

sets. Unsupervised learning aims at finding structure in a data set, as it is common in clustering problems. Semi-supervised learning has emerged as an exciting approach that combines supervised and unsupervised techniques to take advantage of non-annotated datasets that can improve supervised learning by matching specific characteristics of the non-annotated data set. An example of semi-supervised learning is image-to-image translation using GANs.[16] Finally, reinforcement learning is based on agents that learn from their environments through trial and error while optimizing some objective functions. An example of reinforcing learning in medical imaging is landmark detection methods.[17]

Finally, machine learning approaches define the model parameters using training data to solve an optimization problem. The proper definition of the training data set in terms of characteristics, sample size, and image conditions are key to converge to a solution that can be generalized to other data sets beyond the training examples. This implies that machine learning needs a thorough validation and testing of the models using data points that have been employed in training. Different techniques known as cross-validation are used to check the stability of the model when the training data change. It is essential to understand the conditions under which the model was derived, and the modelers need to follow good practices and careful documentation of the training process.[18]

AI IN FUNCTIONAL IMAGE RECONSTRUCTION
Magnetic resonance imaging (MRI)
MRI has been a primary modality in functional lung imaging because of its safety characteristics and the exceptional ability to discover functional properties.[19] The early challenges due to a lack of protons and signal inhomogeneities in the lungs have been overcome, and now MRI can be used for static and dynamic lung imaging.[20] The arrival of ultrashort TE (UTE) MRI with sophisticated clinical hardware has advanced lung imaging, both at the structural and functional levels.[21] From oxygen-enhanced and hyperpolarized gases MRI for ventilation imaging[21] to Fourier Decomposition proton MRI for ventilation/perfusion (V/Q) imaging and dynamic contrast enhancement (DCE) MRI for microvascular perfusion,[19] MRI has become the modality of choice to examine the complex ventilation and perfusion functions in different pathological conditions.[22] Essential to MRI pulse sequence design is the need for short echo times and the balance between acquisition time and signal-to-noise ratio (SNR) that can be achieved with parallel imaging.[23] Many of the computational approaches in MRI applications have been focused on improving optimal phase encoding from an under sampled version of the k-space that could reduce the acquisition time while keeping SNR levels compatible with image quality.[24] Compressed sensing techniques were developed two decades ago for fast MRI reconstruction, and using diffusion MRI with hyperpolarized ^{129}Xe.[25,26] In the past few years, CNNs and Recurrent NNs have taken a prominent role in improving static and dynamic MRI reconstruction to learning the spatiotemporal dependencies in heavily under sampled k-space data.[27–31] Duan et al[32] showed improved ventilation imaging using a coarse-to-fine neural network from under sampled

k-space.[32] Reconstruction can be achieved with higher SNR values than compressed sensing reconstruction, paving the way for real-time reconstruction of contrast-enhanced MRI of the lung. Unlike compressed sensing, CNN reconstruction models rely on incorporating prior information learned as part of the training process to solve the inverse reconstruction problem.[33] However, the reliance on data to define a model implies that rigorous validation is needed.[34]

Another area where deep learning can impact is the inherent need to perform motion correction in dynamic MRI acquisitions. For example, Fourier Decomposition MRI for V/Q Imaging relies on registration techniques as a critical step in their reconstruction paradigm. Likewise, different approaches have been proposed based on traditional functional optimization that shows stable quality results.[35] Deep learning registration offers an alternative with low computational cost during the inference stage once the registration model is trained.[36,37] Deep learning in MRI also has been attempted to estimate quantitative tissue parameters using quantitative susceptibility mapping (QMS) and MRI fingerprinting to achieve more standardized biomarkers.[38] Although these techniques are yet to be applied in both preclinical and clinical MRI lung imaging, deep learning could catalyze the translation of these advanced quantitative tools.

Computed tomography (CT)
Volumetric CT (VCT) has high-density contrast between air and tissue and is a mainstay of clinical chest radiology. The introduction of helical multislice CT scanning facilitated spatiotemporal 4DCT as a tool in radiation oncology for measuring and managing overall respiratory motion.[39] Patient safety is increased because only low dose radiation is required when combined with advanced iterative reconstruction techniques, and hence functional CT imaging (both 4D and dual-energy) is preferred for broader clinical use. Like MRI reconstruction, new AI methods are pushing ultra-low-dose CT image reconstruction to another level. Major manufacturers are introducing new deep learning schemes that show higher SNR and contrast and improved object detectability than standard statistical or model-based iterative techniques.[40–42] New techniques under development and current iterative reconstruction approaches capable of denoising CNNs promise to improve the image SNR further.[43] In addition to supporting low-dose image reconstruction, deep neural networks have also been employed to reduce breathing artifacts and enhance image quality.[44] All these advances will make temporal ultra-low CT a safer and more versatile functional modality in clinical applications of CT.

Cone-beam CT (CBCT) system is becoming a key device in the interventional suite due to portability and high reconstruction quality for volumetric images. In addition, deep learning is catalyzing dynamic applications with real-time reconstruction from sparse projection data permitting real-time ventilation imaging in image-guided radiotherapy.[45,46] The combination of these improvements can open the door for these preclinical CBCT applications to broader adoption as a lung functional imaging modality.[47]

Dual-energy CT scanning (DECT) with contrast agents (iodine or Xenon) has also enabled the assessment of regional ventilation and perfusion by taking advantage of the difference in linear attenuation coefficient at different X-ray energies.[48–52] CNNs are being applied to improve DECT imaging fundamentals related to material decomposition,[53,54] simplify dual-energy acquisitions based on single-energy material decomposition[55] and combine virtual single-energy structural imaging from dual-energy acquisitions. The translation of these techniques can expand the role of DECT in ventilation and perfusion imaging as dual-energy is more readily available.

Positron emission tomography (PET-CT) and single-photon emission computed tomography (SPECT) have also been employed to perform V/Q imaging to improve planar lung scintigraphy[56] and assess pulmonary inflammation.[57] Deep learning solutions are being developed to enhance PET reconstruction and attenuation correction[58,59]; however, up to date, no validation studies have been performed to show the impact of AI-enhanced molecular imaging in the lung. Thus, this area remains an exciting opportunity for AI in the years to come.

AI IN FUNCTIONAL QUANTIFICATION

Automated lung segmentation in functional modalities

For a functional imaging modality, it is important to define the structural components of the lung, such as lung field, lobar compartments, fissures, and the bronchovascular tree, to locate and quantitate image-based data. Deep learning is significantly evolving and transforming the post-acquisition upstream operations necessary to resolve the lung's structural components to interpret and quantify regional functional markers. Deep learning is indeed replacing the rule-based approaches to segment the lung[60] and the lobes with more precise and reliable mapping methods based on CNNs that have shown more consistent results across modalities.[6] In particular, the use of U-nets, a specialized neural network architecture for semantic segmentation, has provided compelling results in multiple medical and biomedical imaging segmentation tasks.[61,62] These new approaches to image segmentation are superior in part due to their enhanced ability to encode shape priors of the segmented organ based on the provided training data without explicitly modeling the shape. One example of the application of U-Nets to functional modalities is the use of a 2D U-Net to perform volumetric lung segmentation from UTE proton MRI in a multiplane fashion.[11] Despite reduced contrast around the lung boundaries, the lung volume estimates in a set of asthmatic and cystic fibrotic patients closely matched the reference values (Figure 2). One caveat for the application of deep learning is the limited availability of training data. Recently, Guo and colleagues showed increased robustness in UTE MRI lung segmentation by including an adaptive k-mean after the initial U-net segmentation.[63] Robust lung segmentation in MRI is essential for quantitative analysis of functional parameters and its use in clinical studies. Similarly, a multi resolution U-Net architecture has been proposed for robust lobar segmentation in CT images to enable regional quantification of dynamic CT series.[64,65]

Deformable image registration (DIR)

Ventilation imaging. DIR is one of the most employed methods to assess ventilation defects from temporal imaging modalities like 4DCT, CBCT, and MRI. DIR-enabled CT-based ventilation assessment has been successfully used in radiation oncology to avoid damage from radiation therapy as well as performing dose–response assessment.[39] Recently, MRI-based mechanical assessment of the lung via elastic registration has also been used to assess SSc-related fibrosis.[66] Ventilation assessment using tissue expansion metrics based on the deformation fields generated by DIR or the differences in tissue density between the coregistered image pairs have shown reasonable correlation with the regional assessment of ventilation using Xenon CT[48,67] and Xenon-MRI.[68] However, variability between registration approaches has led to a poor correlation between DIR-based ventilation metrics and reference modalities at the voxel level.[69]

Traditional DIR approaches describe the mapping of two images via a deformable field by finding the elastic transformation parameters that minimize the difference between images acquired at different moments during the respiratory cycle. Deformable registration in the lung has been challenging by the complexities of describing the transformation in a parametric way when dealing with large displacements commonly found in registration between images acquired between TLC and FRC while preserving known invariants like lung mass.[70] Nevertheless, traditional methods have partially addressed lung registration with reasonable accuracy performance, albeit with complex methodologies that lack robustness and require long computation times due to the numerical minimization needed for each registration instance.[71]

Deep learning-based deformation image registration (DLDIR) has emerged in the last 5 years as a new paradigm for registration. One of the main advantages of DLDIR approaches is the explicit or implicit definition of the deformation field via a CNN that can better capture the complexities of the deformation in a particular problem with relatively low computational needs during the inference step. DLDIR can be classified into supervised and unsupervised registration methods. Supervised approaches that regress the displacement vector field between two images using a CNN model were initially employed in DLDIR[72,73]R These methods were trained with previously aligned images using either a reference method[72] or synthetic deformations.[73,74] Although these approaches improve the registration computing times from minutes to just a few seconds, their accuracy is defined by the characteristics of the reference approach used for learning. The reported registration errors on reference data sets are on par with their traditional techniques that have been extensively used in ventilation studies. Unsupervised registration approaches have been explored to overcome the limitation of using an explicit reference deformation. Among them, unsupervised DLDIR has captured the attention in the last few years because it needs only limited training data.[75] Unsupervised techniques use a mismatch metric between the moving image and the reference image within the training data, as occurs in a traditional registration framework. A CNN model encodes the deformation parameters, and the optimization is done over

Figure 2. Segmentation of the lung field on oxygen-enhanced UTE MRI images using a multiplane (axial and coronal and final consensus) U-Net approach in a 37-year-old female with cystic fibrosis. The delineation of the lung boundaries can be achieved despite the reduced contrast. Adapted from Zha et al[11] with permission. UTE, ultrashort TE.
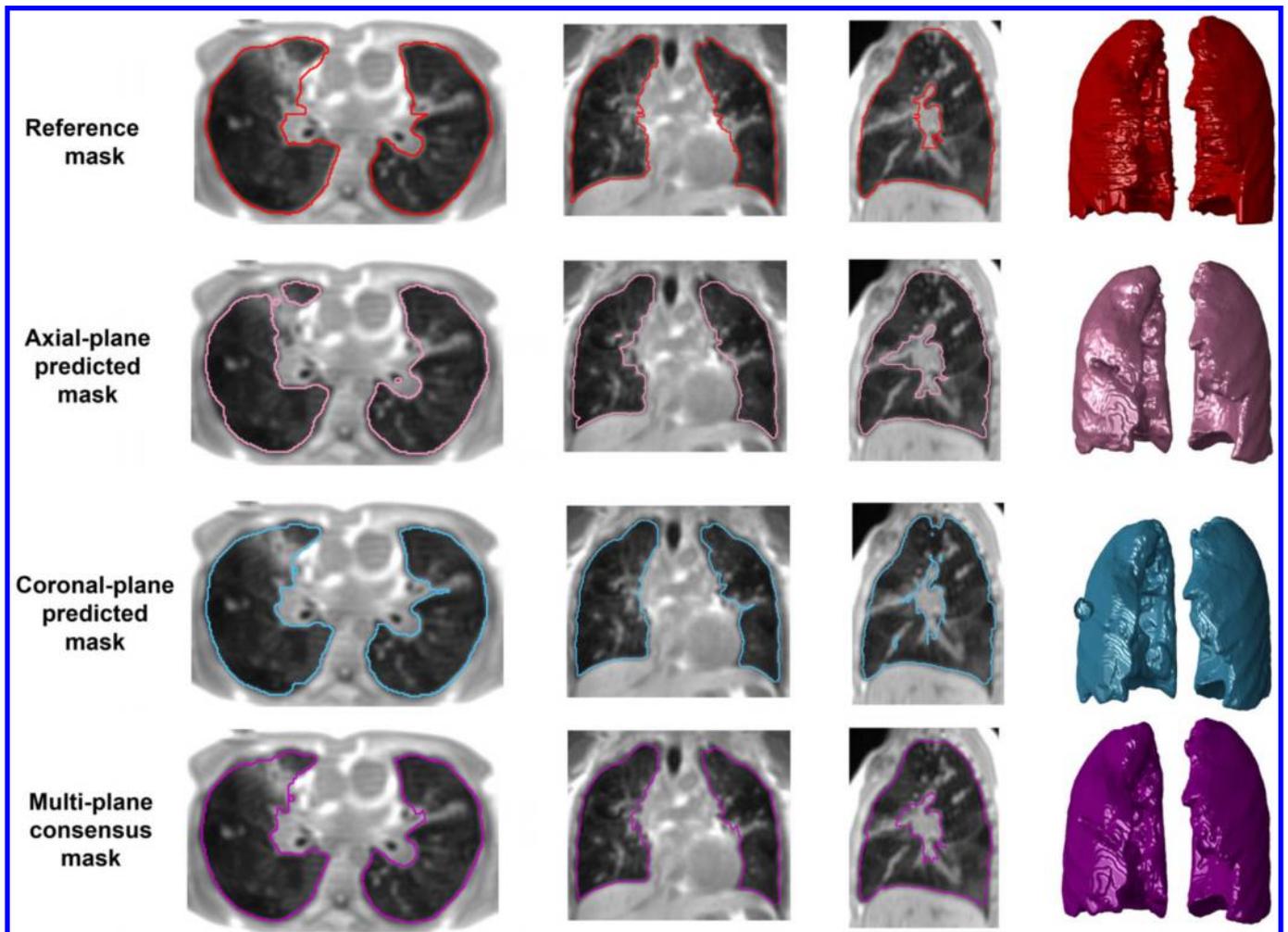

the parameters of the CNNs rather than the deformation parameters. Once the training is completed, the CNN is employed to generate new deformation parameters from unseen data sets. De Vos and colleagues[37] pioneered this framework in lung registration using the multiscale ConvNet architecture (Figure 3). A similar approach has been shown to be feasible to register CT to CBCT and CBCT to CBCT[76] and one-shot methods have been tailored to track periodic breathing motion patterns.[77] Finally, Fu et al[77] proposed a LungRegtNet for 4DCT registration that employs vascular landmarks to achieve superior performance compared to current methods based on unsupervised registration in the DIRLAb data set.[78]

The new breed of lung DLDIR approaches can lead to higher accuracy and more robust registration results that could improve the assessment of regional ventilation at the voxel level; however, extensive validation studies in larger prospective samples should be conducted to confirm this possibility.[69,79] Inaccurate registrations can result in lung tissue being mapped to blood vessel voxels which will cause artifacts in the CT-ventilation image in both the Jacobian and HU formulations. Without any doubt,

the most exciting characteristic of DLDIR is the need for low computation to resolve a deformation field once the method has been trained. This opens the opportunity for bringing DIR-based ventilation metrics closer to the patient point-of-care when applied to lower-cost setups like 4D CBCT. These exciting techniques are potential modalities for ventilation assessment during treatment in the near future.[39]

Multiparametric assessment. Registration is also a fundamental processing component of multiparametric structural and functional imaging analyses to correlate structural changes with functional defects in lung pathophysiology.[80–82] MacNeil et al[83] used volume-matched CT and hyperpolarized helium-3 (^{3}He) MRI using static and diffusion-weighted imaging to define a multiparametric response map (mPRM). Structural changes measured on CT were coupled with regional MRI-based ventilation and microstructure based on the apparent diffusion coefficient (ADC) as shown in Figure 4. mPRM metrics were able to reveal emphysema and small airways disease not otherwise identified with CT or MRI, reflecting the power of multimodal approaches in disease characterization. Registration approaches

Figure 3. An example of Unsupervised Deep Learning Deformable Image Registration from an expiratory (moving) to an inspiratory (fixed) CT scan. The CNN models the deformation field depicted as a warped grid. The Jacobian map estimates the volume change and can be used to compute ventilation maps. Registration inference can be performed in a few seconds in comparison to classical techniques enabling real-time deployment. Adapted from Vos BD de et al[37] with permission. CNN, convolutional neural network.
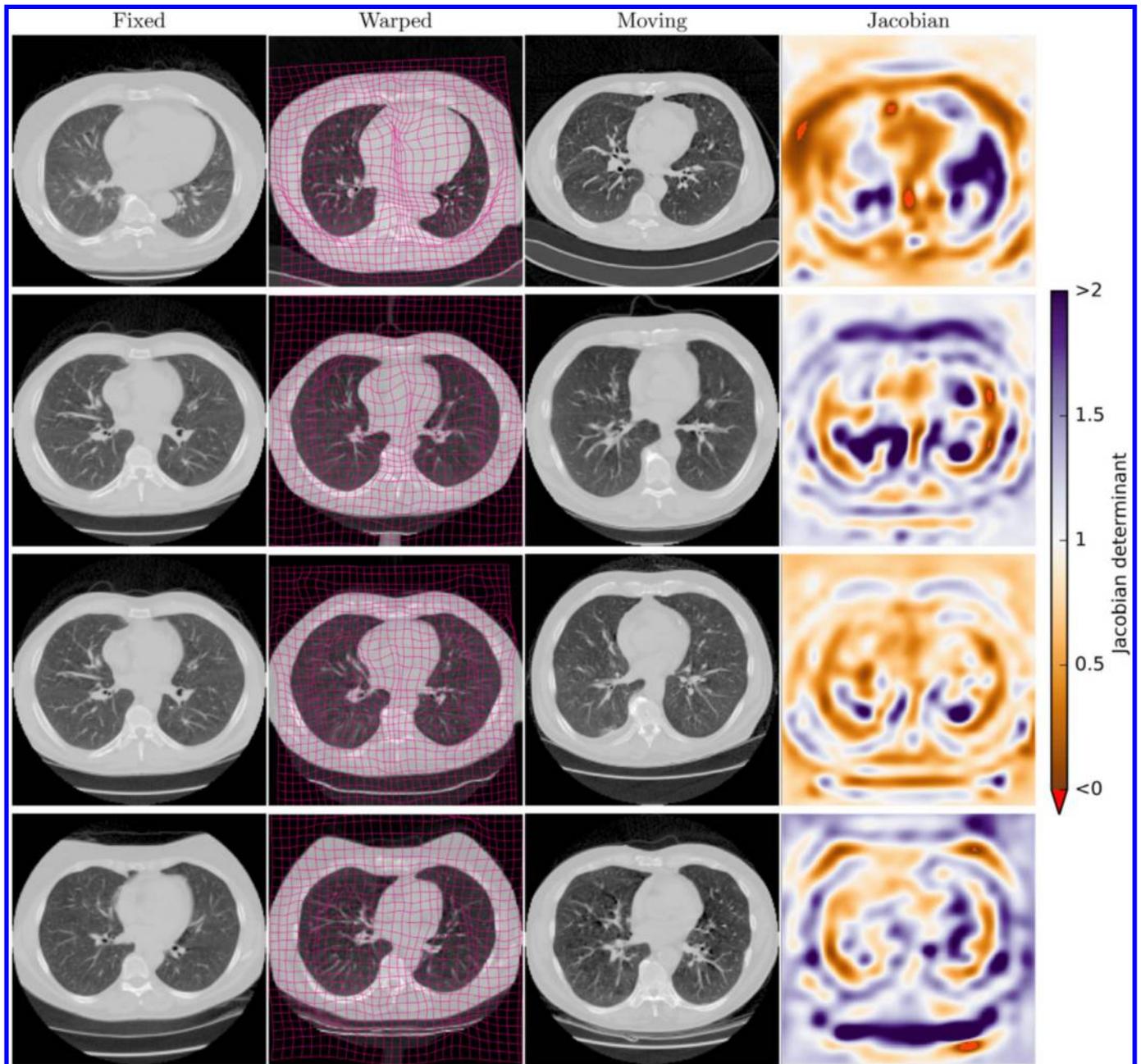

that include deep learning schemes will likely translate these upcoming multiparametric approaches to clinical applications beyond correlative studies.

Functional prediction

Deep learning approaches have been also postulated to predict the functional parameters from structural modalities. Westcott and colleagues showed how textural-based features extracted from a volume of interest on CT scans can predict regional ventilatory effects in subjects with COPD.[84] The method was trained with ventilation defect labels obtained from ^{1}H and ^{3}He MRI using a k-mean approach. Different classifiers were compared, and the most relevant features were selected in a cross-validation experimental setup. The AUC of the best model was 82%, with high specificity (91%) and moderate sensitivity (49%). Ventilation-defect percentage (VPD) predicted by the model and the one computed using the reference MRI modality show a strong correlation (90%); an encouraging sign of the ability of these approaches to offer patient-specific information on functional impairment conditions. However, it is hard to ascertain

Figure 4. Multiparametric imaging mapping from ^{3}He MRI and CT in COPD. Functional and structural images (top) are aligned to produce a multiparametric Response Map (bottom). DL Registration techniques can enable accurate and real-time response mapping assessment. Adapted from MacNeil et al[83] with permission. COPD, chronic obstructive pulmonary disease.
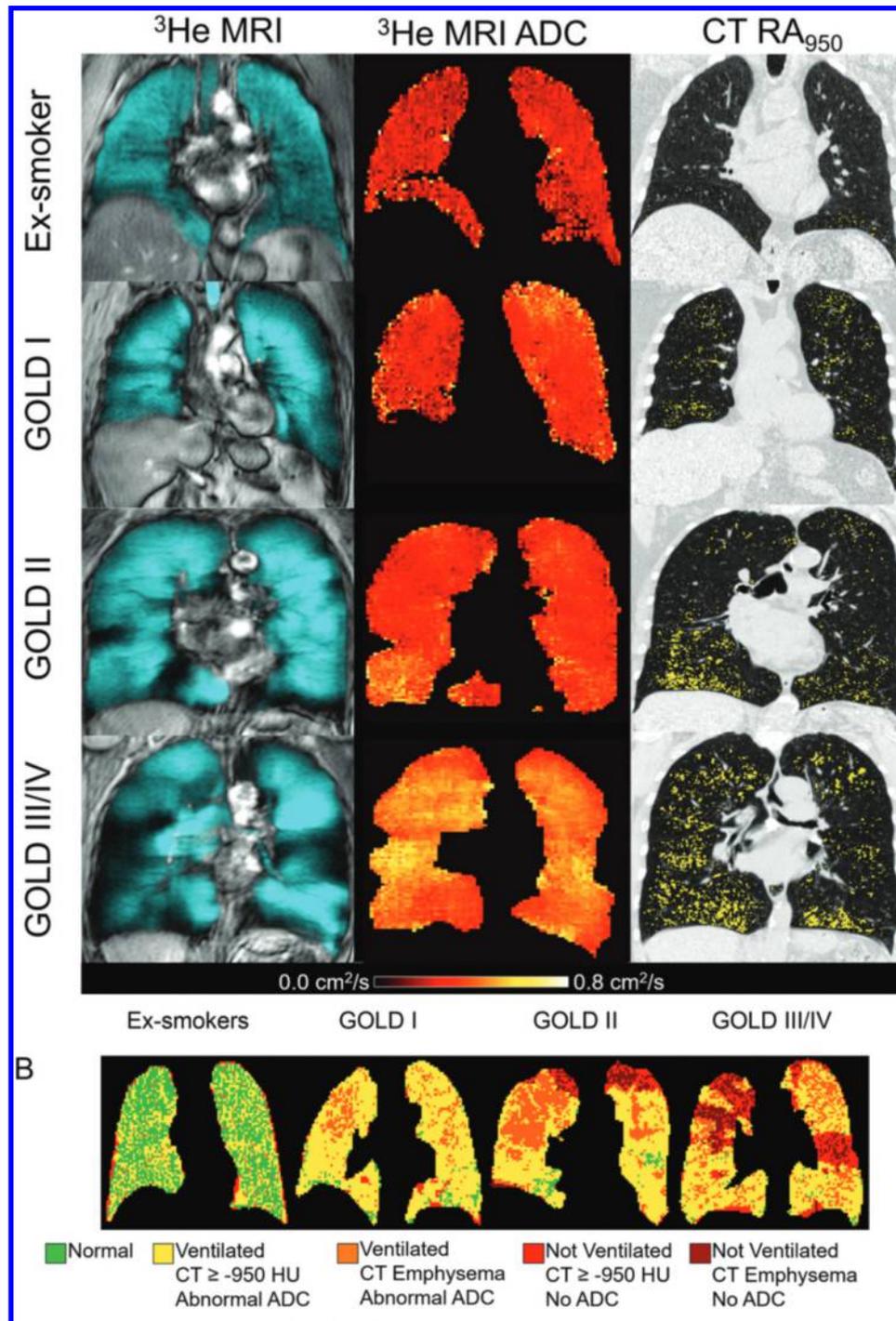
how stable the features proposed by this study could be generalized to a larger COPD population with milder disease conditions because of the limited sample size used for training. Larger sample size and reproducibility studies are needed to define the generalization power of the proposed features.

CNNs have also been used to extract features from CT images that can define spirometric status in smokers with and without COPD. Gonzalez et al[85] used a three-layer feed-forward CNN to predict COPD functional status based on spirometry. The correlation between FEV1 measurements and deep learning CT-based measurements was 73%. Tang and colleagues used a more complex network—a residual Network with 152 layers—to diagnose COPD from CT volumetric imaging.[86] The AUC in the testing cohort for the best model was 86%. This result was consistent with the performance reported by Gonzalez and colleagues.

These results suggest that different architectures can extract complementary feature information from CT imaging to predict an outcome.

In sum, the best network architecture design in terms of combinations of neural layers must strike a trade-off between model complexity and the ability to generalize to different populations and imaging characteristics. Meta-learning techniques are being actively researched and developed to improve upon the prediction of single learning techniques in multiple learning episodes that integrates different approaches.[87]

AI IN FUNCTIONAL ASSESSMENT

Function assessment is one of the most exciting emerging applications of AI where a direct functional response is synthesized to mimic a target functional modality, *e.g.* dual energy CT pulmonary perfusion, from source modalities that require simpler or a more direct imaging reconstruction setup. These techniques aim to resolve intrinsic relations across functional modalities or even the resolution of functional information from structural modalities like CT. These approaches are referred to as image-to-image translation within the AI community. They are based on an array of supervised and semi-supervised techniques that range from fully CNNs like convolutional generators based on autoencoders and U-nets[61] to Generative Adversarial Networks (GAN)[13] that combine a generator and a discriminator network. Image-to-image translation techniques were borne off in the context of computer graphics applications[88] and one prominent application is artificial style representation from natural images using paired (conditional) or unpaired (cycle) GANs.[16,89] In paired approaches, the training is performed in a data set containing paired instances of the target and source modality, while unpaired approaches can use instances from the source and the target modalities that are not matched or even belong to the same population of subjects.

Supervised functional synthesis

One of the first demonstrations of image translation approaches in functional lung images has been synthesizing ventilation imaging from 4DCT without explicit use of DIR. Unfortunately, 4DCT-derived ventilation images are sensitive to the choice of DIR algorithm and its accuracy.[90] Direct approaches can overcome this limitation by directly learning tissue expansion characteristics from multiple snapshots across a breathing cycle. Zhong et al[91] proposed a fully convolutional model composed of seven layers without any downsampling step to preserve the image resolution. Despite reasonable results, fully convolutional networks are limited to local relations between the inspiratory and expiratory images around a voxel that could lead to inconsistent results if the mismatch between inspiratory and expiratory images is significant.

To overcome some of the limitations of fully convolutional approaches, encoder–decoder convolutional like the U-net architecture have been extensively applied in image-to-image reconstruction tasks. The U-Net architecture includes multiple convolutional steps followed by a data down-sampling step in the encoder phase and up-sampling layers in the decoder phase. Also, information from the encoding phase at a given level is transferred to the decoder phase, similar to the fully convolutional approach. These architectures have shown promising results in synthesizing different functional ventilation images.[92,93] Capaldi et al[93] demonstrated the use of U-nets to estimate hyperpolarized noble gas MRI ventilation maps from free-breathing proton (^{1}H) MRI after breathing phase sorting and interpolation (Figure 5). Training, validation, and testing were done in a set of 114 subjects with different pulmonary conditions, *i.e.* asthma, COPD, bronchiectasis, and NSCLC, and healthy volunteers. The deep learning-based VDP estimation showed good agreement with reference values based on hyperpolarized ^{3}He MRI. In a similar fashion to Zhong et al.,[89] Gerard et al[92] used a multi resolution U-net to provide a direct estimation of the

Figure 5. Deep learning ventilation MRI for the synthesis of ^{3}He MRI ventilation imaging from free-breathing proton (^{1}H) MRI. (A) Illustration of the MRI pipeline to register and sort out free-breathing MRI images before consumption by the image-to-image U-Net network. The training was performed to predict ventilation maps from ^{3}He MRI. (B) Comparison between the reference ventilation maps and DL ventilation MRI synthetic imaging for subjects with different types of obstructive airway diseases. Agreement in ventilation defect percentage between modalities was high with good correspondence. Adapted from Capaldi et al[93] with permission.
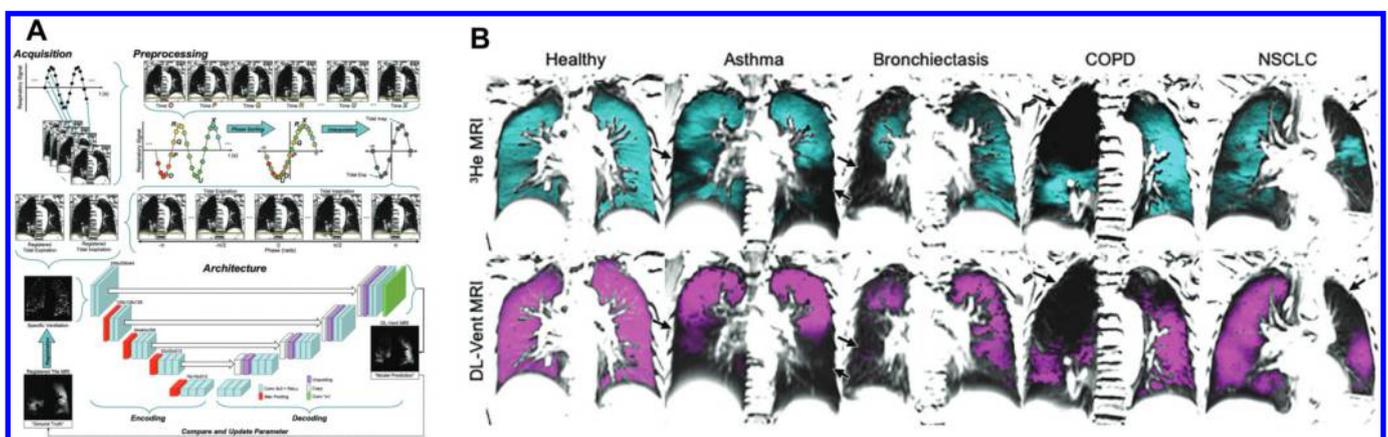
Figure 6. Illustration of image-to-image translation techniques for synthetic ventilation and perfusion assessment based on single energy CT. (A) Direct Jacobian ventilation map estimation using a multi resolution deep learning approach without deformation image registration from inspiratory and expiratory CT scans. (B) Estimation of dual-energy perfusion maps from single energy CT angiograms to assess perfusion defects using a *functional consistency* CycleGAN.
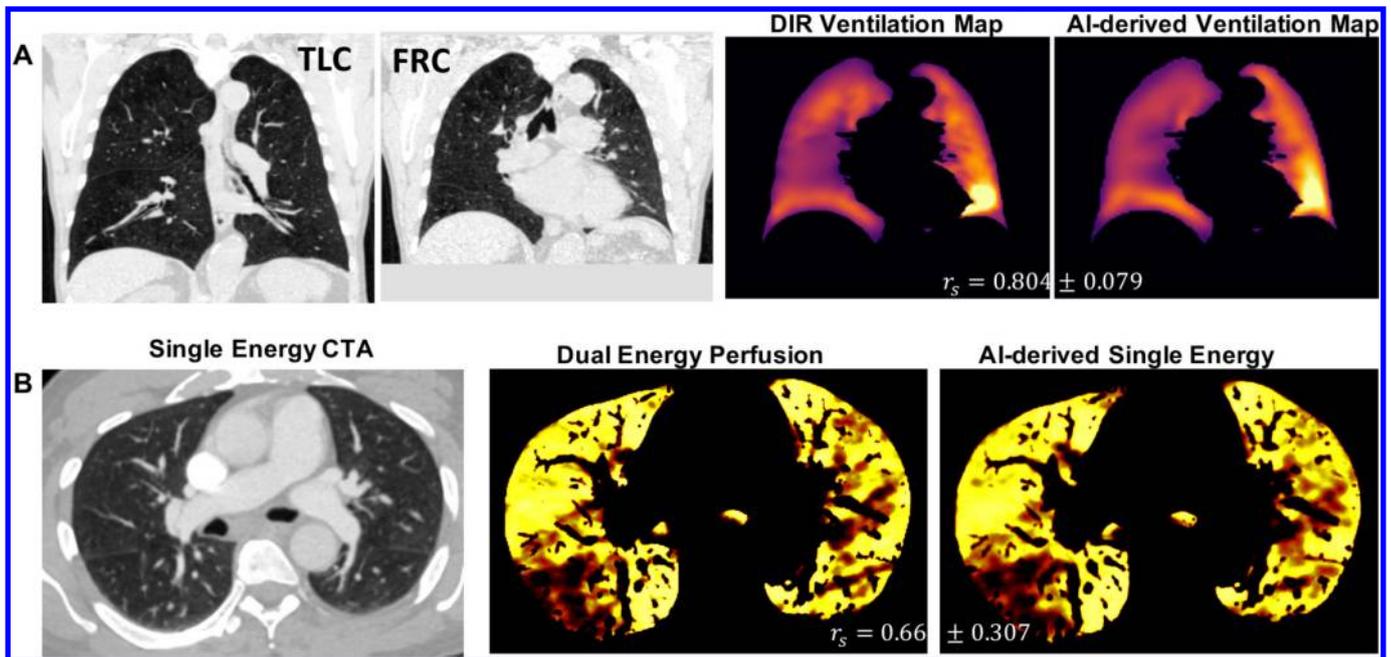
ventilation response based on the deformation Jacobian without a DIR (Figure 6A). Unlike prior approaches, this network was trained in an extensive database of inspiratory and expiratory CT scans from the COPDGene cohort and showed high voxel-wide correlations with ventilation images based on a classical mass-preserving DIR approach.

Like ventilation imaging, recent studies have also shown the use of CNN approaches to estimate lung perfusion from single energy CT scans. Ren et al[94] employed an attention U-net architecture to synthesize albumin SPECT/CT perfusion mapping from non-contrast CT scans to enable functional lung avoidance in radiotherapy planning.[94] Their proposed neural network is superior to the traditional U-Net architecture and is able to identify features from the CT domain that are compatible with perfusion defects with moderate correlation. Despite the limited size of the training (31 subjects) and testing data (11 subjects), these results illustrate the ability of deep learning approaches to estimate both ventilation and perfusion functional imaging from routine non-contrast CT scans under a common imaging platform.

Adversarial functional imaging

Semi-supervised approaches based on GANs are also under development as an improved alternative in image translation that aims at increasing the stability of the results of multi layered neural networks.[15] For example, Nardelli et al[95] illustrated the use of a modified conditional CycleGAN to synthesize dual-energy-derived iodine perfusion maps from single energy contrast CT scans (Figure 6B). The cycleGAN leverages both CT imaging and structural vascular information in a setup with

three encoding CNNs and three discriminators to generate the functional output with moderate local correlations (0.52 and 0.66 in the core and peel lung regions, respectively). Although unpaired GAN approaches are more complex and more challenging to train due to the need to find an equilibrium point in a min-max optimization problem, they seem relevant to approximate the statistical characteristics of the image that is being estimated. Unpaired GAN approaches can be employed with larger databases of unpaired datasets to predict the target functional modality without the need for scanning the same subject with both modalities as required by plain convolutional approaches.[93,94] Thus, the application of GANs presents a greater opportunity in the context of functional imaging. GAN-based learning can also be applied in various domains related to image reconstruction and preprocessing stages like super-resolution and multimodal registration and modality synthesis for multi-parametric analysis.

Opportunities and challenges

AI applications in medical imaging have exploded over the past 5 years, driven by multiple factors. First, the maturity of the deep learning approaches exploiting non-linear relations in the data has been instrumental. Second, advances in optimization and regularization techniques have made it tractable to fit models with a large number of parameters to a limited set of training data points. Third, the availability of methods in well-maintained open-source libraries has empowered a broad community with AI techniques, including non-experts in the field with limited skills. Finally, specialized computing architectures based on Graphics Processing Units (GPUs) have delivered the necessary

computing power to train advanced models within a reasonable amount of time.

While AI is still an emerging discipline in functional lung imaging, there are clear and tangible opportunities worth mentioning:

(1) **Multifunctional assessment**: AI has the potential to unleash the power of multiple functional assessments under a single imaging platform. Currently, ventilation and perfusion imaging require the use of different imaging contrast agents in CT. One potential integrated solution could be the emerging combination of 4D CBCT and simulated dual-energy imaging for functional imaging. The benefits of synthetic multifunctional assessment include reduced radiological tests that require hard-to-obtain radioactive contrast agents, reduced radiation exposure, and improved care delivery as imaging synthesis is performed without the patient as part of the radiological and clinical evaluation. However, realizing these opportunities will require an extensive validation process to define the interval confidence in which the synthetic images are consistent with the underlying functional ground truth. The outcome of the validation studies will also determine the potential of AI-enabled synthetic imaging for clinical adoption and whether it could eventually be circumscribed to narrower clinical scenarios where an initial triage based on a sub optimal approach might be useful.

(2) **Clinical translation to low footprint radiological setups:** current functional imaging relies on advanced modalities that require specialized equipment like hyperpolarizers. The potential use of AI-driven image-to-image translation could bring the benefit of functional information to standard radiological imaging modalities that are available in primary and secondary care facilities.

(3) **Novel biomarkers:** functional modalities provide voluminous multiparametric data that need to be laboriously synthesized into specific markers of disease. AI provides an alternative computational approach to define novel biomarkers of the disease. Supervised CNNs can be used to extract relevant image features that are associated with a specific outcome. Unsupervised autoencoder techniques can also be applied for dimensionality reduction to define novel biomarkers from multiparametric imaging sources.

(4) **Unraveling lung structure and function:** the relationship between structure and function of the lung has been well-described, but we are still limited in linking the structural changes to the functional impairment and achieving a better characterization of the disease. Studies that combine structural and functional modalities[83,96] can take advantage of AI as an exploratory tool to gain further insight into the structure–function relationship.

Despite the exciting and compelling preliminary evidence promising a more significant and elaborate role for AI in pulmonary functional imaging, several challenges remain that need to be carefully evaluated and resolved before realizing AI as a reliable component of clinical functional lung imaging:

(1) **Validation**: data-driven approaches require rigorous validation studies to gauge the generality and robustness of the methods. Until now, most of the studies that apply AI to functional lung imaging were performed with small datasets. Although they provide early evidence of what AI can do, they lack the rigor needed to qualify as bonafide approaches. Large databases on diverse populations will be required to train and validate the techniques before translating them into clinical use.

(2) **Model transparency**: one of the major criticisms of deep learning is a lack of transparency and interpretability. In other words, users (clinicians and researchers) should be able to understand the "reasoning" of the AI model; why it renders one verdict and not the other. Model developers and data scientists must make didactic efforts to teach the users how the models operate and decide outcomes. Transparency is crucial to defining a modality's operational realm and proactively restricting deviations from the model that can affect image quality and diagnostic interpretability.

(3) **Model robustness**: one collateral effect of the lack of model transparency is model instability to adversarial attacks (negligible input variations resulting in significant changes of the model output) and intrinsic model biases. Adversarial attack prevention is an oft-discussed topic in AI and they pose a substantial barrier to the use of AI for image synthesis in critical applications like diagnostic imaging.[97] Careful model design and training considerations must be taken to avoid adversarial attacks overall if models are trained with off-the-shelve components.[98] In a similar fashion, biases and disparity in functional expression may be translated into AI systems trained with imaging data in which those underlying biases exist.[99] Understanding the specific performance characteristics of each model is crucial to move beyond the preclinical scenario and successfully introduce it into clinical practice.

(4) **Unlocking data silos**: the unresolved complexities of functional imaging imply that the number of training cases is limited compared to training scenarios available for modalities like CT and CXR. Training sample size is a key factor in deep learning that depends on the specific characteristics of the problem begin addressed and the model that is used. Unlocking the available data silos is paramount for implementing new data-driven advances in functional lung imaging. Open data repositories and challenges like VAMPIRE[79] are crucial for developing machine learning-centric approaches that improve functional lung imaging quality and performance reasonably and reproducibly. Issues about data integrity and privacy could be overcome with federated solutions that enable de-centralized AI modeling to exploit pan-institutional datasets.[100,101]

CONCLUSION

AI continues to evolve rapidly and push the limits in many spheres, and its interest in medicine is growing exponentially in recent years, especially in the functional imaging domain. Public and private entities recognize this as a thrust area, and their initiatives have begun to catalyze this field.[100] The pulmonary functional imaging community can benefit from this

frenetic activity in data science as novel approaches using rich data sets are proposed to redefine disease conditions. Machine learning models that link imaging, functional, biomarkers, and multi omics data can advance our understanding of the complex and intimate connection between structure and function.[102] AI can also play a transformative role in adopting functional imaging approaches to clinical settings that are now restricted to preclinical scenarios due to their complexity. The use of functional modalities in diseases like chronic obstructive pulmonary disease (COPD), asthma, Interstital Lung Disease (ILD), or Cystic Fibrosis (CF) can bring a new dimensionality to define relevant markers of disease heterogeneity and progression.[82,103,104] At the same time, the application of AI is not free of limitations and perils stemming from the experimental nature of current techniques. The reliance on vast amounts of data exemplars rather than well-understood "fixed" models could act as a double-edged sword if AI is applied without careful methodological consideration. This issue is even more relevant in functional imaging scenarios where functional metrics describe subtle pathophysiological processes that need to be well-understood by the AI developers. Therefore, a multidisciplinary approach is essential to introduce AI in functional pulmonary imaging. Successful incorporation of AI in functional imaging holds promise to transform the field, delivering significant benefits in the coming years.

COMPETING INTERESTS

Dr. San José Estépar has no conflicts of interest to disclose related to the context of this manuscript. He is the founder and co-owner of Quantitative Imaging Solutions, a company that provides image-based consulting and develops software for data sharing and artificial intelligence applications.

FUNDING

REFERENCES

1. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus Photographs. *JAMA* 2016; **316**: 2402–10. doi: https://doi.org/10.1001/jama.2016.17216

2. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; **542**: 115–8. doi: https://doi.org/10.1038/nature21056

3. Ohno Y, Seo JB, Parraga G, Lee KS, Gefter WB, Fain SB, et al. Pulmonary functional imaging: part 1-State-of-the-Art technical and physiologic underpinnings. *Radiology* 2021; **299**: 508–23. doi: https://doi.org/10.1148/radiol.2021203711

4. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer* 2018; **18**: 500–10. doi: https://doi.org/10.1038/s41568-018-0016-5

5. Savadjiev P, Chong J, Dohan A, Vakalopoulou M, Reinhold C, Paragios N, et al. Demystification of AI-driven medical image interpretation: past, present and future. *Eur Radiol* 2019; **29**: 1616–24. doi: https://doi.org/10.1007/s00330-018-5674-x

6. Farhat H, Sakr GE, Kilany R. Deep learning applications in pulmonary medical imaging: recent updates and insights on COVID-19. *Mach Vis Appl* 2020; **31**: 53. doi: https://doi.org/10.1007/s00138-020-01101-5

7. Hatcher WG, Yu W. A survey of deep learning: platforms, applications and emerging research trends. *IEEE Access* 2018; **6**: 24411–32. doi: https://doi.org/10.1109/ACCESS.2018.2830661

8. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. 1943. *Bull Math Biol* 1990; **52**(1-2): 115–33.

9. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**: 436–44. doi: https://doi.org/10.1038/nature14539

10. Chartrand G, Cheng PM, Vorontsov E, Drozdzal M, Turcotte S, Pal CJ, et al. Deep learning: a primer for radiologists. *Radiographics* 2017; **37**: 2113–31. doi: https://doi.org/10.1148/rg.2017170077

11. Zha W, Fain SB, Schiebler ML, Evans MD, Nagle SK, Liu F. Deep convolutional neural networks with multiplane consensus labeling for lung function quantification using UTE proton MRI. *J Magn Reson Imaging* 2019; **50**: 1169–81. doi: https://doi.org/10.1002/jmri.26734

12. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw* 2015; **61**: 85–117. doi: https://doi.org/10.1016/j.neunet.2014.09.003

13. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S. Generative Adversarial networks. *arXiv* 2014.

14. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN. Attention is all you need. *Arxiv* 2017.

15. Wolterink JM, Mukhopadhyay A, Leiner T, Vogl TJ, Bucher AM, Išgum I. Generative Adversarial networks: a primer for radiologists. *Radiographics* 2021; **41**: 200151. doi: https://doi.org/10.1148/rg.2021200151

16. Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-Image translation with conditional Adversarial networks. *arXiv* 2016.

17. Ghesu F-C, Georgescu B, Zheng Y, Grbic S, Maier A, Hornegger J, et al. Multi-Scale deep reinforcement learning for real-time 3D-Landmark detection in CT scans. *IEEE Trans Pattern Anal Mach Intell* 2019; **41**: 176–89. doi: https://doi.org/10.1109/TPAMI.2017.2782687

18. Haibe-Kains B, Adam GA, Hosny A, Khodakarami F, Waldron L, et al.Massive Analysis Quality Control (MAQC) Society Board of Directors Transparency and reproducibility in artificial intelligence. *Nature* 2020; **586**: E14–16. doi: https://doi.org/10.1038/s41586-020-2766-y

19. Biederer J, Mirsadraee S, Beer M, Molinari F, Hintze C, Bauman G, et al. MRI of the lung (3/3)-current applications and future perspectives. *Insights Imaging* 2012; **3**: 373–86. doi: https://doi.org/10.1007/s13244-011-0142-z

20. Mills GH, Wild JM, Eberle B, Van Beek EJR, Beek E. Functional magnetic resonance imaging of the lung. *Br J Anaesth* 2003; **91**: 16–30. doi: https://doi.org/10.1093/bja/aeg149

21. Kruger SJ, Nagle SK, Couch MJ, Ohno Y, Albert M, Fain SB. Functional imaging of the lungs with gas agents. *J Magn Reson Imaging* 2016; **43**: 295–315. doi: https://doi.org/10.1002/jmri.25002

22. Biederer J, Heussel C, Puderbach M, Wielpuetz M. Functional magnetic resonance imaging of the lung. *Semin Resp Crit Care* 2014; **35**: 74–82. doi: https://doi.org/10.1055/s-0033-1363453

23. Wild JM, Marshall H, Bock M, Schad LR, Jakob PM, Puderbach M, et al. MRI of the lung (1/3): methods. *Insights Imaging* 2012; **3**: 345–53. doi: https://doi.org/10.1007/s13244-012-0176-x

24. Ouriadov A, Guo F, McCormack DG, Parraga G. Accelerated 129Xe MRI morphometry of terminal airspace enlargement: feasibility in volunteers and those with alpha-1 antitrypsin deficiency. *Magnet Reson Med* 2020; **84**: 416–26.

25. Zhang H, Xie J, Xiao S, Zhao X, Zhang M, Shi L, et al. Lung morphometry using hyperpolarized 129Xe Xe multi-b diffusion MRI with compressed sensing in healthy subjects and patients with COPD. *Med Phys* 2018; **45**: 3097–108. doi: https://doi.org/10.1002/mp.12944

26. Collier GJ, Hughes PJC, Horn FC, Chan H, Tahir B, Norquay G, et al. Single breath-held acquisition of coregistered 3D 129Xe lung ventilation and anatomical proton images of the human lung with compressed sensing. *Magnet Reson Med* 2019; **82**: 342–7.

27. Wang S, Su Z, Ying L, Peng X, Zhu S, Liang F, et al. Accelerating Magnetic Resonance Imaging via Deep Learning. *2016 Ieee 13th Int Symposium Biomed Imaging Isbi*. 2016; **2016**: 514–7.

28. Schlemper J, Caballero J, Hajnal JV, Price AN, Rueckert D. A deep cascade of Convolutional neural networks for dynamic Mr image reconstruction. *IEEE Trans Med Imaging* 2018; **37**: 491–503. doi: https://doi.org/10.1109/TMI.2017.2760978

29. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. *Nature* 2018; **555**: 487–92. doi: https://doi.org/10.1038/nature25988

30. Strack R. AI transforms image reconstruction. *Nat Methods* 2018; **15**: 309. doi: https://doi.org/10.1038/nmeth.4678

31. Lin DJ, Johnson PM, Knoll F, Lui YW. Artificial intelligence for MR image reconstruction: an overview for clinicians. *J Magn Reson Imaging* 2021; **53**: 1015–28. doi: https://doi.org/10.1002/jmri.27078

32. Duan C, Deng H, Xiao S, Xie J, Li H, Sun X, et al. Fast and accurate reconstruction of human lung gas MRI with deep learning. *Magn Reson Med* 2019; **82**: 2273–85. doi: https://doi.org/10.1002/mrm.27889

33. Wu Y, Ma Y, Capaldi DP, Liu J, Zhao W, Du J, et al. Incorporating prior knowledge via volumetric deep residual network to optimize the reconstruction of sparsely sampled MRI. *Magn Reson Imaging* 2020; **66**: 93–103. doi: https://doi.org/10.1016/j.mri.2019.03.012

34. Antun V, Renna F, Poon C, Adcock B, Hansen AC. On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proc Natl Acad Sci U S A* 2020; **117**: 30088–95. doi: https://doi.org/10.1073/pnas.1907377117

35. Ljimani A, Hojdis M, Stabinska J, Valentin B, Frenken M, Appel E, et al. Analysis of different image-registration algorithms for Fourier decomposition MRI in functional lung imaging. *Acta Radiol* 2021; **62**: 028418512094490. doi: https://doi.org/10.1177/0284185120944902

36. Yang X, Kwitt R, Styner M, Niethammer M. Quicksilver: Fast predictive image registration - A deep learning approach. *Neuroimage* 2017; **158**: 378–96. doi: https://doi.org/10.1016/j.neuroimage.2017.07.008

37. de Vos BD, Berendsen FF, Viergever MA, Sokooti H, Staring M, Išgum I. A deep learning framework for unsupervised affine and deformable image registration. *Med Image Anal* 2019; **52**: 128–43. doi: https://doi.org/10.1016/j.media.2018.11.010

38. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Z Med Phys* 2019; **29**: 102–27. doi: https://doi.org/10.1016/j.zemedi.2018.11.002

39. Vinogradskiy Y. Ct-Based ventilation imaging in radiation oncology. *BJR Open* 2019; **1**: 20180035. doi: https://doi.org/10.1259/bjro.20180035

40. Park C, Choo KS, Jung Y, Jeong HS, Hwang J-Y, Yun MS. Ct iterative vs deep learning reconstruction: comparison of noise and sharpness. *Eur Radiol* 2021; **31**: 3156–64. doi: https://doi.org/10.1007/s00330-020-07358-8

41. Lenfant M, Chevallier O, Comby P-O, Secco G, Haioun K, Ricolfi F, et al. Deep learning versus iterative reconstruction for CT pulmonary angiography in the emergency setting: improved image quality and reduced radiation dose. *Diagnostics* 2020; **10**: 558. doi: https://doi.org/10.3390/diagnostics10080558

42. Brady SL, Trout AT, Somasundaram E, Anton CG, Li Y, Dillman JR. Improving image quality and reducing radiation dose for pediatric CT by using deep learning reconstruction. *Radiology* 2021; **298**: 180–8. doi: https://doi.org/10.1148/radiol.2020202317

43. Hata A, Yanagawa M, Yoshida Y, Miyata T, Tsubamoto M, Honda O, et al. Combination of deep Learning–Based denoising and iterative reconstruction for Ultra-Low-Dose CT of the chest: image quality and Lung-RADS evaluation. *American Journal of Roentgenology* 2020; **215**: 1321–8. doi: https://doi.org/10.2214/AJR.19.22680

44. Mori S, Hirai R, Sakata Y. Using a deep neural network for four-dimensional CT artifact reduction in image-guided radiotherapy. *Phys Med* 2019; **65**: 67–75. doi: https://doi.org/10.1016/j.ejmp.2019.08.008

45. Chen G, Zhao Y, Huang Q, Gao H. 4D-AirNet: a temporally-resolved CBCT slice reconstruction method synergizing analytical and iterative method with deep learning. *Phys Med Biol* 2020; **65**: 175020. doi: https://doi.org/10.1088/1361-6560/ab9f60

46. Jiang Z, Chen Y, Zhang Y, Ge Y, Yin F-F, Ren L. Augmentation of CBCT reconstructed from Under-Sampled projections using deep learning. *IEEE Trans Med Imaging* 2019; **38**: 2705–15. doi: https://doi.org/10.1109/TMI.2019.2912791

47. Murrie RP, Werdiger F, Donnelley M, Lin Y-wei, Carnibella RP, Samarage CR, et al. Real-Time in vivo imaging of regional lung function in a mouse model of cystic fibrosis on a laboratory X-ray source. *Sci Rep* 2020; **10**: 447. doi: https://doi.org/10.1038/s41598-019-57376-w

48. Kong X, Sheng HX, Lu GM, Meinel FG, Dyer KT, Schoepf UJ, et al. Xenon-enhanced dual-energy CT lung ventilation imaging: techniques and clinical applications. *AJR Am J Roentgenol* 2014; **202**: 309–17. doi: https://doi.org/10.2214/AJR.13.11191

49. Hoey ETD, Mirsadraee S, Pepke-Zaba J, Jenkins DP, Gopalan D, Screaton NJ. Dual-Energy CT angiography for assessment of regional pulmonary perfusion in patients with chronic thromboembolic pulmonary hypertension: initial experience. *AJR Am J Roentgenol* 2011; **196**: 524–32. doi: https://doi.org/10.2214/AJR.10.4842

50. Goo HW, Goo JM. Dual-Energy CT: new horizon in medical imaging. *Korean J Radiol* 2017; **18**: 555–69. doi: https://doi.org/10.3348/kjr.2017.18.4.555

51. Fuld MK, Halaweish AF, Haynes SE, Divekar AA, Guo J, Hoffman EA. Pulmonary perfused blood volume with dual-energy CT as surrogate for pulmonary perfusion assessed with dynamic multidetector CT. *Radiology* 2013; **267**: 747–56. doi: https://doi.org/10.1148/radiol.12112789

52. Iyer KS, Newell JD, Jin D, Fuld MK, Saha PK, Hansdottir S, et al. Quantitative dual-energy computed tomography supports a vascular etiology of smoking-induced inflammatory lung disease. *Am J Respir Crit Care Med* 2016; **193**: 652–61. doi:

https://doi.org/10.1164/rccm.201506-
1196OC

53. Zhang W, Zhang H, Wang L, Wang X, Hu
X, Cai A, et al. Image domain dual material
decomposition for dual-energy CT using
butterfly network. *Med Phys* 2019; **46**:
2037–51. doi: https://doi.org/10.1002/mp.
13489

54. Abascal JFPJ, Ducros N, Pronina V, Rit S,
Rodesch P-A, Broussaud T, et al. Material
decomposition in spectral CT using deep
learning: a Sim2Real transfer approach.
IEEE Access 2021; **9**: 25632–47. doi:
https://doi.org/10.1109/ACCESS.2021.
3056150

55. Lyu T, Zhao W, Zhu Y, Wu Z, Zhang Y, Chen
Y, et al. Estimating dual-energy CT imaging
from single-energy CT data with material
decomposition convolutional neural
network. *Med Image Anal* 2021; **70**: 102001.
doi: https://doi.org/10.1016/j.media.2021.
102001

56. Dimastromatteo J, Charles EJ, Laubach VE.
Molecular imaging of pulmonary diseases.
Respir Res 2018; **19**: 17. doi: https://doi.org/
10.1186/s12931-018-0716-0

57. Vass L, Fisk M, Lee S, Wilson FJ, Cheriyan
J, Wilkinson I. Advances in PET to assess
pulmonary inflammation: a systematic
review. *Eur J Radiol* 2020; **130**: 109182. doi:
https://doi.org/10.1016/j.ejrad.2020.109182

58. Dong X, Lei Y, Wang T, Higgins K, Liu
T, Curran WJ, et al. Deep learning-based
attenuation correction in the absence of
structural information for whole-body
positron emission tomography imaging.
Phys Med Biol 2020; **65**: 055011. doi: https://
doi.org/10.1088/1361-6560/ab652c

59. Gong K, Guan J, Kim K, Zhang X, Yang J,
Seo Y. Iterative PET image reconstruction
using Convolutional neural network
representation. *IEEE Trans Med Imaging*
2019; **38**: 675–85. doi: https://doi.org/10.
1109/TMI.2018.2869871

60. Sluimer I, Schilham A, Prokop M,
van Ginneken B, van GB. Computer analysis
of computed tomography scans of the lung:
a survey. *IEEE Trans Med Imaging* 2006; **25**:
385–405. doi: https://doi.org/10.1109/TMI.
2005.862753

61. Ronneberger O, Fischer P. Brox T. U-Net:
Convolutional Networks for Biomedical
Image Segmentation. In: Navab N, Joachim
H, Wells William M, Frangi A. F, eds.; 2015.
pp.. 234–41.

62. Isensee F, Jaeger PF, Kohl SAA, Petersen
J, Maier-Hein KH. nnU-Net: a self-
configuring method for deep learning-
based biomedical image segmentation. *Nat
Methods* 2021; **18**: 203–11. doi: https://doi.
org/10.1038/s41592-020-01008-z

63. Guo F, Capaldi DP, McCormack DG,
Fenster A, Parraga G. Ultra-Short echo-
time magnetic resonance imaging lung
segmentation with under-Annotations and
domain shift. *Med Image Anal* 2021; **72**:
102107. doi: https://doi.org/10.1016/j.media.
2021.102107

64. Gerard SE, Patton TJ, Christensen GE,
Bayouth JE, Reinhardt JM. FissureNet: a
deep learning approach for pulmonary
fissure detection in CT images. *IEEE Trans
Med Imaging* 2019; **38**: 156–66. doi: https://
doi.org/10.1109/TMI.2018.2858202

65. Gerard SE, Herrmann J, Kaczka DW, Musch
G, Fernandez-Bustamante A, Reinhardt
JM. Multi-resolution convolutional neural
networks for fully automated segmentation
of acutely injured lungs in multiple species.
Med Image Anal 2020; **60**: 101592. doi:
https://doi.org/10.1016/j.media.2019.101592

66. Chassagnon G, Martin C, Marini R,
Vakalopoulou M, Régent A, Mouthon L, et al.
Use of elastic registration in pulmonary
MRI for the assessment of pulmonary
fibrosis in patients with systemic sclerosis.
Radiology 2019; **291**: 487–92. doi: https://
doi.org/10.1148/radiol.2019182099

67. Reinhardt JM, Ding K, Cao K, Christensen
GE, Hoffman EA, Bodas SV. Registration-
based estimates of local lung tissue
expansion compared to xenon CT measures
of specific ventilation. *Med Image Anal* 2008;
12: 752–63. doi: https://doi.org/10.1016/j.
media.2008.03.007

68. Tahir BA, Marshall H, Hughes PJC,
Brightling CE, Collier G, Ireland RH, et al.
Comparison of CT ventilation imaging and
hyperpolarised gas MRI: effects of breathing
manoeuvre. *Phys Med Biol* 2019; **64**: 055013.
doi: https://doi.org/10.1088/1361-6560/
ab0145

69. Hegi-Johnson F, de Ruysscher D, Keall P,
Hendriks L, Vinogradskiy Y, Yamamoto
T, et al. Imaging of regional ventilation:
is CT ventilation imaging the answer? A
systematic review of the validation data.
Radiother Oncol 2019; **137**: 175–85. doi:
https://doi.org/10.1016/j.radonc.2019.03.010

70. Yin Y, Hoffman EA, Lin C-L. Mass
preserving nonrigid registration of CT lung
images using cubic B-spline. *Med Phys* 2009;
36: 4213–22. doi: https://doi.org/10.1118/1.
3193526

71. Murphy K, van Ginneken B, Reinhardt JM,
Kabus S, Ding K, Deng X, et al. Evaluation
of registration methods on thoracic CT:
the EMPIRE10 challenge. *IEEE Trans Med
Imaging* 2011; **30**: 1901–20. doi: https://doi.
org/10.1109/TMI.2011.2158349

72. Onieva JO, Marti-Fuster B. Image analysis
for moving organ, breast, and thoracic

images, third International workshop,
RAMBO 2018, fourth International
workshop, BIA 2018, and first International
workshop, TIA 2018, held in conjunction
with MICCAI 2018, Granada, Spain,
September 16 and 20, 2018, proceedings.
Lect Notes Comput Sc 2018; **11040**: 284–94.
doi: https://doi.org/10.1007/978-3-030-
00946-5

73. Sokooti H, de VB, Berendsen F, Lelieveldt
BPF, Išgum I, Staring M. Medical image
computing and computer assisted
intervention – MICCAI 2017: 20th
International Conference, Quebec City,
Qc, Canada, September 11-13, 2017,
proceedings, part I. *Lect Notes Comput Sc*
2017; **3**: 232–9. doi: https://doi.org/10.1007/
978-3-319-66182-7_27

74. Eppenhof KAJ, Pluim JPW. Pulmonary CT
registration through supervised learning
with Convolutional neural networks. *IEEE
Trans Med Imaging* 2019; **38**: 1097–105. doi:
https://doi.org/10.1109/TMI.2018.2878316

75. Haskins G, Kruger U, Yan P. Deep learning
in medical image registration: a survey.
Mach Vis Appl 2020; **31**(1-2): 8. doi: https://
doi.org/10.1007/s00138-020-01060-x

76. Jiang Z, Yin F-F, Ge Y, Ren L. A multi-scale
framework with unsupervised joint training
of convolutional neural networks for
pulmonary deformable image registration.
Phys Med Biol 2020; **65**: 015011. doi: https://
doi.org/10.1088/1361-6560/ab5da0

77. Fechter T, Baltas D. One-Shot learning for
deformable medical image registration and
periodic motion tracking. *IEEE Trans Med
Imaging* 2020; **39**: 2506–17. doi: https://doi.
org/10.1109/TMI.2020.2972616

78. Castillo R, Castillo E, Fuentes D, Ahmad M,
Wood AM, Ludwig MS, et al. A reference
dataset for deformable image registration
spatial accuracy evaluation using the
COPDgene study archive. *Phys Med Biol*
2013; **58**: 2861–77. doi: https://doi.org/10.
1088/0031-9155/58/9/2861

79. Kipritidis J, Tahir BA, Cazoulat G, Hofman
MS, Siva S, Callahan J, et al. The vampire
challenge: a multi-institutional validation
study of CT ventilation imaging. *Med Phys*
2019; **46**: 1198–217. doi: https://doi.org/10.
1002/mp.13346

80. Guo F, Svenningsen S, Kirby M, Capaldi
DP, Sheikh K, Fenster A, et al. Thoracic CT-
MRI coregistration for regional pulmonary
structure-function measurements of
obstructive lung disease. *Med Phys* 2017; **44**:
1718–33. doi: https://doi.org/10.1002/mp.
12160

81. Hoffman EA, Lynch DA, Barr RG, van Beek
EJR, Parraga G. IWPFI Investigators
Pulmonary CT and MRI phenotypes that

help explain chronic pulmonary obstruction disease pathophysiology and outcomes. *J Magn Reson Imaging* 2016; **43**: 544–57. doi: https://doi.org/10.1002/jmri.25010

82. Washko GR, Parraga G. COPD biomarkers and phenotypes: opportunities for better outcomes with precision imaging. *Eur Respir J* 2018; **52**: 52. doi: https://doi.org/10.1183/13993003.01570-2018

83. MacNeil JL, Capaldi DPI, Westcott AR, Eddy RL, Barker AL, McCormack DG, et al. Pulmonary imaging phenotypes of chronic obstructive pulmonary disease using multiparametric response maps. *Radiology* 2020; **295**: 227–36. doi: https://doi.org/10.1148/radiol.2020191735

84. Westcott A, Capaldi DPI, McCormack DG, Ward AD, Fenster A, Parraga G. Chronic obstructive pulmonary disease: thoracic CT texture analysis and machine learning to predict pulmonary ventilation. *Radiology* 2019; **293**: 676–84. doi: https://doi.org/10.1148/radiol.2019190450

85. Serrano GG, Ash SY, Vegas-Sánchez-Ferrero G, Onieva JO, Rahaghi FN, Ross JC. Disease staging and prognosis in smokers using deep learning in chest computed tomography. *Am J Respir Crit Care Med* 2017; **197**: 193–203. doi: https://doi.org/10.1164/rccm.201705-0860OC

86. Tang LYW, Coxson HO, Lam S, Leipsic J, Tam RC, Sin DD. Towards large-scale case-finding: training and validation of residual networks for detection of chronic obstructive pulmonary disease using low-dose CT. *Lancet Digit Health* 2020; **2**: e259–67. doi: https://doi.org/10.1016/S2589-7500(20)30064-9

87. Hospedales T, Antoniou A, Micaelli P, Storkey A. Meta-Learning in neural networks: a survey. *Arxiv* 2020.

88. Ruan C, Yuan L, Hu H, Chen D. Image translation with dual-directional generative adversarial networks. *IET comput. vis* 2021; **15**: 73–83. doi: https://doi.org/10.1049/cvi2.12011

89. Zhu J-Y, Park T, Isola P, Efros AA. Unpaired Image-to-Image translation using Cycle-Consistent Adversarial networks. *arXiv.org* 2017;: cs.CV:.

90. Latifi K, Forster KM, Hoffe SE, Dilling TJ, van Elmpt W, Dekker A, et al. Dependence of ventilation image derived from 4D CT on deformable image registration and ventilation algorithms. *J Appl Clin Med Phys* 2013; **14**: 150–62. doi: https://doi.org/10.1120/jacmp.v14i4.4247

91. Zhong Y, Vinogradskiy Y, Chen L, Myziuk N, Castillo R, Castillo E, et al. Technical note: deriving ventilation imaging from 4DCT by deep convolutional neural network. *Med Phys* 2019; **46**: 2323–9. doi: https://doi.org/10.1002/mp.13421

92. Gerard SE, Reinhardt JM, Christensen GE, Estépar RSJ. Estimating local tissue expansion in thoracic computed tomography images using Convolutional neural networks. 2020; 1856–60. doi: https://doi.org/10.1109/ISBI45749.2020.9098413

93. Capaldi DPI, Guo F, Xing L, Parraga G. Pulmonary ventilation maps generated with Free-breathing proton MRI and a deep Convolutional neural network. *Radiology* 2021; **298**: 427–38. doi: https://doi.org/10.1148/radiol.2020202861

94. Ren G, Lam S-K, Zhang J, Xiao H, Cheung AL-Y, Ho W-Y, et al. Investigation of a novel deep Learning-Based computed tomography perfusion mapping framework for functional lung avoidance radiotherapy. *Front Oncol* 2021; **11**: 644703. doi: https://doi.org/10.3389/fonc.2021.644703

95. Nardelli P, Estépar RSJ, Rahaghi FN, Estépar RSJ. Functional-Consistent CycleGAN for CT to iodine perfusion map translation. *Lecture Notes In Computer Science* 2020; **12502**: 109–17. doi: https://doi.org/10.1007/978-3-030-62469-9_10

96. Eddy RL, Parraga G. Pulmonary xenon-129 MRI: new opportunities to unravel enigmas in respiratory medicine. *Eur Respir J* 2020; **55**: 1901987. doi: https://doi.org/10.1183/13993003.01987-2019

97. Ma X, Niu Y, Gu L, Wang Y, Zhao Y, Bailey J, et al. Understanding adversarial attacks on deep learning based medical image analysis systems. *Pattern Recognit* 2021; **110**: 107332. doi: https://doi.org/10.1016/j.patcog.2020.107332

98. Bortsova G, González-Gonzalo C, Wetstein SC, Dubost F, Katramados I, Hogeweg L, et al. Adversarial attack vulnerability of medical image analysis systems: unexplored factors. *Med Image Anal* 2021; **73**: 102141. doi: https://doi.org/10.1016/j.media.2021.102141

99. Larrazabal AJ, Nieto N, Peterson V, Milone DH, Ferrante E. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proc Natl Acad Sci U S A* 2020; **117**: 201919012. doi: https://doi.org/10.1073/pnas.1919012117

100. Langlotz CP, Allen B, Erickson BJ, Kalpathy-Cramer J, Bigelow K, Cook TS, et al. A roadmap for foundational research on artificial intelligence in medical imaging: from the 2018 NIH/RSNA/ACR/The Academy workshop. *Radiology* 2019; **291**: 190613. doi: https://doi.org/10.1148/radiol.2019190613

101. Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. *Nat Mach Intell* 2020; **2**: 305–11. doi: https://doi.org/10.1038/s42256-020-0186-1

102. Weibel ER. It takes more than cells to make a good lung. *Am J Respir Crit Care Med* 2013; **187**: 342–6. doi: https://doi.org/10.1164/rccm.201212-2260OE

103. Hatabu H, Ohno Y, Gefter WB, Parraga G, Madore B, Lee KS, et al. Expanding applications of pulmonary MRI in the clinical evaluation of lung disorders: Fleischner Society position paper. *Radiology* 2020; **297**: 286–301. doi: https://doi.org/10.1148/radiol.2020201138

104. Woods JC, Wild JM, Wielpütz MO, Clancy JP, Hatabu H, Kauczor H-U, et al. Current state of the art MRI for the longitudinal assessment of cystic fibrosis. *J Magn Reson Imaging* 2020; **52**: 1306–20. doi: https://doi.org/10.1002/jmri.27030

FULL PAPER

Application of deep learning image reconstruction in low-dose chest CT scan

HUANG WANG, LU-LU LI, JIN SHANG, JIAN SONG and BIN LIU, PhD

Department of Radiology, The Fourth Affiliated Hospital of Anhui Medical University, Hefei, China

Address correspondence to: **Mr Bin Liu**
E-mail: *Lbhyz32@126.com*

The authors Huang Wang and Lu-Lu Li contributed equally to the work.

Objective: Deep learning image reconstruction (DLIR) is a new reconstruction method for maintaining image quality at reduced radiation dose. The purpose of this study was to compare image quality of reduced-dose DLIR images with the standard-dose adaptive statistical iterative reconstruction (ASIR-V) images in chest CT.
Methods: Our prospective study included 48 adult patients (30 women and 18 men, mean age ±SD, 49.8 ± 14.3 years) who underwent both the standard-dose CT (SDCT) and low-dose CT (LDCT) on a GE Revolution CT scanner. All patients gave written informed consent. All scans were reconstructed with ASIR-V40%. Additionally, LDCT scans were reconstructed with DLIR with high-setting (DLIR-H) and medium-setting (DLIR-M). Image noise and contrast-noise-ratio (CNR) of thoracic aorta with different reconstruction modes were measured and compared.
Results: LDCT reduced radiation dose by 96% compared with SDCT (CTDIvol: 0.54mGy *vs* 12.46mGy). In LDCT, DLIR significantly reduced image noise compared with

the state-of-the-art ASIR-V40% with DLIR-H provided the lowest image noise and highest image quality score. In addition, the image noise, CNR of aorta and overall image quality of the low-dose DLIR-H images did not have significant difference compared with the SDCT ASIR-V40% images (all *p* > 0.05).
Conclusion: DLIR significantly reduces image noise in LDCT chest scans and provides similar image quality as the SDCT ASIR-V images at 4% of the radiation dose.
Advances in knowledge: DLIR uses high-quality FBP data to train deep neural networks to learn how to distinguish between signal and noise, and effectively suppresses noise without affecting anatomical and pathological structures. It opens a new era of CT image reconstruction. DLIR significantly reduces image noise and improves image quality compared with ASIR-V40% under same radiation dose condition. DLIR-H achieves similar image quality at 4% radiation dose as ASIR-V40% at standard-dose level in non-contrast chest CT.

INTRODUCTION

In recent years, lung cancer has become the number one cause of death in several countries. The American National Lung Screening Trial (NLST) has demonstrated that compared with screening using chest X-rays, using low-dose CT (LDCT) has led to 20% reduction in lung cancer mortality by screening heavy smokers.[1,2] CT has replaced conventional chest radiographs as the preferred method for chest examination. Thus, disease can be controlled or treated in its asymptomatic state.[3] However, compared to chest X-ray, patients may receive 10–100 times of radiation dose when a standard-dose chest CT scan is used. The potential risk of radiation-induced malignancy related to the burgeoning use of CT and radiation exposure has attracted a lot of attention.[2] In CT examination, the radiation dose delivered to the patients is a public health concern.[4] The

radiation related risk may be increasing due to the increase of CT examination. Low-dose chest CT is considered as a screening method for early detection of lung cancer in the population at risk.[5] So low-dose chest CT scan has been used more frequently than the standard-dose chest CT in clinical application in the early lung cancer screening,[6] and reducing the radiation dose while maintaining or improving image quality is a goal many people have been pursuing. Filtered back projection (FBP) was the standard reconstruction for CT which has the fastest image reconstruction time. However, when images are reconstructed using conventional FBP under low-dose scan conditions, high image noise and artifacts cannot be ignored.[7,8] Iterative reconstruction techniques have been introduced that can significantly reduce the image noise and provide more possibilities for reducing the radiation dose, compared with

the FBP reconstruction.[6] The adaptive statistical iterative reconstruction (ASIR-V) is a new generation iterative reconstruction algorithm that has advantages of further improving image quality and/or reducing radiation dose compared with its predecessor.[8] However, in general, iterative reconstructions at high levels may cause a plastic-looking, blotchy, and unnatural image appearance that will eventually reduce the image quality and affect clinicians' diagnosis of diseases, limiting the ability for deep radiation dose reduction.

Recently, a deep learning image reconstruction (DLIR) algorithm (TrueFidelity[TM], GE Healthcare Waukesha, WI) has been introduced to address some of the unsolved difficult scientific and technical problems of iterative reconstruction algorithms. DLIR is a CT image reconstruction method applied with a deep convolutional neural network (DCNN) to improve image quality.[9] DLIR uses highly selected, essentially artifact-free FBP image sets of both phantoms and patients to train the software and has been shown to improve image quality or maintain image quality under lower radiation doses in the abdominal and coronary CT applications.[2,10] The purpose of our study was to evaluate image quality in terms of image noise, contrast-to-noise ratio (CNR) of DLIR chest CT images under an extremely low-dose scan condition and compare with those of ASIR-V40% images under both standard- and low-dose conditions.

METHODS AND MATERIALS

Study population
This was a prospective study approved by the internal review board of our hospital and all patients gave written informed consent for participating the study. From May 10 to September 28, 2020, 49 consecutive patients (18 men, 31 women; with mean age ± standard deviation of 49.8 ± 14.3 years; body mass index (BMI) of 18.5–24.0 kg/m²) underwent an extremely LDCT scan of the chest after a clinically indicated standard-dose CT (SDCT). One patient was later excluded from analysis due to motion artifacts. Thus, 48 patients (18 men, 30 women) were finally included in this study.

Scan technology and image acquisition
All patients were scanned on a 256-slice CT scanner (Revolution CT, GE Healthcare) while in supine position with arms raised overhead to prevent artifacts. All patients were instructed to avoid any voluntary motion and to carefully follow the breath-hold instructions. The standard- and low-dose scan protocols were used in an inspiratory breath-hold cycle to ensure that the lesions were in the same position in both scans. The scan parameters were as follows: (1), standard-dose scanning: voltage, 120 kV; and automatic tube current; gantry rotation time, 0.5 s; helical pitch, 0.992:1. (2), low dose scanning: voltage, 80 kV; tube current, 50mA; gantry rotation time, 0.5 s; helical pitch, 0.992:1. Both LDCT and SDCT images were reconstructed at a slice thickness of 1.25 mm and with ASIR-V at a strength level of 40% (ASIR-V40%). In addition, the LDCT scan data sets were reconstructed with DLIR at the medium (DLIR-M) and high (DLIR-H) levels.

To assess the radiation dose, the volume CT dose index (CTDI$_{vol}$) and dose–length product (DLP) were recorded for the SDCT and LDCT imaging series. The estimated effective dose (ED) was calculated as DLP multiplied by a k-factor of 0.014 mSv·mGy^{-1}·cm^{-1} for the chest.

Objective image analysis
All reconstructed (ASIR-V40%@LDCT, ASIR-V40%@SDCT, DLIR-M@LDCT, DLIR-H@LDCT) images were transmitted to a GE AW 4.7 workstation for data measurement and image analysis. Images were reviewed in both a lung setting (window level, −600 HU; window width, 1500 HU) and mediastinal setting (window level, 40 HU; window width, 350 HU). One radiologist with 3 years of working experience in medical imaging performed an objective image analysis on the axial images. The reconstructions were linked so that identical region of interest (ROI) could be drawn in the same location on each reconstruction. Three ROIs with area of 150 mm² were drawn in the aorta, paraspinal musculature, and subcutaneous fat. For each reconstruction, the contrast-to-noise ratio (CNR) relative to muscle was calculated for the aorta as (ROIi – ROIm) / SD, where ROIi is the mean attenuation for the anatomy of interest (aorta), ROIm is the mean attenuation of paraspinal muscles, and SD is the mean image noise based on the measurement for subcutaneous fat, calculated as the mean SD of attenuation in HUs. These measurements were performed in all four reconstruction modes.[2]

Subjective image analysis
Two other radiologists with more than 10 years of experience in medical imaging, separately performed the lesion identification and objective image analysis of the reconstructed images. The patients' information and the image reconstruction modes were hidden. The radiologists were blinded to the patients' data and image reconstruction techniques. A 5-point scoring system for subjective evaluation of image quality,[6] including aspects of morphological display, visibility for surrounding lung tissue and diagnostic confidence for lung lesions (including solid nodules and ground-glass nodules) was used (Table 1).

Statistical analysis
Data were recorded in Excel (Microsoft Office 2016) and analyzed with SPSS statistical software (v. 22.0, IBM SPSS Statistics). The objective data were expressed as mean ± SD. Radiation dose between LDCT and SDCT was compared using the Student's *t*-test. The differences among the CT images reconstructed with ASIR-V40% (with LDCT and SDCT data), DLIR-M, DLIR-H (with LDCT data) were evaluated. The one-way ANOVA with Bonferroni Correction was used to compare the quantitative CT measurements and the Kruskal–Wallis Wilcoxon rank sum test and Dunnett' *t*-test was used to compare image quality across the different dose levels and different reconstruction techniques for chest CT. For the subjective analysis, we calculated the interobserver agreement using the κ statistic to evaluate the agreement between the two readers. A *p*-value of less than 0.05 was considered statistically significant.

RESULTS

Basic information of patients and radiation dose
A total of 48 patients (18 men, 30 women) were finally included in this study. A total of 97 solid nodules and 30 ground-glass

Table 1. Subjective score criteria for image quality evaluation

Grading score	Qualitative image analysis		
	Morphological display of all nodules	Visibility for surrounding lung tissue	Artifacts and diagnostic confidence
1	Very poor display, unclear edge	Unacceptable visibility, cannot distinguish small structures	Severe artifacts, insufficient confidence
2	Poor display, fuzzy edge	Small structures are not displayed very well, seriously impact diagnosis	Substantial artifacts, insufficient confidence
3	Moderate display, not very clear edge	Small structures can be displayed, and enough for diagnosis	Moderate artifacts, low confidence but diagnosis possible
4	Better display, still clear edge	Small structures can be clearly displayed with good contrast	Minor artifacts, good diagnostic confidence
5	Excellent display, clear edge	Small structures can be clearly displayed with excellent contrast	No artifacts, excellent diagnostic confidence

nodules were identified in the study in SDCT. There was no significant difference in the detection rate of nodules between the SDCT, LDCT, DLIR-M and DLIR-H. The difference was in the appearance of the nodules.

From the images, the average X-ray tube current of SDCT was 413.78 ± 4.81 mA. As for the radiation dosage, the mean CTDIvol, DLP, ED were 12.46 ± 1.16 mGy, 447.32 ± 34.51 mGy*cm, 6.26 ± 0.48 mSv in SDCT and 0.54 ± 0.00 mGy, 19.44 ± 1.37 mGy*cm, 0.27 ± 0.20 mSv in LDCT, respectively with about 96% dose reduction in LDCT (all $p < 0.001$) (Table 2).

Objective analysis
The objective image analysis results are presented in Table 3. The image signals (CT number) were conformed to normal distribution and did not have significantly difference across different reconstructions ($p = 0.2$). But the other parameters, including image noise (SD), and CNR, differed significantly. For image noise, the SD value did not have significantly difference between ASIR-V40%@SDCT and DLIR-H@LDCT ($p = 1.000$), while the differences between any other reconstruction pairs were all statistically significant (DLIR-M@LDCT *vs* ASIR-V40%@LDCT, $p = 0.006$ and ASIR-V40%@SDCT *vs* DLIR-M@LDCT, ASIR-V40%@SDCT *vs* ASIR-V40%@LDCT, DLIR-H@LDCT *vs* DLIR-M@LDCT, DLIR-H@LDCT *vs* ASIR-V40%@LDCT, all $p < 0.001$ (Figure 1)). The CNR for the aorta did not have statistically significant difference between ASIR-V40%@SDCT and DLIR-H@LDCT, ($p = 1.000$), DLIR-M@LDCT and ASIR-V40%@LDCT, ($p = 0.625$), and ASIR-V40%@SDCT and DLIR-M@LDCT, ($p = 0.163$), and DLIR-H@LDCT and DLIR-M@LDCT, ($p = 0.181$); while there was statistically significant difference between ASIR-V40%@SDCT and ASIR-V40%@

LDCT and between DLIR-H@LDCT and ASIR-V40%@LDCT (all $p < 0.05$).

Subjective analysis
The subjective analysis is summarized in Tables 4 and 5 and Figure 2 . The results showed that, There was no significant difference in the detection rate of nodules among the different reconstructions. The difference was in the appearance of the nodules.The DLIR images had a better quality than the ASIR-V40% at LDCT. Moreover, compared with SDCT, DLIR-H images at LDCT had similar image quality for evaluating the nodules in terms of morphological display of nodules, visibility for surrounding lung tissue, artifacts, and diagnostic confidence (Tables 4 and 5). There was substantial agreement between the two readers (Kappa>0.7).

DISCUSSION
During the history of CT development, people have made unremitting efforts to reduce radiation dose while maintaining the image quality and diagnostic accuracy. FBP was the standard reconstruction for CT. However, with the decrease of radiation dose, image quality was greatly affected, prominent noise and artifacts occurred.[11] Then, several methods of maintaining image quality while reducing radiation dose were introduced in succession, such as IR, model-based IR (MBIR), ASIR and ASIR–V (GE Healthcare).[12] IR techniques have been introduced to reduce image noise or maintain good CT image quality on reduced-dose CT scans. ASIR (GE Healthcare) was the first commercially available IR algorithm.[13] One can choose the percentages of blending IR with FBP to obtain the desired balance between noise reduction, spatial resolution and image appearance for clinical application, and is a major advance in the development

Table 2. Comparison of radiation dose between the two scanning modes

|	CTDIvol (mGy)	DLP (mGy.cm)	ED (mSv)
Standard-dose scan	12.46±1.16	447.32±34.51	6.26±0.48
Low-dose scan	0.54±0.00	19.44±1.37	0.27±0.20
p	<0.001	<0.001	<0.001

CTDIvol, volume CT dose index; DLP, dose–length product; ED, effective dose.

Table 3. Comparison of quantitative measurements among ASIR-V and DLIR under different radiation doses

| Variables | SDCT | LDCT | DLIR-M | DLIR-H | P | *p* | | | | | |
						SDCT *vs* DLIR-H	SDCT *vs* DLIR-M	SDCT *vs* LDCT	DLIR-H *vs* DLIR-M	DLIR-H *vs* LDCT	DLIR-M *vs* LDCT
Image noise	12.4 ± 2.1	28.6 ± 3.9	20.4 ± 2.6	12.5 ± 7.2	<0.001	1.000	<0.001	<0.001	<0.001	<0.001	0.006
CNR aorta	0.9 ± 0.6	0.4 ± 0.3	0.6 ± 0.5	0.9 ± 0.7	<0.001	1.000	0.163	0.004	0.181	0.007	0.625
Image signal	46.9 ± 6.6	49.0 ± 9.5	50.2 ± 8.9	50.4 ± 8.0	0.2	0.098	0.124	0.325	0.905	0.497	0.576

ASIR, adaptive statistical iterative reconstruction; DLIR, deep learning image reconstruction; DLIR-H: DLIR-H with low-dose; DLIR-M: DLIR-M with low-dose;LDCT, low-dose CT; SDCT, standard-dose CT; SDCT: ASIR-V40% with standard-dose.

Figure 1. Comparison of chest CT scan in axial soft tissue window images of mediastinum in 43-year-old male (A), ASIR-V40% at SDCT; (B), ASIR-V40% at LDCT; (C), DLIR-M at LDCT; and (D), DLIR-H at LDCT. In different reconstructions, the image attenuation values (CT numbers) did not have statistically significant difference. For image noise, the SD value did not have significantly difference between SDCT and DLIR-H (*p* = 1.000), while there were statistically significant differences between any other reconstruction pairs: DLIR-M *vs* LDCT, *p* = 0.006, and SDCT *vs* DLIR-M, SDCT *vs* LDCT, DLIR-H *vs* DL-M, DLIR-H *vs* LDCT, all *p* < 0.001. ASIR, adaptive statistical iterative reconstruction; DLIR, deep learning image reconstruction; LDCT, low-dose CT; SDCT, standard-dose CT; SD, standard deviation.
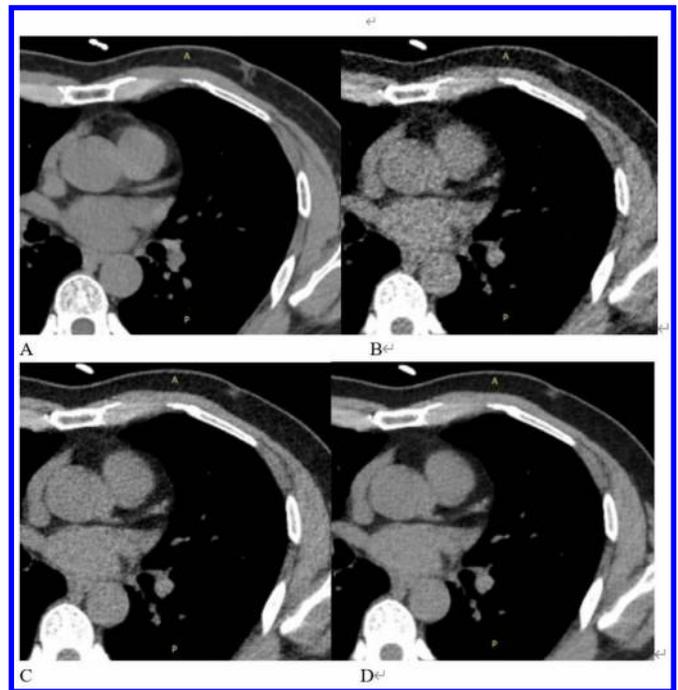
of reconstruction technology.[11] MBIR is a more advanced iterative algorithm than ASIR, using both backward and forward projections. MBIR can reduce image noise more effectively than ASIR, through many complex models, such as system noise model, object model, physics model. Recent studies have shown that MBIR allows significant reduction of radiation dose without affecting image quality and has the potential to further increase the detection rate of some subtle lesions at the expense of longer reconstruction time.[14] ASIR-V is the vendor's third-generation IR algorithm, and replaces its first-generation IR algorithm, ASIR. ASIR-V contains improved noise and object modeling compared with ASIR. But the image noise reduction potential of ASIR-V is lower than that of the MBIR. However, compared with MBIR, the reconstruction time of ASIR-V is substantially reduced, which is one of the major limitations for clinical use of MBIR.[15] Deep learning is a subset of machine learning in artificial intelligence. In general, deep learning consists of massive multilayer networks of artificial neurons. And, the deep convolutional neural networks (DCNN) method, is commonly used in image recognition.[16] The DCNN is trained with virtual low- and high-quality images, the former is obtained with LDCT and

Table 4. Subjective image analysis results with ASIR-V and DLIR under different radiation doses

| Variables	SDCT	LDCT	DLIR-M	DLIR-H	P						
					P	SDCT *vs* DLIR-H	SDCT *vs* DLIR-M	SDCT *vs* LDCT	DLIR-H *vs* DL-M	DLIR-H *vs* LDCT	DLIR-M *vs* LDCT
Solid nodule	4.20 ± 0.40	2.72 ± 0.62	3.31 ± 0.73	4.08 ± 0.49	<0.001	0.43	<0.001	<0.001	<0.001	<0.001	<0.001
Ground-glass nodule	3.70 ± 0.47	2.10 ± 0.31	2.33 ± 0.48	3.47 ± 0.82	<0.001	0.34	<0.001	<0.001	<0.001	<0.001	0.16
Mediastinal tissue	4.12 ± 0.33	2.04 ± 0.20	3.02 ± 0.44	3.85 ± 0.41	<0.001	<0.005	<0.001	<0.001	<0.001	<0.001	<0.001

ASIR, adaptive statistical iterative reconstruction; DLIR, deep learning image reconstruction; LDCT, low-dose CT; SD, standard deviation; SDCT, standard-dose CT. Data are presented as mean ± SD. *p* < 0.05; SDCT: ASIR-V40% with standard-dose; LDCT: ASIR-V40% with low-dose; DLIR-M: DLIR-M with low-dose; DLIR-H: DLIR-H with low-dose.

Table 5. Subjective scoring for the whole image quality of all pulmonary nodules among ASIR-V and DLIR under different radiation doses

Variables	SDCT	LDCT	DLIR-M	DLIR-H	P						
					P	SDCT *vs* DLIR-H	SDCT *vs* DLIR-M	SDCT *vs* LDCT	DLIR-H *vs* DL-M	DLIR-H *vs* LDCT	DLIR-M *vs* LDCT
Morphological display of nodules	4.37 ± 0.49	2.71 ± 0.51	3.37 ± 0.49	4.32 ± 0.52	<0.001	0.66	<0.001	<0.001	<0.001	<0.001	<0.001
Visibility for surrounding lung tissue	4.41 ± 0.50	2.44 ± 0.55	3.24 ± 0.58	4.29 ± 0.56	<0.001	0.32	<0.001	<0.001	<0.001	<0.001	<0.001
Artifacts and diagnostic confidence	4.36 ± 0.49	2.46 ± 0.55	3.34 ± 0.62	4.27 ± 0.50	<0.001	0.41	<0.001	<0.001	<0.001	<0.001	<0.001

ASIR, adaptive statistical iterative reconstruction; DLIR, deep learning image reconstruction; DLIR-H: DLIR-H with low-dose; DLIR-M: DLIR-M with low-dose; LDCT, low-dose CT; SD, standard deviation; SDCT, standard-dose CT; SDCT: ASIR-V40% with standard-dose.

Figure 2. Comparison of chest CT scan in axial soft tissue window images of lung in 42-year-old male. Images were (A), ASIR-V40% at SDCT; (B), ASIR-V40% at LDCT; (C), DLIR-M at LDCT; and (D), DLIR-H at LDCT. There was no significant difference in the detection rate of nodules among the different reconstructions. The difference was in the appearance of the nodules. ASIR, adaptive statistical iterative reconstruction; DLIR, deep learning image reconstruction; LDCT, low-dose CT; SDCT, standard-dose CT; SD, standard deviation.
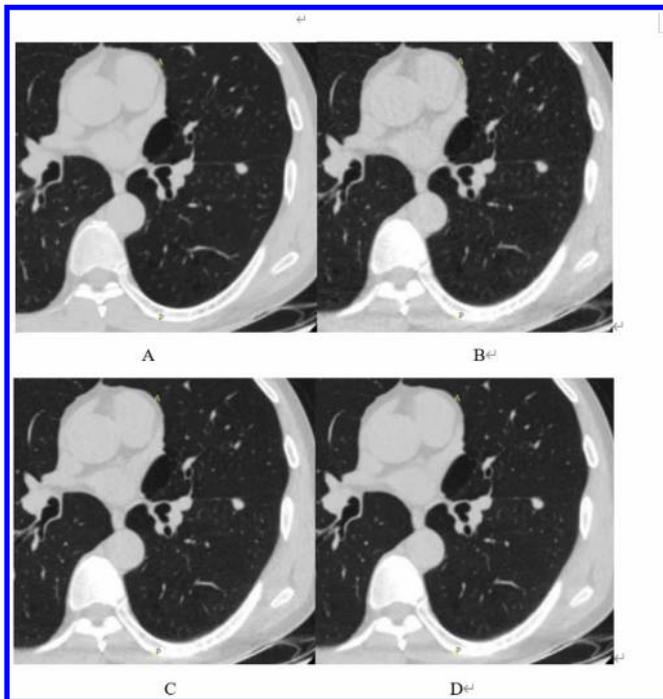
the latter with SDCT. Compared with conventional machine learning methods, a distinctive feature of deep learning is that it can generate appropriate models for tasks directly from the raw data, removing the need for human-led feature extraction, and images reconstructed with DLIR have the property of reduced image noise without blurring.[17,18] The DLIR technique assessed in our study represents a major advancement in the pursuit of CT radiation dose optimization. Studies have shown that compared with both conventional FBP and IR techniques, DLIR-based images delivered better qualitative and quantitative image quality while enabling superior lesion detection ability on chest LDCT[19] and that deep learning approaches offer the exciting potential to more complex image analysis, detect subtle holistic imaging findings and unify methodologies for image evaluation.[20]

In our study, we assessed the use of 80 kV and 50 mA for a low-dose chest CT, which reduced radiation dose by 96% compared with the SDCT. We evaluated the image quality of DLIR chest CT images under such a low-dose scan condition and compared with that of ASIR-V40% images under both the standard- and low-dose conditions. Our study demonstrated that, the attenuation values (CT numbers) in images had no significant difference among different reconstructions (SDCT, LDCT, DLIR-M and DLIR-H); However, under the same low-dose condition, DLIR significantly reduced image noise, resulting in higher CNR compared with ASIR-V40%; In addition, DLIR at the 4% radiation dose level provided similar image quality as ASIR-V40% at the standard dose level.

Our study also indicated that although all lesions could be displayed in all the reconstructions studied, the overall image quality, lesion diagnostic confidence, artifacts, image noise and texture and the details of the lesion among the different image reconstructions varied greatly. DLIR showed improved image quality compared to ASIR-V40% in low-dose chest CT scans. For the images at LDCT, the lesion diagnostic confidence was significantly higher with DLIR-H than with ASIR-V40% or DLIR-M. There was no significant difference in lesion diagnostic confidence between ASIR-V40% at SDCT and DLIR-H at LDCT for solid nodules ($p = 0.43$) and for ground-glass nodules ($p = 0.34$). While there were statistically significant differences between any other reconstruction pairs: DLIR-M@LDCT *vs* ASIR-V40%@LDCT; ASIR-V40%@SDCT *vs* DLIR-M@LDCT; ASIR-V40%@SDCT *vs* ASIR-V40%@LDCT; DLIR-H@LDCT *vs* DLIR-M@LDCT; and DLIR-H@LDCT *vs* ASIR-V40%@LDCT, all $p < 0.05$.

There were limitations in our study. We used a lower tube voltage (at 80 kVp) in LDCT than the 120 kVp in SDCT to dramatically reduce radiation dose. In the future, we plan to investigate the dose saving and image quality improvement potential using the same tube voltage to reduce variables. Another limitation of our study pertains to the fact that we only had small number of patients and we only included non-contrast chest CT examinations. Studies with more patients and contrast-enhanced CT scans need to be carried out in the future to generalize our conclusion.

CONCLUSIONS
In summary, DLIR significantly reduces image noise in low-dose chest CT scans and DLIR-H provides similar image quality as the SDCT ASIR-V40% images with only 4% of the radiation dose.

REFERENCES
1. Ciompi F, Chung K, van Riel SJ, Setio AAA, Gerke PK, et al. Towards automatic pulmonary nodule management in lung cancer screening with deep learning. *Sci Rep* 2017; **7**: 46479. https://doi.org/10.1038/srep46479

2. Jensen CT, Liu X, Tamm EP, Chandler AG, Sun J, et al. Image quality assessment of abdominal ct by use of new deep learning image reconstruction: initial experience. *AJR Am J Roentgenol* 2020; **215**: 50–57. https://doi.org/10.2214/AJR.19.22332

3. Kim JH, Yoon HJ, Lee E, Kim I, Cha YK, et al. Validation of deep-learning image reconstruction for low-dose chest computed tomography scan: emphasis on image quality and noise. *Korean J Radiol* 2021; **22**: 131–38. https://doi.org/10.3348/kjr.2020.0116

4. Greffier J, Hamard A, Pereira F, Barrau C, Pasquier H, et al. Image quality and dose reduction opportunity of deep learning image reconstruction algorithm for ct: a phantom study. *Eur Radiol* 2020; **30**: 3951–59. https://doi.org/10.1007/s00330-020-06724-w

5. Du Y, Li Q, Sidorenkov G, Vonder M, Cai J, et al. Computed tomography screening for early lung cancer, copd and cardiovascular disease in shanghai: rationale and design of a population-based comparative study. *Acad Radiol* 2021; **28**: 36–45. https://doi.org/10.1016/j.acra.2020.01.020

6. Tang H, Liu Z, Hu Z, He T, Li D, et al. Clinical value of a new generation adaptive statistical iterative reconstruction (asir-v) in the diagnosis of pulmonary nodule in low-dose chest ct. *Br J Radiol* 2019; **92**(1103): 20180909. https://doi.org/10.1259/bjr.20180909

7. den Harder AM, de Boer E, Lagerweij SJ, Boomsma MF, Schilham AMR, et al. Emphysema quantification using chest ct: influence of radiation dose reduction and reconstruction technique. *Eur Radiol Exp* 2018; **2**: 30. https://doi.org/10.1186/s41747-018-0064-3

8. Chen L, Jin C, Li J, Wang G, Jia Y, et al. Image quality comparison of two adaptive statistical iterative reconstruction (asir, asir-v) algorithms and filtered back projection in routine liver ct. *BJR* 2018; **91**: 20170655. https://doi.org/10.1259/bjr.20170655

9. Higaki T, Nakamura Y, Zhou J, Yu Z, Nemoto T, et al. Deep learning reconstruction at ct: phantom study of the image characteristics. *Acad Radiol* 2020; **27**: 82–87. https://doi.org/10.1016/j.acra.2019.09.008

10. Benz DC, Benetos G, Rampidis G, von Felten E, Bakula A, et al. Validation of deep-learning image reconstruction for coronary computed tomography angiography: impact on noise, image quality and diagnostic accuracy. *Journal of Cardiovascular Computed Tomography* 2020; **14**: 444–51. https://doi.org/10.1016/j.jcct.2020.01.002

11. Goodenberger MH, Wagner-Bartak NA, Gupta S, Liu X, Yap RQ, et al. Computed tomography image quality evaluation of a new iterative reconstruction algorithm in the abdomen (adaptive statistical iterative reconstruction-v) a comparison with model-based iterative reconstruction, adaptive statistical iterative reconstruction, and filtered back projection reconstructions. *J Comput Assist Tomogr* 2018; **42**: 184–90. https://doi.org/10.1097/RCT.0000000000000666

12. Jensen CT, Wagner-Bartak NA, Vu LN, Liu X, Raval B, et al. Detection of colorectal hepatic metastases is superior at standard radiation dose ct versus reduced dose ct. *Radiology* 2019; **290**: 400–409. https://doi.org/10.1148/radiol.2018181657

13. Lee NK, Kim S, Hong SB, Kim TU, Ryu H, et al. Low-dose ct with the adaptive statistical iterative reconstruction v technique in abdominal organ injury: comparison with routine-dose ct with filtered back projection. *AJR Am J Roentgenol* 2019; **213**: 659–66. https://doi.org/10.2214/AJR.18.20827

14. Benz DC, Fuchs TA, Gräni C, Studer Bruengger AA, Clerc OF, et al. Head-to-head comparison of adaptive statistical and model-based iterative reconstruction algorithms for submillisievert coronary ct angiography. *Eur Heart J Cardiovasc Imaging* 2018; **19**: 193–98. https://doi.org/10.1093/ehjci/jex008

15. Euler A, Solomon J, Marin D, Nelson RC, Samei E, et al. A third-generation adaptive statistical iterative reconstruction technique: phantom study of image noise, spatial resolution, lesion detectability, and dose reduction potential. *AJR Am J Roentgenol* 2018; **210**: 1301–8. https://doi.org/10.2214/AJR.17.19102

16. Nakamura Y, Higaki T, Tatsugami F, Honda Y, Narita K, et al. Possibility of deep learning in medical imaging focusing improvement of computed tomography image quality. *J Comput Assist Tomogr* 2020; **44**: 161–67. https://doi.org/10.1097/RCT.0000000000000928

17. Higaki T, Nakamura Y, Tatsugami F, Nakaura T, Awai K, et al. Improvement of image quality at ct and mri using deep learning. *Jpn J Radiol* 2019; **37**: 73–80. https://doi.org/10.1007/s11604-018-0796-2

18. Lee SM, Seo JB, Yun J, Cho Y-H, Vogel-Claussen J, et al. Deep learning applications in chest radiography and computed tomography: current state of the art. *J Thorac Imaging* 2019; **34**: 75–85. https://doi.org/10.1097/RTI.0000000000000387

19. Singh R, Digumarthy SR, Muse VV, Kambadakone AR, Blake MA, et al. Image quality and lesion detection on deep learning reconstruction and iterative reconstruction of submillisievert chest and abdominal ct. *AJR Am J Roentgenol* 2020; **214**: 566–73. https://doi.org/10.2214/AJR.19.21809

20. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 2019; **25**: 954–61. https://doi.org/10.1038/s41591-019-0447-x

THE ROLE OF IMAGING IN SCREENING SPECIAL FEATURE: FULL PAPER

Screening for lung cancer using sub-millisievert chest CT with iterative reconstruction algorithm: image quality and nodule detectability

[1]MIAO ZHANG, [1]WEIWEI QI, [1]YE SUN, [2]YAN JIANG, [1]XIAOYI LIU and [1]NAN HONG

[1]Department of Radiology, Peking University People's Hospital, Beijing, China
[2]Clinical Science IS, Philips Healthcare, Shanghai, China

Address correspondence to: Dr Nan Hong
E-mail: *hongnan@pkuph.edu.cn*

Objective: To investigate the image quality and nodules detectability using ultra-low dose (ULD) protocol with iterative model reconstruction (IMR) algorithm when compared to routine low dose (LD) chest CT in lung cancer screening.

Methods: Chest CT scans were acquired using a 256-slice scanner for 300 subjects. The scan protocol for the ULD group was 120 kVp/17 mAs while for the LD group was 120 kVp/30 mAs. All images were reconstructed with filtered back projection (FBP), hybrid iterative reconstruction (HIR) and IMR algorithms. Effective dose was recorded. Image quality assessments were performed by two radiologists. SD of CT attenuation was measured as objective image noise. The number of non-calcified nodules detected in both groups with different reconstruction algorithms were calculated and compared.

Results: The effective dose of ULD group (0.67 ± 0.08 mSv) was about 44% reduced compared with LD group (1.20 ± 0.08 mSv) (*p* < 0.01). IMR improved image quality and reduced image noise significantly than HIR and FBP in both groups (all, *p* < 0.01). IMR enabled a higher number of nodule detected compared to FBP and HIR in both LD and ULD groups, especially for solid nodules less than 4 mm.

Conclusion: IMR may improve the diagnostic accuracy of ULD CT lung screening with potential nodule detectability improvement.

Advances in knowledge: IMR enables significant reduction of the image noise and improvement of image quality in sub-mSv (66% reduction) chest scans.

INTRODUCTION

Lung cancer is the leading cause of cancer death worldwide. Its 5-year survival rate after diagnosis is merely 15.6%, despite the advances in surgical, medical and radio-therapeutic treatments.[1] In other words, lung cancer has a good prognosis only if detected at a very early tumour stage. Recently, American National Lung Screening Trial (NLST) has suggested that as compared with chest radiography, low-dose CT (LDCT) screening exhibited more sensitivity in early-stage lung cancer detecting and may contribute to a reduction in mortality from lung cancer,[2,3] which indicates that LDCT has the potential to be widely used in lung cancer screening. However, the LDCT scans performed in NLST involves an approximate dose of 2 mSv, whereas the full-dose chest CT scans, used for nodules follow-up in the major diagnostic, may involve a dose up to 8 mSv, which may present an independent risk of lung cancer and remains a

concern of CT lung screening.[4] Thus, it is valuable to find an approach that can further reduce radiation dose in CT scans meanwhile maintaining the image quality and diagnostic accuracy.

With the development of CT techniques, iterative reconstruction (IR) algorithms were introduced to help reduce the quantum noise associated with the filtered back projection (FBP) algorithm thus to offer better image quality with less radiation dose;[5–7] however, most of the commercially available and widely used IR techniques are hybrid iterative reconstruction (HIR) algorithms which have been reported with certain limits in image noise and artefacts suppression.[8] In recent years, a new model-based IR algorithm, iterative model reconstruction (IMR), has been reported to enable further dose reduction and image quality improvement in chest CT.[9] Thus, we assumed that IMR has the potential to help for further dose reduction in

Table 1.Comparisons of patient characteristics and radiation doses between groups

Characteristics	LD group	ULD group	p
Age in years, mean ± SD	56.4 ± 6.8	55.4 ± 7.2	0.462
Males/females, *n/n*	42/89	42/97	0.196
Body weight in kg, mean ± SD	67.3 ± 8.7	66.2 ± 9.8	0.115
Body mass index in kg m^{-2}, mean ± SD	24.9 ± 2.6	24.6 ± 2.5	0.729
Smokers, *n/N* (%)	42/130 (32)	45/139 (32)	0.891
Passive smokers, *n/N* (%)	69/130 (53)	63/139 (45)	0.989
Occupational expose, *n/N* (%)	13/130 (10)	10/139 (7)	0.549
Effective dose in mSv, mean ± SD	1.20 ± 0.08	0.67 ± 0.08	0.003[a]

LD, low dose; ULD, ultra-low dose.
[a]Significant difference between groups.

LDCT scans used for lung cancer screening, and designed this study to investigate the image quality and nodule detectability of the ultra-low dose (ULD)-CT scans with IMR algorithm by comparing with the routine LDCT scans with FBP and HIR algorithms, to determine whether ULD-CT scans with the use of IMR could achieve diagnostic acceptable in lung cancer screening.

METHODS AND MATERIALS

Study design and population

This prospective study received institutional review board approval; prior informed consent was obtained from all patients. We prospectively enrolled 300 consecutive patients who underwent chest CT during a 4-week period in July and August 2013. All had suspected or confirmed risk of lung cancer. The inclusion criteria were (1) at the age of 40 to 74; (2) Smokers with a smoking history of more than 10 pack-years, including those who had quitted smoking but not more than 10 years; (3) Passive smokers; (4) Occupationally exposed to asbestos, beryllium, uranium and radon. Exclusion criteria included (1) confirmed histologic diagnosis of lung cancer; (2) had previous surgery or radiotherapy in chest; (3) with current respiratory symptoms; (4) pregnancy or lactation status; (5) severe chronic life-threatening disease with a life expectancy less than 6 monthsand (6) body mass index (BMI) larger than 30 kg m^{-2}.

In the first 2 weeks, 150 patients underwent chest CT using routine low dose protocols (LD-group); 20 of them were excluded due

to large BMI and previous surgery in chest and severe chronic disease. In the second 2 weeks, 150 patients underwent CT using a further reduced radiation dose protocol (ULD group) and 11 of them were excluded due to BMI over 30 kg m^{-2} and severe chronic disease.

CT acquisition and image reconstruction

All CT examinations were performed on a 256-slice CT scanner (Brilliance iCT; Philips Healthcare, Cleveland, OH). The data acquisition parameters were as follows: detector configuration, 128 × 0.625 mm; beam pitch, 0.99; rotation time, 0.5 s; field of view, 350 mm; slice thickness, 1.0 mm; slice increment, 0.5 mm, matrix 512*512; tube voltage, 120 kVp; tube current time products, 30 mAs for LD group and 17 mAs for ULD group. Both raw data from LD and ULD groups were reconstructed with FBP, HIR (iDose[4], Level 4, Philips Healthcare, Cleveland, OH) and IMR algorithms, respectively, using identical parameters of 1.0 mm thickness at 0.5 mm increment, and a sharp reconstruction filter (Y-sharp) for lung structures as well as a standard reconstruction filter (B) for mediastinum structures.

Image quality assessment

All images were reviewed and interpreted on a commercially available workstation (Intellispace portal 5.0, Philips Healthcare, Cleveland, OH). Objective image assessment was performed in lung window as follows: A 200 mm^{2}region of interest was placed within the ascending aorta, the CT value (in Hounsfield units) of the region of interest was recorded and its SD was used as image noise. Measurements were performed three times and expressed as the mean value. On the other hand, two thoracic radiologists who were not aware of any image reconstruction settings with 3 and 7 years of experience were asked to perform subjective image assessment independently. Images were displayed in the lung window setting (window width, 1400 HU; window level, −450 HU) and in the mediastinum window setting (window width, 360 HU; window level, 60 HU) for evaluation. The image quality was evaluated for the following structures: lesion margins, visibility of small structures, noise, artefacts and diagnostic confidence. It was determined using a five-point rating scale to image quality (5 = excellent image quality with very good demarcation of structures, noise free; 4 = good image quality with good demarcation of structures, slight increase in noise or artefact; 3 = moderate image quality with reduction of sharpness, moderate increase in noise or artefact; 2 = poor image quality with blurred demarcation of structures, severe increase in noise or artefact; 1 = unassessable). When they disagreed, a third thoracic radiologist with more than 15 years of experience was asked to adjudicate the differences in order to obtain a consensus score.

Table 2.Objective image quality comparison

	LD group			ULD group		
	FBP	iDose[4]	IMR	FBP	iDose[4]	IMR
CT attenuation (HU)	47.6 ± 7.7	47.9 ± 7.3	47.4 ± 7.0	47.6 ± 7.5	46.8 ± 7.0	47.4 ± 7.3
Noise (HU)	62.2 ± 13.1	37.4 ± 7.4	12.3 ± 1.5	82.1 ± 19.5	46.9 ± 9.1	14.4 ± 1.9

FBP, filtered backprojection; HU, Hounsfield units; IMR, iterative model reconstruction; LD, low dose; ULD, ultra-low dose.

Figure 1.Comparison of image noise among LD and ULD groups with different reconstruction algorithms. FBP, filtered backprojection; IMR, iterative model reconstruction; LD, low dose; ULD, ultra-low dose.
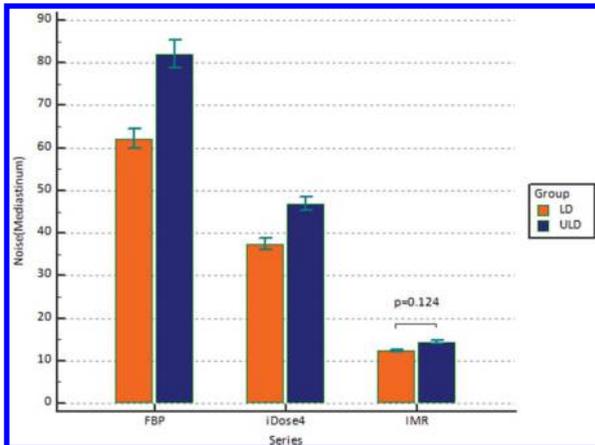
Nodule detection

All non-calcified nodules were recorded and classified as solid and ground-glass opacity categories. The solid nodules were further classified by long-axis diameters in axial plane to three groups (less than 4 mm, 4–8 mm and greater than 8 mm), as well as ground-glass opacity nodules to two groups (less than 5 mm and not less than 5 mm).[10] The number of nodules detected in both LD and ULD images with different reconstruction algorithms were recorded and compared.

Radiation dose analysis

Total dose-length product, which represented the total absorbed dose for all the scans, were recorded from CT dose report. Estimated effective dose was calculated from dose-length product using a conversion factor of 0.014.[11]

Statistical analysis

All continuous values were expressed as mean ± SD. To compare the invariable relationships of the patients' demographic and pathological characteristics between groups, we used χ^2 test when the predictor was categorical and independent t-test when the predictor was quantitative. The objective image noise were compared with ANOVA analysis; if there was a significant difference, pairwise comparisons would be performed with student-Newman-Keuls (SNK) test. The subjective scores were compared by using the Friedman test; if there was a significant difference, pairwise comparisons would be performed with the Steel–Dwass test. Interobserver agreement for subjective image scores was measured using Kappa test. The number of detected nodules

Figure 2. Image quality score for lung (a) and mediastinum (b) structures in LD and ULD groups. Only IMR images enabled a diagnostic acceptable image quality in both groups for both lung and mediastinum structures (red line, score ≥3). FBP, filtered backprojection; IMR, iterative model reconstruction; LD, low dose; ULD, ultra-low dose.
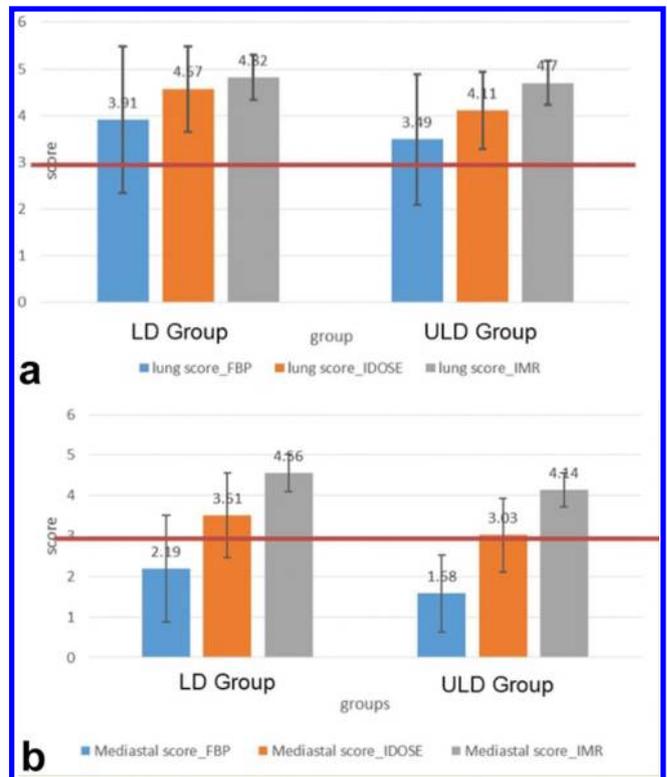
were compared by using χ^2 test. All statistical analyses were performed with commercially available software (SPSS v. 20.0; SPSS Inc, Ill, Excel 2013, Microsoft, Chicago, IL). A value of $p < 0.05$ was considered a statistical significant difference.

RESULTS

Patients demographics and radiation dose

The results of patient demographics and radiation dose are summarized in Table 1. There was no significant difference between the two groups with respect to age, gender, BMI and the clinical characteristics including history of smoking and occupational expose. The effective dose of ULD group was significantly reduced compared to LD group (1.20 mSv ± 0.08, 0.67 mSv ± 0.08, $p = 0.003$).

Table 3.Subjective image quality score

	LD group			ULD group		
	FBP	**iDose^4**	**IMR**	**FBP**	**iDose^4**	**IMR**
Lung	3.91 ± 0.76	4.57 ± 0.52	4.82 ± 0.29	3.42 ± 0.65	4.11 ± 0.47	4.70 ± 0.42
Mediastinum	2.19 ± 0.81	3.51 ± 0.92	4.56 ± 0.58	1.58 ± 0.62	3.03 ± 0.91	4.14 ± 0.69

FBP, filtered backprojection; IMR, iterative model reconstruction; LD, low dose; ULD, ultra-low dose.

Table 4.Number of detected nodules using different algorithms in two dose groups

		SN (<4 mm)	SN (4–8 mm)	SN (>8 mm)	GGN (<5 mm)	GGN (≥5 mm)
LD	FBP	40	39	0	3	11
	iDose[4]	50	49	0	5	13
	IMR	52[a]	51	0	6	13
ULD	FBP	37	31	1	12	31
	iDose[4]	48	38	1	14	31
	IMR	54	36	1	15	32

FBP, filtered backprojection; GGN, ground-glass opacity nodules; IMR, iterative model reconstruction; LD, low dose; SN, solid nodules; ULD, ultra-low dose.
[a]Significant difference in detected nodule numbers compared to FBP using χ^2 test.

Objective image assessment

There was significant difference for all comparison combinations among the different dose groups with different reconstruction algorithms, except LD-IMR *vs* ULD IMR ($p = 0.124$). No difference was found in CT attenuation among all six series ($p = 0.883$). IMR images in both groups showed significant noise reduction compared to FBP and iDose[4]. Details are demonstrated in Table 2 and Figure 1.

Subjective image assessment

There was no significant disagreement between the two radiologists ($\kappa = 0.57–0.87$), except the ULD-FBP images for lung structures ($\kappa = 0.38$). All the subjective image quality scores for each series were summarized in Table 3 and Figure 2. IMR significantly improved subjective image quality compared to iDose[4] and FBP, especially for mediastinum structures in ULD group. There was significant difference in lung structure scores for all comparison combinations among all six series, except for LD-IMR *vs* ULD IMR images, and LD-iDose[4] *vs* ULD-iDose[4] (both, $p > 0.05$). Significant differences were found in mediastinum structure scores for all comparison combinations among the six series.

Nodule detection

The number of nodules detected in both groups with different reconstruction settings were summarized in Table 4. IMR enabled a higher number of nodules detected in both LD and ULD groups for all kinds of nodules, except for solid nodules between 4 and 8 mm. No difference was found in the number of nodules detected among the three algorithms in both groups for different kinds of nodules, except between IMR and FBP in ULD groups for solid nodules less than 4 mm(54 *vs* 37, $p = 0.048$).

Figure 3.Transverse chest CT images of a 52-year-old female (BMI = 18.29 kg m^{-2}) who had two GGOs in the apicoposterior segment of left upper lobe (arrow); the larger one (32 × 26 mm) had a solid component. (a) FBP image of LD group; (b) iDose[4] image of LD group; (c) IMR image of LD group; (d) iDose[4] image of stand-dose (120 kVp/200 mAs) in 15 days later from the same patient. There was no difference of lesion detection among different algorithms (especially for iMR) in low-dose CT and stand-dose CT. BMI, body mass index; FBP, filtered back projection; GGOs, ground-glass opacities; IMR, iterative model reconstruction; LD, low dose.
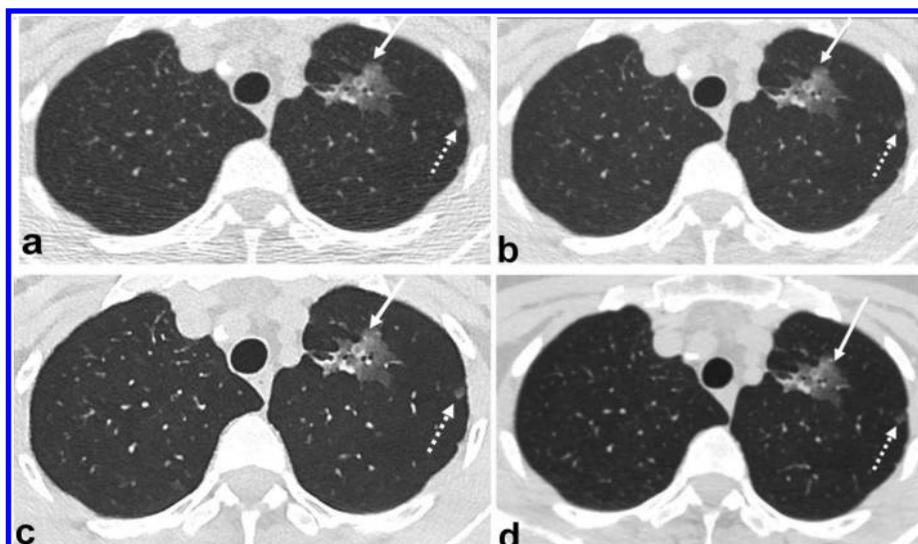
Figure 4. Transverse chest CT images of a 54-year-old female (BMI = 21.48 kg m^{-2}) who had a nodule with calcification in the left lobe of thyroid gland (arrow) in LD group. Images were obtained with (a) FBP, (b) iDose[4] and (c) IMR. The diagnostic confidence of IMR image was much higher than iDose[4] and FBP. The lesion was apparent in IMR image, and available in iDose[4] image, but missed in FBP image. BMI, body mass index; FBP, filtered back projection; IMR, iterative model reconstruction; LD, low dose.
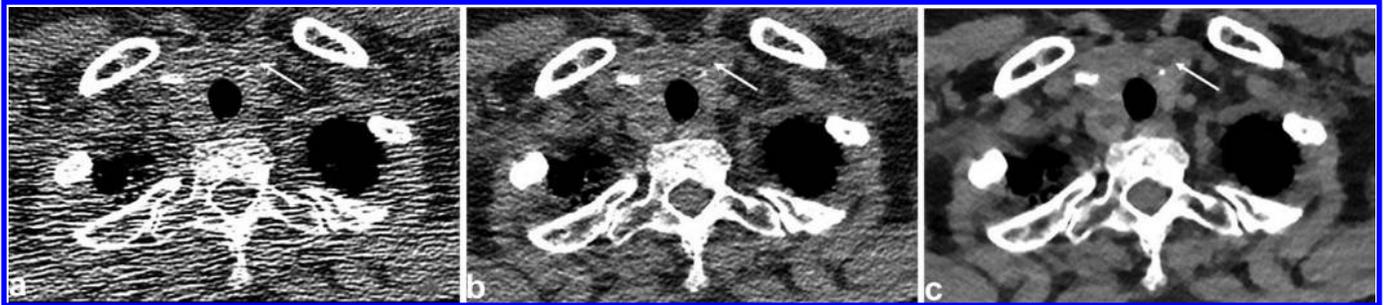
DISCUSSION

The LDCT scans performed in our study involves a further dose reduction of 40% than the screening CT performed in NLST study (1.2 mSv *vs* 2 mSv). Diagnostic image quality of both lung and mediastinum structures were achieved in the LDCT scans with the use of iDose[4] and IMR. IMR offered further objective noise reduction and better subjective image quality scores compared to iDose[4], while FBP exhibited increased noise and failed in diagnostic image quality of mediastinum structures. Nodule detectability was found not significantly associated with reconstruction algorithms in LDCT scans (Figures 3 and 4). Moreover, ULD-CT scans performed in our study reduced the radiation dose to 0.67 mSv, which was approximately 44% further reduction than LDCT. In ULD-CT scans, only IMR enabled diagnostic image quality of both lung and mediastinum structures, neither FBP nor iDose,[4] enabled diagnostic image quality of mediastinum. In addition, IMR was found superior than FBP in nodule detectability for solid nodules less than 4 mm in ULD group (Figures 5 and 6).

To our knowledge, IMR is an advanced IR algorithm that differs from hybrid IR algorithms for its use of system optics to model the acquisition process as accurately as possible in addition to photon and noise statistics.[12] In theory, IMR enables lower image noise and better low-contrast detectability thus to optimize dose protocol further. Previous study[13] demonstrated that with the use of IMR, diagnostic image quality can be achieved on sub-mSv (0.9 mSv) chest scans with even better delineation of lesion margins. Our study observed similar results that IMR enabled superior image quality compared to iDose[4] and FBP, especially for mediastinum structures at ULD scans. Moreover, we observed in ULD scans, iDose[4] enabled diagnostic acceptable image quality in lung but failed in mediastinum, and similarly, in LD scans, FBP enabled diagnostic image quality in lung but failed in mediastinum.

This indicates that it is more difficult for LD scans to achieve diagnostic image quality in mediastinum compared to in lung and further noise reduction is needed for diagnostic image

Figure 5. Transverse chest CT images of a 55-year-old female (BMI = 20.43 kg m^{-2}) who had a GGO (3 × 3 mm) in the apicoposterior segment of left upper lobe (arrow) in the ULD group. (a) FBP image; (b) iDose[4] image; (c) IMR image from the same patient. This lesion was visible on IMR and iDose[4] image, but missed on FBP image in this study. As compared with images a and b image c shows reduced artefacts and higher diagnosis confidence. BMI, body mass index; FBP, filtered back projection; GGO, ground-glass opacity; IMR, iterative model reconstruction; ULD, ultra-low dose.
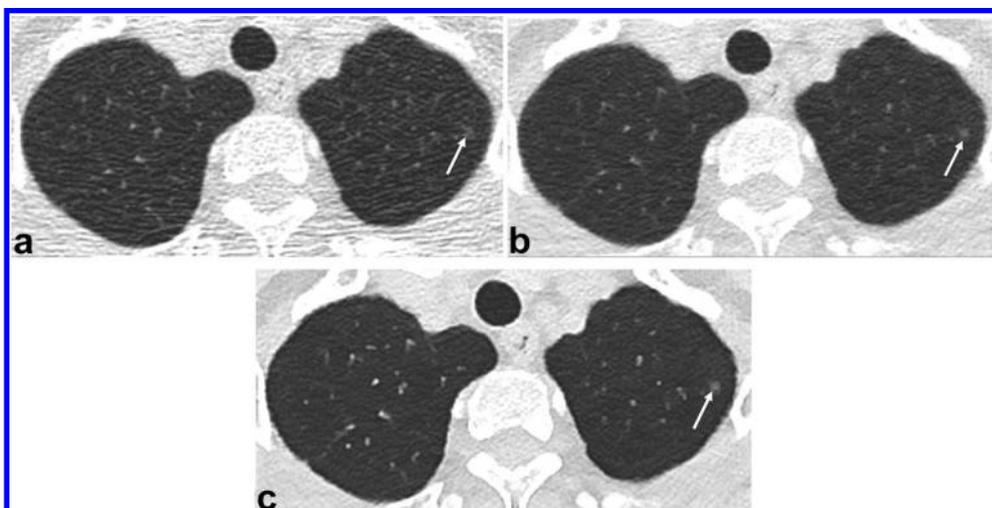
Figure 6. Transverse chest CT through the ascending aorta in a 61-year-old female (BMI = 26.37 kg m^{-2}) with mediastinum lymph node enlargement (arrow). Images were obtained with FBP (a), iDose^4 (b) and IMR (c) in ULD group. Note the excellent depiction of mediastinum lymph node on the IMR image (score 5), compared with FBP (score 3) and iDose^4 (score 4). SNR on IMR image is 3.32 dB, showing higher than those on FBP (0.55 dB) and iDOSE^4 (0.89 dB) in this patient. BMI, body mass index; FBP, filtered back projection; IMR, iterative model reconstruction; SNR, signal-noise ratio; ULD, ultra-low dose.
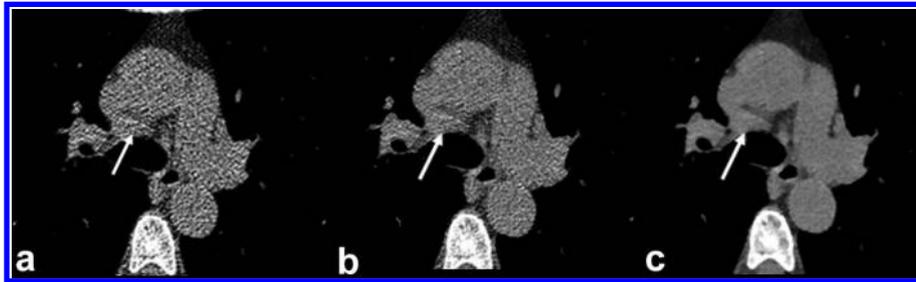
quality of mediastinum images. The main reason could be that it is easier for image quality of mediastinum to be deteriorated by substantial increased image noise at LD conditions, considering there is relatively lower contrast between different tissues in mediastinum as compared to lung.[14]

However, to find suspicious nodules by observing lung structures plays the main role in lung screening chest scans, which indicates that ULD scans with iDose^4 may enable the key demands of lung screening scans by providing diagnostic image quality of lung structures, despite it failed in diagnostic image quality of mediastinum structures. As to the lung nodule detection, we found that there was no significant difference in the number of nodules detected between IMR and iDose^4 in both LD and ULD scans for each kind of nodule, while FBP detected lower number of solid nodules with a diameter less than 4 mm in ULD scans. We attribute this to inadequate image quality of FBP for both lung and mediastinum structures in ULD scans. Moreover, diagnostic information acquired by observation of mediastinum structures such as lymph node and pleural is necessary and in favour of evaluating other complications in lung screening chest scans.[15,16] Hence, it is of practical importance because sub-mSv chest CT with IMR are able to help reduce the risk of radiation exposure without any compromising of diagnostic information including both lung and mediastinum information for patients who undergo lung screening scans.

In addition, it is worth to note that the LDCT protocol combined with iDose^4 was practiced as reference standard instead of full-dose protocol with FBP reconstruction in our study, considering iDose^4, the hybrid IR, has already been used routinely with robust reconstruction speed, as well as the hybrid IRs such as iDose^4, sinogram-affirmed iterative reconstruction, and adaptive statistical iterative reconstruction were observed yielding diagnostic image quality in LD chest CT with similar dose setting at around 1 mSv.[13,17,18]

Our study has several limitations. First, a relatively small number of positive cases were reviewed in the study. Second, the pathological results have not come out for the positive cases. There was no gold standard to verify lesion detection accuracy of all the algorithms. Third, protocols with fixed tube current products were used in our study for stable scan dose; automatic tube current modulation techniques can be used in further study to maintain the image noise at consistent for patients with different sizes. Fourth, the overweight subjects (BMI ≥30 kg m^{-2}) were excluded; further studies will need to investigate the effect of IMR in chest CT on obese patients.

CONCLUSION

In conclusion, both IMR and iDose^4 enables diagnostic image quality in 40% reduction LD chest scans; meanwhile, IMR enables significant reduction of the image noise and improvement of image quality in sub-mSv (66% reduction) chest scans. IMR with significant better image quality may emphasize its potential to better nodule detectability in ULD scans and help the sub-mSv protocols become the clinical routine in lung cancer screening.

FUNDING

REFERENCES

1. American Lung Association. Providing guidance on lung cancer screening to patients and physicians. 2013. Available from: http://www.lung.org/lung-disease/lung-cancer/lung-cancer-screening-guidelines/lung-cancer screening.pdf

2. Church TR, Black WC, Aberle DR, Berg CD, Clingan KL, Duan F, et al. Results of initial low-dose computed tomographic screening for lung cancer. *N Engl J Med* 2013; **368**: 1980–91. doi: https://doi.org/10.1056/NEJMoa1209120

3. Aberle DR, DeMello S, Berg CD, Black WC, Brewer B, Church TR, et al.Results of the two incidence screenings in the national lung screening trial. *N Engl J Med* 2013; **369**: 920–31. doi: https://doi.org/10.1056/NEJMoa1208962

4. McCunney RJ, Li J. Radiation risks in lung cancer screening programs: a comparison with nuclear industry workers and atomic bomb survivors. *Chest* 2014; **145**: 618–24. doi: https://doi.org/10.1378/chest.13-1420

5. Thibault JB, Sauer KD, Bouman CA, Hsieh J. A three-dimensional statistical approach to improved image quality for multislice helical CT. *Med Phys* 2007; **34**: 4526–44. doi: https://doi.org/10.1118/1.2789499

6. Prakash P, Kalra MK, Digumarthy SR, Hsieh J, Pien H, Singh S, et al. Radiation dose reduction with chest computed tomography using adaptive statistical iterative reconstruction technique: initial experience. *J Comput Assist Tomogr* 2010; **34**: 40–5. doi: https://doi.org/10.1097/RCT.0b013e3181b26c67

7. Nelson RC, Feuerlein S, Boll DT. New iterative reconstruction techniques for cardiovascular computed tomography: how do they work, and what are the advantages and disadvantages? *J Cardiovasc Comput Tomogr* 2011; **5**: 286–92. doi: https://doi.org/10.1016/j.jcct.2011.07.001

8. Oda S, Utsunomiya D, Funama Y, Yonenaga K, Namimoto T, Nakaura T, et al. A hybrid iterative reconstruction algorithm that improves the image quality of low-tube-voltage coronary CT angiography. *AJR Am J Roentgenol* 2012; **198**: 1126–31. doi: https://doi.org/10.2214/AJR.11.7117

9. Mehta D, Thompson R, Morton T, Dhanantwari A. Iterative model reconstruction: simultaneously lowered computed tomography radiation dose and improved image quality. *Int J Med Phys* 2013;: 147–54.

10. 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: astatement from the Fleischner Society. *Radiology* 2013; **266**: 304–17. doi: https://doi.org/10.1148/radiol.12120628

11. McCollough C, Cody D and Edyvean S, and American Association of Physicists in Medicine. the measurement, reporting, and management of radiation dose in CT-report of AAPM Task Group 23 of the Diagnostic Imaging Council CT Committee Report No: AAPM report no. 96. 2012.

12. Yuki H, Utsunomiya D, Funama Y, Tokuyasu S, Namimoto T, Hirai T, et al. Value of knowledge-based iterative model reconstruction in low-kV 256-slice coronary CT angiography. *J Cardiovasc Comput Tomogr* 2014; **8**: 115–23. doi: https://doi.org/10.1016/j.jcct.2013.12.010

13. Khawaja RD, Singh S, Gilman M, Sharma A, Do S, Pourjabbar S, et al. Computed tomography (CT) of the chest at less than 1 mSv: an ongoing prospective clinical trial of chest CT at submillisievert radiation doses with iterative model image reconstruction and iDose4 technique. *J Comput Assist Tomogr* 2014; **38**: 613–9. doi: https://doi.org/10.1097/RCT.0000000000000087

14. Willemink MJ, Takx RA, de Jong PA, Budde RP, Bleys RL, Das M, et al. Computed tomography radiation dose reduction: effect of different iterative reconstruction algorithms on image quality. *J Comput Assist Tomogr* 2014; **38**: 815–23. doi: https://doi.org/10.1097/RCT.0000000000000128

15. Lambert L, Banerjee R, Votruba J, El-Lababidi N, Zeman J. Ultra-low-dose CT Imaging of the Thorax: decreasing the radiation dose by one order of magnitude. *Indian J Pediatr* 2016; **83**: 1479–81. doi: https://doi.org/10.1007/s12098-016-2175-2

16. Chen JH, Chan S, Lu NH, Li Y, Tsai YC, Huang PY, et al. Opportunistic breast density assessment in women receiving low-dose chest computed tomography screening. *Acad Radiol* 2016; **23**: 1154–61. doi: https://doi.org/10.1016/j.acra.2016.05.003

17. Yamada Y, Jinzaki M, Tanami Y, Shiomi E, Sugiura H, Abe T, et al. Model-based iterative reconstruction technique for ultralow-dose computed tomography of the lung: a pilot study. *Invest Radiol* 2012; **47**: 482–9. doi: https://doi.org/10.1097/RLI.0b013e3182562a89

18. Minehiro K, Takata T, Hayashi H, Sakuda K, Nunome H, Kawashima H, et al. Phantom study on dose reduction using iterative reconstruction in low-dose computed tomography for lung cancer screening. *Nihon Hoshasen Gijutsu Gakkai Zasshi* 2015; **71**: 1201–8. doi: https://doi.org/10.6009/jjrt.2015_JSRT_71.12.1201

FULL PAPER

A method for evaluating the performance of computer-aided detection of pulmonary nodules in lung cancer CT screening: detection limit for nodule size and density

[1,2]HAJIME KOBAYASHI, MS, [1]MASAKI OHKUBO, PhD, [1]AKIHIRO NARITA, MS, [1,3]JANAKA C MARASINGHE, PhD, [4]KOHEI MURAO, PhD, [5]TORU MATSUMOTO, PhD, [6,7]SHUSUKE SONE, MD, PhD and [1]SHINICHI WADA, PhD

[1]Graduate School of Health Sciences, Niigata University, Niigata, Japan
[2]Department of Radiology, Sannocho Hospital, Niigata, Japan
[3]Department of Radiography and Radiotherapy, Faculty of Allied Health Sciences, University of Peradeniya, Peradeniya, Sri Lanka
[4]Fujitsu Ltd, Tokyo, Japan
[5]Chiba Kensei Hospital, Chiba, Japan
[6]JA Nagano Azumi General Hospital, Nagano, Japan
[7]Present Address: Chest Imaging Division, Nagano Health Promotion Corporation, Nagano, Japan

Address correspondence to: Dr Shinichi Wada
E-mail: *swada@clg.niigata-u.ac.jp*

Objective: We propose the application of virtual nodules to evaluate the performance of computer-aided detection (CAD) of lung nodules in cancer screening using low-dose CT.
Methods: The virtual nodules were generated based on the spatial resolution measured for a CT system used in an institution providing cancer screening and were fused into clinical lung images obtained at that institution, allowing site specificity. First, we validated virtual nodules as an alternative to artificial nodules inserted into a phantom. In addition, we compared the results of CAD analysis between the real nodules ($n = 6$) and the corresponding virtual nodules. Subsequently, virtual nodules of various sizes and contrasts between nodule density and background density (ΔCT) were inserted into clinical images ($n = 10$) and submitted for CAD analysis.

Results: In the validation study, 46 of 48 virtual nodules had the same CAD results as artificial nodules (kappa coefficient = 0.913). Real nodules and the corresponding virtual nodules showed the same CAD results. The detection limits of the tested CAD system were determined in terms of size and density of peripheral lung nodules; we demonstrated that a nodule with a 5-mm diameter was detected when the nodule had a ΔCT > 220 HU.
Conclusion: Virtual nodules are effective in evaluating CAD performance using site-specific scan/reconstruction conditions.
Advances in knowledge: Virtual nodules can be an effective means of evaluating site-specific CAD performance. The methodology for guiding the detection limit for nodule size/density might be a useful evaluation strategy.

INTRODUCTION

Screening by low-dose CT has been shown to reduce mortality from lung cancer in high-risk individuals.[1] Computer-aided detection (CAD) systems are increasingly being used to assist radiologists in the detection of lung nodules. Performance evaluations of CAD systems for the detection of lung nodules have been carried out using clinical images including actual nodules.[2–4] When a CAD system is introduced in an institution providing CT screening, its performance should be evaluated by image data obtained under the same scan and image reconstruction conditions as those used at that site, because of the dependence of CAD performance on scan/reconstruction conditions.[5–7] However, it would be difficult for end users of a CAD system to archive a large

database of CT screenings with sufficient numbers of nodules at each screening site. Public image databases are available,[8] but such images are usually obtained under a limited range of scan/reconstruction conditions, which are not always equivalent across institutions. One way to overcome this limitation is with the use of a lung phantom containing artificial nodules.[9,10] Images of the phantom acquired with the same scan/reconstruction parameters as those used for screening are then subjected to CAD. This approach is effective for assessing site-specific CAD performance. Another method is the use of virtual (computer-simulated) nodules.[11,12] As virtual nodules can be fused into the clinical images obtained in each institution, this approach can also lead to the evaluation of site-specific CAD performance. However, virtual nodules are obtained

by arbitrarily selected modelling and filtering such that the resultant nodules appear to be similar to real nodules (*i.e.* they do not accurately depend on the characteristics of the spatial resolution in the CT system). This is a large disadvantage of virtual nodules compared with artificial nodules.

Another approach to generating nodules has been reported.[13–15] In this method, virtual nodules are computed with a dependence on the spatial resolution characteristics measured for each CT system, which, in turn, are based on the image-generating system itself. Therefore, this method is the most appropriate computational technique for nodule generation. We have explored the feasibility of using this method to assess CAD performance; in this study, the virtual nodules were suggested to be useful for generating free-response receiver-operating characteristic curves for a CAD system.[16] However, in this preliminary study, the validity of the use of virtual nodules as an alternative to artificial nodules included in a lung phantom was not verified. A comparative study between virtual and artificial nodules is necessary. In addition, the use of virtual nodules has the potential advantage of being suitable for assessing the dependence of CAD performance on nodule size and density. This is because of the predetermined (known) size and density of the nodules. When using real nodules in CT images obtained from patients, the true values of the size and density of the nodules are not known. Real nodules have heterogeneous densities and various complicated shapes; therefore, it is difficult to measure their size and density accurately. By using virtual nodules, we are able to clarify the detection limit of the CAD system in terms of nodule size and density. This might lead to a useful evaluation strategy of CAD performance.

In this study, we first validated the application of virtual nodules[13–16] regarding their use for the evaluation of CAD performance in place of artificial nodules included in a lung phantom. Next, as a further validation study, virtual nodules were made to appear to be comparable with real nodules with regard to size and density and were applied to the detection test of the CAD system. Finally, we demonstrated the detection limit of the CAD system in terms of nodule size and density.

METHODS AND MATERIALS
Virtual nodules
We assume that CT image blurring is described by a two-dimensional (2D) point spread function (PSF) in the *x*–*y* scan plane and a slice sensitivity profile (SSP) in the *z* direction perpendicular to the scan plane.[13,17,18] The three-dimensional CT image $I(x,y,z)$ is expressed as

$$I(x, y, z) = [O(x, y, z) ** \text{PSF}(x, y)] * \text{SSP}(z), \quad (1)$$

where $O(x,y,z)$ is the object function and $\text{PSF}(x,y)$ and $\text{SSP}(z)$ are the 2D PSF and SSP, respectively. The operators $*$ and $**$ are one-dimensional and 2D convolutions, respectively. The image simulation based on Equation (1) has been used widely and its validity has been verified.[13–15]

The PSF and SSP are measured in a CT scanner. Object functions are generated numerically as ideal spheres with uniform density

on the assumption of typical solitary pulmonary nodules (Figure 1a). The image $I(x,y,z)$ is calculated from Equation (1) (Figure 1b). Subsequently, the computer-simulated nodule is resampled in three dimensions with intervals equal to the pixel size and slice interval found in clinical CT images (Figure 1c).[19] The resultant image is a virtual nodule that can be fused into practical images by the process shown in Figure 1d and can be used for clinical evaluation.[19,20] In image fusion, the nodule is added into multiple slices to cover the whole virtual nodule in the *z* direction (only the centre slice is shown in Figure 1d).

Validation using artificial nodules in a phantom
Phantom and measurements of point spread function and slice sensitivity profile
We used a commercially available anthropomorphic chest phantom (LSCT001; Kyoto Kagaku Co. Ltd, Kyoto, Japan) (Figure 2). CT images of this phantom are considered to be similar to actual CT images obtained in clinical examinations.[21] The artificial nodules included in the phantom were placed at three levels: the lung apex, the tracheal bifurcation and the lung base. In each slice level, there were artificial nodules with a high density of −630 HU in the left lung and with a low density of −800 HU in the right lung.[21] The background density was −900 HU; then, the contrast between nodule density and background density (ΔCT) was 270 HU for high-contrast nodules and 100 HU for low-contrast nodules. The example image shown in Figure 2 is of the lung base. In this study, we used four high-contrast nodules with diameters of 4 mm, 6 mm, 8 mm and 10 mm and four low-contrast nodules with diameters of 6 mm, 8 mm, 10 mm and 12 mm.

A four-detector row CT scanner (Asteion; Toshiba Medical Systems, Tokyo, Japan) was used for the validation study. The PSF and SSP were measured for the FC50 (for standard lung imaging) kernel and for slice thickness of 8 mm, respectively. The PSF was determined by scanning a high-contrast CT test phantom (MHT-type; Kyoto Kagaku Co. Ltd, Kyoto, Japan); this method includes a means of verifying the obtained PSF and is therefore considered to give accurate values.[22,23] The SSP was measured with the use of a Gold Disk Delta phantom (Kyoto Kagaku Co. Ltd), comprising a gold disc 50-μm thick and 1.0 mm in diameter placed in a tissue-equivalent material (acrylic).

Comparison of virtual nodules with artificial nodules
The lung phantom was scanned at 30 mA, 120 kV, 4 × 5 mm collimation, 0.75 s/rotation and pitch factor of 1.375. A targeted image reconstruction was performed using a field of view (FOV) of 60 mm, slice thickness of 8.0 mm, FC50 reconstruction kernel and matrix size of 512 × 512. We used the image through the middle of the artificial nodules at the slice level of the lung base (Figure 2).

The object functions were generated as ideal spheres whose size and ΔCT corresponded to those of eight artificial nodules at the slice level of the lung base. The virtual nodules obtained from the object functions were compared with the images of the corresponding artificial nodules, and the image differences were quantified by the standard deviation (SD) on the subtracted image.
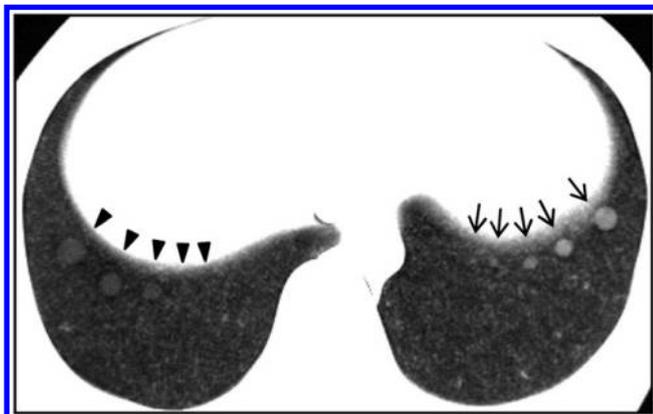
Figure 1. A schematic explanation of virtual nodule generation: (a) the object function of a typical solitary pulmonary nodule with a diameter of 6 mm; (b) a computer-simulated nodule obtained from the object function by Equation (1); (c) a virtual nodule generated by resampling the previous image (b) in three dimensions at clinical CT image resolution; (d) a virtual nodule added to the clinical CT image (arrow).
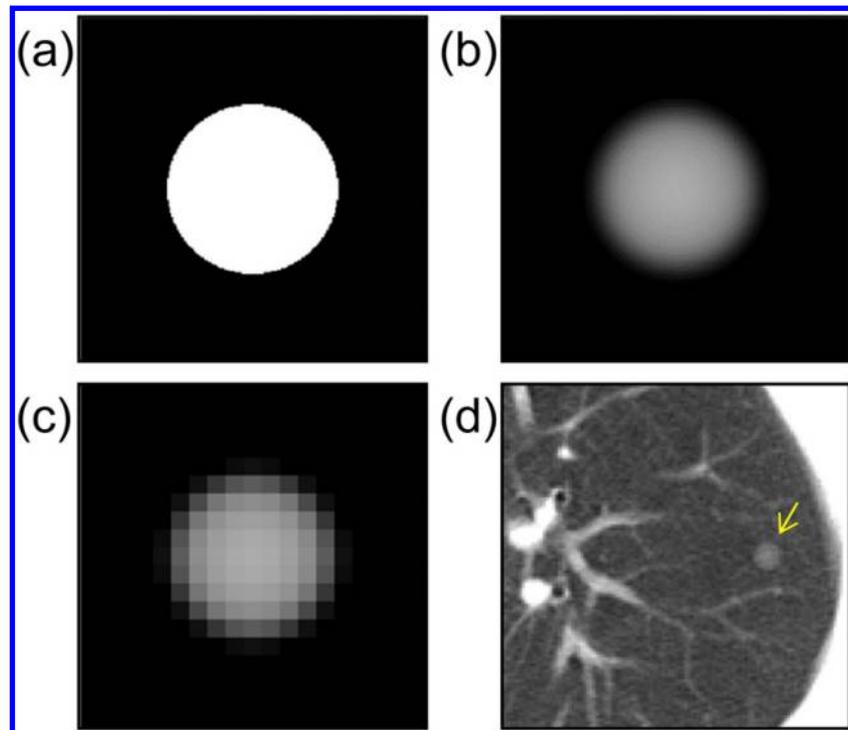
## Comparison of computer-aided detection system detections for virtual nodules and artificial nodules

The lung phantom was scanned and reconstructed with tube currents of 30 mA and 200 mA, a slice thickness of 8.0 mm with an interval of 8.0 mm and an FOV of 320 mm (other conditions were the same as described in the above section). The two tube currents corresponded to different CT scanning conditions and the 320-mm FOV was chosen to mimic a clinical setting for CT screening and CAD analysis. The computer-simulated nodules were generated from object functions having the same size/density as the artificial nodules, as illustrated in Figure 1a,b. The virtual nodules were obtained by resampling the computer-simulated nodules (Figure 1c). The virtual nodules obtained

Figure 2. An image of a chest phantom including artificial nodules: there are five high-contrast nodules in the left lung with diameters of 2 mm, 4 mm, 6 mm, 8 mm and 10 mm (arrows) and five low-contrast nodules in the right lung with diameters of 12 mm, 10 mm, 8 mm, 6 mm and 4 mm (arrowheads). The high-contrast 2-mm nodule and the low-contrast 4-mm nodule were not used in this study because of the difficulty in identifying them in this image.

Figure 3. Virtual nodules added to the phantom image containing artificial nodules: each virtual nodule (arrows) has been placed near the location of the corresponding artificial nodule (arrowheads).
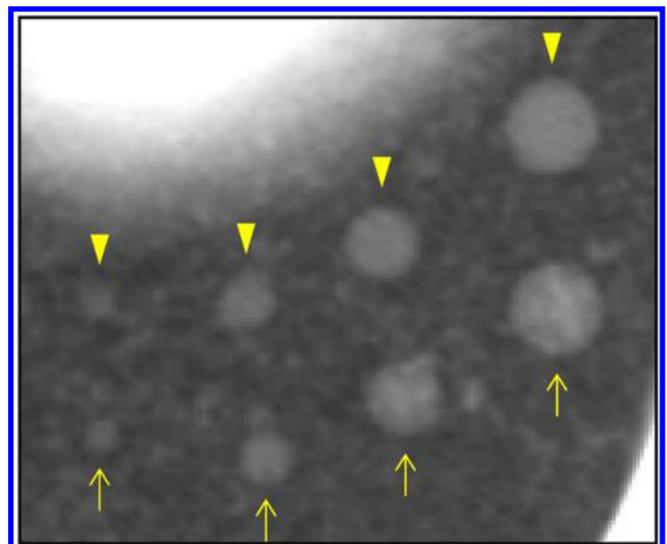
were superimposed onto the phantom images (Figure 3). We chose 24 artificial nodules in the left and right lungs at the 3 slice levels, and we used 2 kinds of phantom images obtained with tube currents of 30 mA and 200 mA; *i.e.* a total of 48 nodules were used. The detection test of the CAD system was performed on the 48 artificial nodules and the corresponding virtual nodules. We applied Cohen's kappa coefficient (κ) to assess the agreement in detection results between the artificial nodules and the virtual nodules using the CAD system. The CAD system used for the detection test was a prototype one developed by our research team,[24] and its development has continued.

Validation with real nodules

An image database archived at an institution providing low-dose CT screening for lung cancer was used. Ethical board approval to access image data was obtained from the institution. Images were obtained with a four-detector row CT scanner (Asteion; Toshiba Medical Systems). This scanner is different from the one described in the above section. The scan was performed at a setting of 30 mA, 120 kV, 4 × 5 mm collimation, 0.75 s/rotation and pitch factor of 1.375. Image reconstruction was performed with a slice thickness of 8.0 mm at an interval of 8.0 mm, an FC50 reconstruction kernel and an FOV of 280–350 mm. The PSF and SSP were measured on this scanner for the kernel and the slice thickness, respectively.

Image data sets in the database were applied to the detection test of the CAD system described above. We chose image data sets of six cases from the database in which a nodular shadow on the image suggested the presence of lung cancer; three of those

Figure 4. Real nodules (arrows) in patient images (a), (c) and (e) are showing computer-aided detection results of true positives (TPs) and corresponding virtual nodules (arrowheads) added into comparable images (b), (d) and (f) of other cases; images were generated using maximum intensity projection of three consecutive sections with the nodule in the centre section. The diameters and contrasts (between nodule density and background density) of object functions used for generating virtual nodules were 6.0 mm and 250 HU (b), 5.5 mm and 350 HU (d) and 5.6 mm and 310 HU (f).
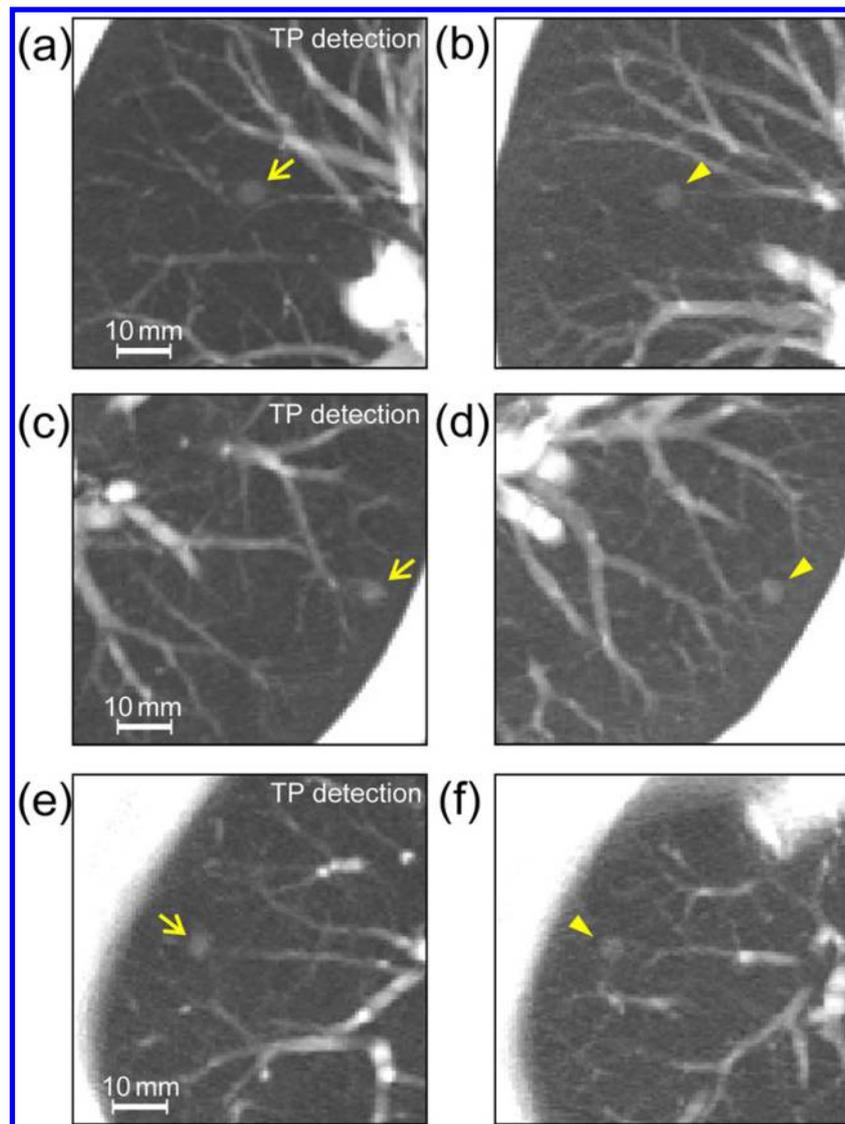
nodules were detected by the CAD system [*i.e.* true positive (TP)] (Figure 4) and the others were not detected [*i.e.* false negative (FN)] (Figure 5). To simulate these nodules, we generated various virtual nodules by changing the diameter and density of the object functions (uniform spheres). Then, we compared the obtained virtual nodules with the real nodules qualitatively in terms of size and density. By visual inspection, we chose virtual nodules that appeared to be most similar to the real nodules overall. Further image data sets of additional cases were selected from the database. In these data sets, the virtual nodules were situated in locations similar to those of the original real nodules with regard to anatomical structures around the nodules, as indicated in Figures 4 and 5. When deciding nodule placement locations, to locate blood vessels existing near

nodules, we used maximum intensity projection images generated from three consecutive sections with the virtual nodule in the middle section. The resulting data sets (plus data sets with subtle changes in positions of the virtual nodules in Figure 5) were submitted for analysis by the CAD system.

Application of virtual nodules to computer-aided detection performance evaluation

The object functions were generated as typical solitary pulmonary nodules of ideal spheres with diameters of 4.0 mm, 4.5 mm, 5.0 mm, 6.0 mm, 7.0 mm and 8.0 mm, and ΔCT was changed from 100 HU to 800 HU in 100-HU increments.[25] A corresponding virtual nodule was obtained from each object function. We used image data sets of 10 cases obtained from the

Figure 5. The analogous comparison of real nodules (left) with corresponding virtual nodules (right) as described in Figure 4, but here with real nodules showing computer-aided detection results of false negatives (FNs): the diameters and contrasts between nodule density and background density of object functions used for generating virtual nodules were 4.5 mm and 360 HU (b), 7.0 mm and 480 HU (d) and 4.3 mm and 320 HU (f).
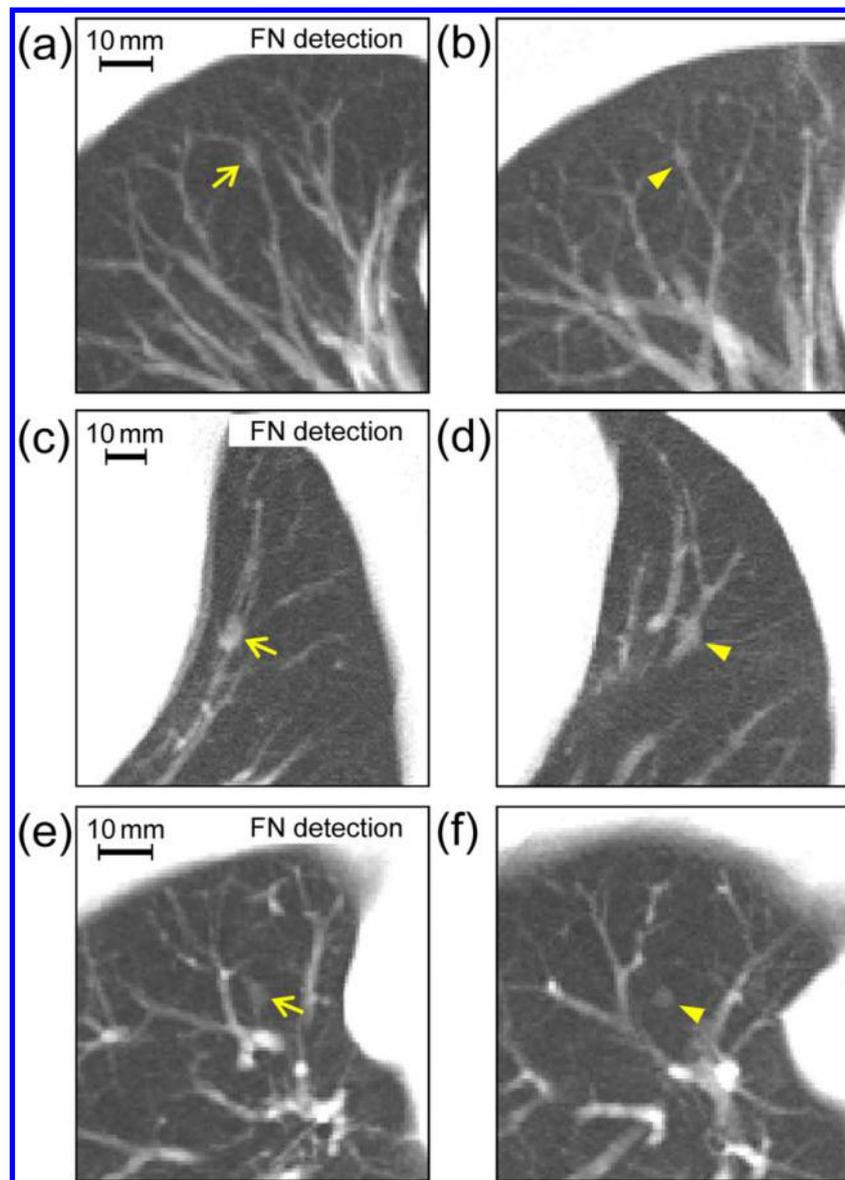
database described in the above section and placed the virtual nodules in the images. Some examples of resultant images are shown in Figure 6. We chose one axial image at the level of the tracheal bifurcation slice in each case, and a virtual nodule obtained with a selected diameter and ΔCT was fused into the image at five locations in the left lung (Figure 6). The three-dimensional virtual nodule occupied multiple 2D slices (only the centre slice is shown in each image in Figure 6). The locations of virtual nodules were selected in the lung periphery so that there was no overlap with large blood vessels, adjacent nodules were sufficiently distant and there was no contact with the lung wall. A total of 48 virtual nodule data sets were obtained per case for the combinations of the 6 diameters and 8 ΔCT values.

We performed CAD system detection in all image data sets. When the total number of TP detections for five nodules with a selected diameter and ΔCT was ≥4 (*i.e.* detection rate ≥80%), the CAD was considered to be able to detect that nodule. In this manner, for all settings of the diameter and ΔCT, we determined the nodule detectability of CAD.
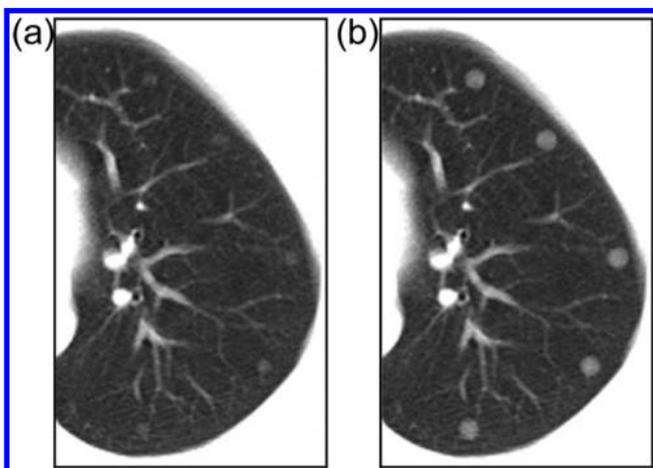
RESULTS

Validation with artificial nodules

Comparison of virtual nodules with artificial nodules
In Figure 7, virtual nodules are compared with the images of high-contrast artificial nodules scanned from the phantom. Subtraction of these images shows little residual intensity around the nodules, except for a component related to noise, artefacts and fine structures in the simulated lung of the phantom (Figure 7c). The equivalent result for low-contrast nodules is obtained (not shown). The SD values measured for high-contrast nodules in the regions of interest on the subtraction image (Figure 7c) are summarized in Table 1, in which these values were averaged. The SD values for the background (measured in the regions of interest indicated in Figure 7a) varied with location in the lung; their mean value is shown in the table. Values were also obtained for low-contrast nodules and summarized in Table 1. The mean

Figure 6. Five virtual nodules have been added to a clinical image at locations in the lung periphery: the object function diameters and contrasts (between nodule density and back-ground density) are (a) 5 mm and 200 HU and (b) 7 mm and 400 HU, respectively.

values of the SD for high-contrast and low-contrast nodules did not exceed those for the corresponding background.

Comparison of computer-aided detection system detections for virtual nodules and artificial nodules
The results of CAD system detections for virtual nodules were compared with those for the corresponding artificial nodules. The total numbers of TP and FN detections for all nodules are summarized in Table 2. For 20 virtual nodules detected by CAD, 18 corresponding artificial nodules were detected. For the 28 virtual nodules not detected by CAD, the corresponding artificial nodules were also not detected. 46 of the 48 virtual nodules showed the same TP and FN detections as those for the artificial nodules, indicating near perfect agreement ($\kappa = 0.913$).

Validation with real nodules

All virtual nodules in Figure 4 were detected by the CAD system, and those in Figure 5 were not detected, indicating the

Figure 7. A comparison of virtual nodules with the phantom image containing high-contrast artificial nodules: (a) artificial nodules—the regions of interest (ROIs), indicated by boxes, were used to determine a representative standard deviation (SD) of the background intensity; (b) virtual nodules are corresponding to artificial nodules; (c) the image obtained by subtracting each nodule image (b) from (a). The SDs of the subtraction residuals were evaluated in the ROIs indicated by the dashed circles. A narrow window width (400 HU) was used for all images.
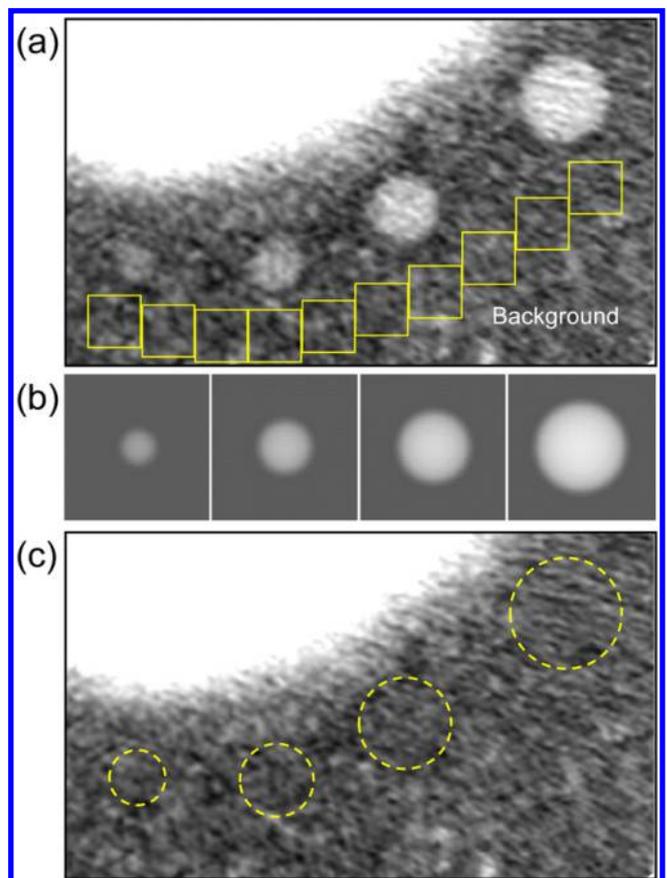
Table 1. Standard deviations (SDs) in the regions of interest (ROIs) on the subtraction image between virtual nodules and artificial nodules, indicated in Figure 7c for high-contrast nodules. SD values measured at various background locations (in the ROIs indicated in Figure 7a) were averaged. These values were also obtained for low-contrast nodules

Region	SD (HU)	
	High-contrast nodules	Low-contrast nodules
Nodule		
12 mm	–	30.8
10 mm	34.4	29.6
8 mm	28.5	26.7
6 mm	31.1	34.6
4 mm	31.9	–
Mean value	31.5	30.5
Background		
Mean value	33.7	32.3

HU, Hounsfield units.

agreement with detection results of corresponding real nodules. For the virtual nodules in Figure 5, the locations of the nodule centres were changed from their initial locations, as shown in Figure 8. When the nodules were repositioned away from the blood vessels (Figure 8a,b), detection results became TP. For the virtual nodules not located close to blood vessels (Figure 8c), changing the location of nodule had no effect on its undetectability (FN).

Evaluation of computer-aided detection performance by virtual nodules

The CAD system detection of virtual nodules in a subject is shown in Figure 9. When the diameter and ΔCT of the object function were set at 5 mm and 200 HU, respectively, only two nodules were detected (Figure 9a). However, when the ΔCT was increased to 300 HU, all five of the virtual nodules were detected (Figure 9b). The summary results for all detections plotted in the diameter–ΔCT space are shown in Figure 9c. The results from Figure 9a,b are denoted as "undetectable" and "detectable" data, respectively, in Figure 9c. When the nodule diameter was increased, the minimum ΔCT required for nodule detection decreased. The solid line delineates the detection limit. The trends were similar for other cases and the overall average detection limit is shown in Figure 10. When the nodule diameter

Table 2. Number of true-positive (TP) detections and false-negative (FN) detections by the computer-aided detection system for virtual nodules and their corresponding artificial nodules (kappa coefficient = 0.913)

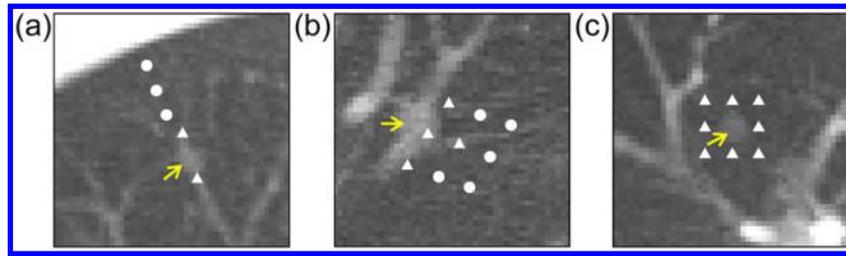| Virtual nodule	Artificial nodule		
	TP	FN	Total
TP	18	2	20
FN	0	28	28
Total	18	30	48

was >5 mm, nodules with a ΔCT more than approximately 220 HU would be detectable with CAD. In addition, the nodule with a diameter of approximately 8 mm was detectable even with a ΔCT as low as approximately 130 HU. For detecting a small nodule 4 mm in diameter, the nodule must have a very high contrast of more than approximately 590 HU. These results for the detection limit in terms of nodule size and density demonstrate the basic performance of the tested CAD system. The mean number of FP detections was 4.8 per case; this value was invariant to the size and density of the virtual nodules.

DISCUSSION

Because CAD performance is affected by scanning and image reconstruction conditions,[5–7] performance evaluation should be implemented using the same site-specific conditions used for lung cancer screening. The virtual nodules proposed in this study were generated from the scan/reconstruction conditions of a particular institution and thus, the CAD performance evaluation intrinsically took into account local factors. The detection limit that was determined for our prototype CAD system, in terms of size and density of peripheral lung nodules (Figure 10), served as a practical guide. As it has been reported that follow-up CT is required for nodules of diameter >5 mm,[26] the evaluation of different detection limits that depend on site-specific scan/reconstruction conditions is an important step towards the clinical use of CAD systems.

The virtual nodule method was validated by comparing virtual nodule images with images of artificial nodules in a lung phantom (Figure 7 and Table 1). In the detection test of the CAD system, the results for virtual nodules showed good agreement with those for artificial nodules (Table 2). There were only two virtual nodules that showed different results from those of artificial nodules. This is attributed to the conditions of the background (*i.e.* simulated lung in the phantom) around the artificial nodules, which were different from those around the fused virtual nodules, as seen in Figure 3. Taking this into account, the agreement ($\kappa = 0.913$) is highly satisfactory. These

Figure 8. Virtual nodules (arrows) not detected by the computer-aided detection (CAD) system: the virtual nodules shown in (a)–(c) are the same as those in Figure 5b,d,f respectively. The locations of the nodule centres were changed from their initial locations to those locations marked with closed triangles "▲" and closed circles "●" in the figure. When nodules were placed on the locations "▲", they were not detected by the CAD system. When nodules were placed on the locations "●", they were detected.

results indicate that virtual nodules can be applied to CAD performance evaluation with similar accuracy to artificial nodules. Moreover, virtual nodules avoid the expense of physical

Figure 9. Results of computer-aided detection (CAD) system detections of virtual nodules fused into clinical lung images: (a) two of the five virtual nodules were detected (indicated by boxes) when the object function was 5 mm in diameter and the contrast between nodule density and background density was (ΔCT) = 200 HU. (b) All five 5-mm virtual nodules were detected when the contrast was ΔCT = 300 HU. (c) The summary of all detection results with the detection limit is indicated by the solid line. When four or more nodules were detected (*i.e.* detection rate ≥80%), the data with the corresponding diameter and ΔCT in the figure are marked as "○" (detectable); otherwise, they are marked as "●" (undetectable).
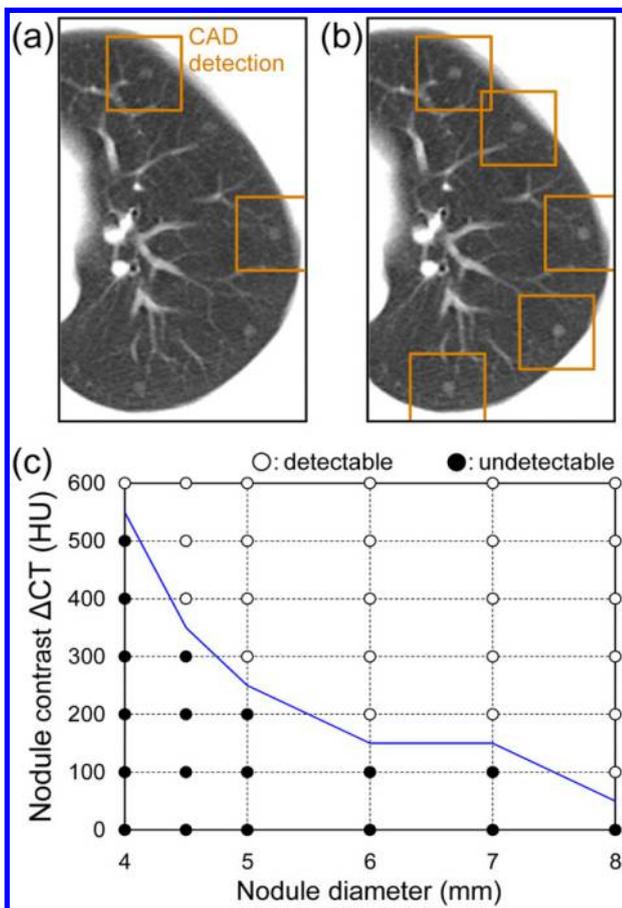
manufacture of artificial nodules and can also be deployed in clinical images. For CAD performance evaluation, virtual nodules present a number of advantages over their artificial counterparts.

Virtual nodules, which were generated and situated to be comparable with real nodules (Figures 4 and 5), generated the same CAD results as the real nodules, further elucidating the dependence of CAD performance on nodule location (Figure 8). This also supports the results depicted in Figure 10. The virtual nodule shown in Figure 5b was generated from an object function having a diameter of 4.5 mm and a ΔCT of 360 HU. Although the nodule was judged to be detectable by the CAD system (Figure 10), it was actually not detected. We surmise that the reason was the nodule proximity to (or overlapping with) a blood vessel; this was evidenced by the result that the nodule was detected by moving it away from the blood vessel (Figure 8a). This explanation also applies to the nodule shown in Figures 5d and 8b (diameter of 7.0 mm and ΔCT of 480 HU). Conversely, the virtual nodule shown in Figure 5f was generated from an object function having a diameter of 4.3 mm and a ΔCT of 320 HU. As shown in Figure 10, this nodule was accurately predicted to be undetectable by the CAD system. In this case, the

Figure 10. The detection limit of the computer-aided detection system averaged over all cases ($n = 10$) (Figure 9c). Error bars are indicating standard deviation. ΔCT, contrast between nodule density and background density.
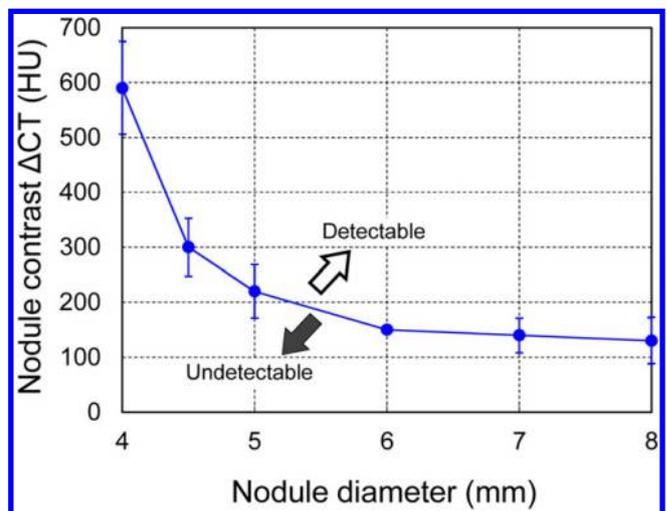
reason for undetectability would be the nodule size and density, not the location; this was evidenced by the result that the nodule remained undetectable after changing its location (Figure 8c). The virtual nodules shown in Figure 4 were accurately predicted to be detectable (Figure 10) as well. Only by using virtual nodules can such detailed analysis about detections by the CAD system be performed. This analysis is not possible when using real nodules.

Our study has some limitations. First, the validation studies were limited by the small number of artificial nodules included in the phantom. And we used only our prototype CAD system throughout our study. A greater number of nodules would be beneficial, as would using additional scan/reconstruction settings and CAD systems. Second, virtual nodules were generated from object functions of ideal spheres with uniform density. For better simulations of real nodules in patients, it is necessary to generate virtual nodules using object functions of heterogeneous density and non-uniform shape. This is theoretically possible, because object functions can be generated numerically with arbitrary shapes and densities. Third, accurate measurement of the spatial resolution in a CT system is essential for generating valid virtual nodules. When scanning a point-source phantom, common methods, such as using a thin wire or a microbead,

have known difficulties in obtaining an accurate PSF.[27] We submit that the 2D PSF measurement method accompanied by verification,[22,23] which was used in the present study, yields sound results for validation.

CONCLUSION

We proposed an application of virtual nodules to evaluate CAD performance using the specific clinical scan/reconstruction conditions of each site. We confirmed that virtual nodules elicited results similar to those of artificial nodules. In addition, the virtual nodules, which were made to be comparable with real nodules, elicited the same CAD results as real nodules, further illustrating the dependence of CAD performance on nodule location. The detection limits of our prototype CAD system were determined in terms of the size and density of peripheral lung nodules; it was demonstrated that a 5-mm nodule was detected when the nodule had a $\Delta CT > 220\,HU$. This methodology of guiding the detection limit might be a useful strategy in the evaluation of CAD performance.

FUNDING

REFERENCES

1. Church TR, Black WC, Aberle DR, Berg CD, Clingan KL, Duan F, et al; National Lung Screening Trial Research Team. Results of initial low-dose computed tomographic screening for lung cancer. *N Engl J Med* 2013; **368**: 1980–91. doi: https://doi.org/10.1056/NEJMoa1209120

2. Armato SG 3rd, Roy AS, Macmahon H, Li F, Doi K, Sone S, et al. Evaluation of automated lung nodule detection on low-dose computed tomography scans from a lung cancer screening program. *Acad Radiol* 2005; **12**: 337–46. doi: https://doi.org/10.1016/j.acra.2004.10.061

3. 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**: 2076–84. doi: https://doi.org/10.1007/s00330-012-2437-y

4. Zinovev D, Duo Y, Raicu DS, Furst J, Armato SG. Consensus *versus* disagreement in imaging research: a case study using the LIDC database. *J Digit Imaging* 2012; **25**: 423–36. doi: https://doi.org/10.1007/s10278-011-9445-3

5. Godoy MC, Kim TJ, White CS, Bogoni L, de Groot P, Florin C, 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**: 74–83. doi: https://doi.org/10.2214/AJR.11.7532

6. Hwang J, Chung MJ, Bae Y, Shin KM, Jeong SY, Lee KS. Computer-aided detection of lung nodules: influence of the image reconstruction kernel for computer-aided detection performance. *J Comput Assist Tomogr* 2010; **34**: 31–4. doi: https://doi.org/10.1097/rct.0b013e3181b5c630

7. 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**: 326–33. doi: https://doi.org/10.1016/j.acra.2007.09.027

8. Armato SG 3rd, 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**: 915–31. doi: https://doi.org/10.1118/1.3528204

9. 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**: e873–8. doi: https://doi.org/10.1016/j.ejrad.2013.08.026

10. Wielpütz MO, Wroblewski J, Lederlin M, Dinkel J, Eichinger M, Koenigkam-Santos M, et al. Computer-aided detection of artificial pulmonary nodules using an *ex vivo* lung phantom: Influence of exposure parameters and iterative reconstruction. *Eur J Radiol* 2015; **84**: 1005–11.

11. Shin HO, Blietz M, Frericks B, Baus S, Savellano D, Galanski M. Insertion of virtual pulmonary nodules in CT data of the chest: development of a software tool. *Eur Radiol* 2006; **16**: 2567–74. doi: https://doi.org/10.1007/s00330-006-0254-x

12. Zhao B, Gamsu G, Ginsberg MS, Jiang L, Schwartz LH. Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm. *J Appl Clin Med Phys* 2003; **4**: 248–60. doi: https://doi.org/10.1120/1.1582411

13. Ohkubo M, Wada S, Kunii M, Matsumoto T, Nishizawa K. Imaging of small spherical structures in CT: simulation study using measured point spread function. *Med Biol Eng Comput* 2008; **46**: 273–82. doi: https://doi.org/10.1007/s11517-007-0283-x

14. Prevrhal S, Fox JC, Shepherd JA, Genant HK. Accuracy of CT-based thickness measurement of thin structures: modeling of limited spatial resolution in all three dimensions. *Med Phys* 2003; **30**: 1–8. doi: https://doi.org/10.1118/1.1521940

15. Rollano-Hijarrubia E, Stokking R, van der Meer F, Niessen WJ. Imaging of small high-density structures in CT: a phantom study. *Acad Radiol* 2006; **13**: 893–908. doi: https://doi.org/10.1016/j.acra.2006.03.009

16. Marasinghe JC, Ohkubo M, Kobayashi H, Murao K, Matsumoto T, Sone S, et al. Feasible method to assess the performance of a lung cancer CT screening CAD system in clinical practice: dependence on nodule size and density. *Int J Med Phys Clin Eng Radiat Oncol* 2014; **3**: 107–16. doi: https://doi.org/10.4236/ijmpcero.2014.32016

17. Ohkubo M, Wada S, Kayugawa A, Matsumoto T, Murao K. Image filtering as an alternative to the application of a different reconstruction kernel in CT imaging: feasibility study in lung cancer screening. *Med Phys* 2011; **38**: 3915–23. doi: https://doi.org/10.1118/1.3590363

18. Polacin A, Kalender WA, Brink J, Vannier MA. Measurement of slice sensitivity profiles in spiral CT. *Med Phys* 1994; **21**: 133–40. doi: https://doi.org/10.1118/1.597251

19. Ohno K, Ohkubo M, Marasinghe JC, Murao K, Matsumoto T, Wada S. Accuracy of lung nodule density on HRCT: analysis by PSF-based image simulation. *J Appl Clin Med Phys* 2012; **13**: 277–92. doi: https://doi.org/10.1120/jacmp.v13i6.3868

20. Funaki A, Ohkubo M, Wada S, Murao K, Matsumoto T, Niizuma S. Application of CT-PSF-based computer-simulated lung nodules for evaluating the accuracy of computer-aided volumetry. *Radiol Phys Technol* 2012; **5**: 166–71. doi: https://doi.org/10.1007/s12194-012-0150-9

21. Muramatsu Y, Tsuda Y, Nakamura Y, Kubo M, Takayama T, Hanai K. The development and use of a chest phantom for optimizing scanning techniques on a variety of low-dose helical computed tomography devices. *J Comput Assist Tomogr* 2003; **27**: 364–74. doi: https://doi.org/10.1097/00004728-200305000-00012

22. Ohkubo M, Wada S, Matsumoto T, Nishizawa K. An effective method to verify line and point spread functions measured in computed tomography. *Med Phys* 2006; **33**: 2757–64. doi: https://doi.org/10.1118/1.2214168

23. Ohkubo M, Wada S, Ida S, Kunii M, Kayugawa A, Matsumoto T, et al. Determination of point spread function in computed tomography accompanied with verification. *Med Phys* 2009; **36**: 2089–97. doi: https://doi.org/10.1118/1.3123762

24. Awai K, Murao K, Ozawa A, Komi M, Hayakawa H, Hori S, et al. Pulmonary nodules at chest CT: effect of computer-aided diagnosis on radiologists' detection performance. *Radiology* 2004; **230**: 347–52. doi: https://doi.org/10.1148/radiol.2302030049

25. Sone S, Hanaoka T, Ogata H, Takayama F, Watanabe T, Haniuda M, et al. Small peripheral lung carcinomas with five-year post-surgical follow-up: assessment by semi-automated volumetric measurement of tumor size, CT value and growth rate on TSCT. *Eur Radiol* 2012; **22**: 104–19. doi: https://doi.org/10.1007/s00330-011-2241-0

26. 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**: 304–17. doi: https://doi.org/10.1148/radiol.12120628

27. Kayugawa A, Ohkubo M, Wada S. Accurate determination of CT point-spread-function with high precision. *J Appl Clin Med Phys* 2013; **14**: 3905. doi: https://doi.org/10.1120/jacmp.v14i4.3905

FULL PAPER

Deep-learning reconstruction for ultra-low-dose lung CT: Volumetric measurement accuracy and reproducibility of artificial ground-glass nodules in a phantom study

[1]RYOJI MIKAYAMA, [1]TAKASHI SHIRASAKA, [1,2]TSUKASA KOJIMA, [1]YUKI SAKAI, [3]HIDETAKE YABUUCHI, [1]MASATOSHI KONDO and [1]TOYOYUKI KATO

[1]Division of Radiology, Department of Medical Technology, Kyushu University Hospital, Fukuoka, Japan
[2]Department of Health Sciences, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan
[3]Department of Health Sciences, Faculty of Medical Sciences, Kyushu University, Fukuoka, Japan

Address correspondence to: Mr ryoji mikayama
E-mail: *mikayama.ryoji.923@m.kyushu-u.ac.jp*

Objectives The lung nodule volume determined by CT is used for nodule diagnoses and monitoring tumor responses to therapy. Increased image noise on low-dose CT degrades the measurement accuracy of the lung nodule volume. We compared the volumetric accuracy among deep-learning reconstruction (DLR), model-based iterative reconstruction (MBIR), and hybrid iterative reconstruction (HIR) at an ultra-low-dose setting.

Methods Artificial ground-glass nodules (6 mm and 10 mm diameters, −660 HU) placed at the lung-apex and the middle-lung field in chest phantom were scanned by 320-row CT with the ultra-low-dose setting of 6.3 mAs. Each scan data set was reconstructed by DLR, MBIR, and HIR. The volumes of nodules were measured semi-automatically, and the absolute percent volumetric error (APEvol) was calculated. The APEvol provided by each reconstruction were compared by the Tukey-Kramer method. Inter- and intraobserver variabilities were evaluated by a Bland-Altman analysis with limits of agreements.

Results DLR provided a lower APEvol compared to MBIR and HIR. The APEvol of DLR (1.36%) was significantly lower than those of the HIR (8.01%, p = 0.0022) and MBIR (7.30%, p = 0.0053) on a 10-mm-diameter middle-lung nodule. DLR showed narrower limits of agreement compared to MBIR and HIR in the inter- and intraobserver agreement of the volumetric measurement.

Conclusions DLR showed higher accuracy compared to MBIR and HIR for the volumetric measurement of artificial ground-glass nodules by ultra-low-dose CT.

Advances in knowledge DLR with ultra-low-dose setting allows a reduction of dose exposure, maintaining accuracy for the volumetry of lung nodule, especially in patients which deserve a long-term follow-up.

INTRODUCTION

The characterization of the properties of pulmonary nodules involves the estimation of nodule volume doubling time on follow-up CT.[1–3] In the CT examinations conducted for the follow-up of lung cancer, computer-aided semi-automated volumetric measurements of pulmonary nodules are used to guide the diagnostic strategy.[4,5] Repeated follow-up CT examinations in clinical practice can provide excessive radiation exposure. Therefore, it is necessary to reduce the radiation exposure dose of CT while maintaining the diagnostic ability. Model-based iterative reconstruction (MBIR) and hybrid iterative reconstruction (HIR) are widely used to reduce radiation exposure and improve diagnostic ability, because these techniques have lower image noise and fewer

artefacts.[6–9] However, the image texture and spatial resolution of these images obtained by MBIR or HIR are inadequate, especially on low-dose CT.[10]

A novel reconstruction algorithm, deep-learning reconstruction (DLR), was recently developed. It incorporates noise and artefact reduction by a deep convolutional neural network.[11,12] The DLR method incorporates a noise and artefact reduction filter with a deep convolutional neural network which is trained with noise-contaminated and noise-free training pairs to extract true signals from noisy images.[11] DLR reduces the image noise and increases the spatial resolution simultaneously, unlike conventional noise reduction methods with trade-offs between spatial

Figure 1. LSCT-001 chest phantom.

resolution and noise reduction.[11] Compared to other traditional reconstruction methods, the DLR technique help improve the image quality, especially at low-dose settings.[13] We have thus speculated that DLR is a suitable reconstruction method for ultra-low-dose CT and that the use of DLR could improve the accuracy of measurements of the volume of pure ground-glass nodules (GGNs) compared to HIR and MBIR. We conducted the present study to compare the accuracy of HIR, MBIR, and DLR in the volumetric measurement of artificial pure GGNs by ultra-low-dose CT.

METHODS AND MATERIALS

Phantom
A chest phantom used for lung cancer screening CT (LSCT-001 chest phantom, Kyoto Kagaku Co. Ltd., Kyoto, Japan) was used[14] (Figure 1). The lung parenchyma of the phantom consists of a combination of urethane foam and styrene resin that shows a CT value of approx. −900 HU. Two sets of spherical artificial pure GGNs (6 mm and 10 mm diameters, −660 HU) were placed at the lung apex level and the middle lung field level of the phantom. Thus, a total of four artificial were placed in the phantom. The volume of the 6-mm-diameter GGN was 113.1 mm^3, and that of the 10-mm-diameter GGN was 523.6 mm^3.

Image acquisition
The chest phantom with artificial pure GGNs was scanned on a 320-row detector CT system (Aquilion ONE/PRISM Edition:

Canon Medical Systems, Otawara, Japan). The following scanning parameters were used: X-ray tube voltage, 120 kV; detector configuration, 0.5 mm × 80 rows, beam pitch, 0.8; gantry rotation time, 0.5 second ; X-ray tube current, 160, 80, and 10 mA. The effective milliampere-second (mAs) values were thus set at 100, 50, and 6.3 mAs. All acquisitions were repeated six times consecutively with shifting the scan start position by 60 degrees each to compensate for random errors in data measurements. Each set of raw data was subjected to HIR (Adaptive Iterative Dose Reduction 3D [AIDR 3D] Standard FC13), MBIR (Forward projected model-based Iterative Reconstruction SoluTion [FIRST] BODY Standard), and DLR (Advanced intelligent Clear-IQ Engine [AiCE] BODY Standard). The lung kernel is generally adopted on detection of lung nodules. However, Wang Y et al reported that the soft tissue kernel was appropriate for accurate quantitative evaluation in lung nodule volume measurement.[15] We thus adopted the soft tissue kernel in this study.

Assessment of radiation doses
Radiation doses of each scan were recorded in terms of the volume CT dose index (CTDIvol [mGy]) and the total dose-length product (DLP [mGy cm]). CTDIvol and DLP values were obtained from the dose page provided from the CT scanner. The effective dose was estimated from the DLP using a normalized effective dose constant of 0.014. The relative percentage compared with the average effective dose of 1.5 mSv in the low-dose screening of National Lung Screening Trial (NLST) in United States were calculated for each tube current.[16]

Volumetric protocols
Two radiological technologists experienced in CT examinations (7 and 10 years experience) performed volumetric measurements for the chest phantom data. A total of 216 nodules (2 size nodules × 2 locations × 3 tube currents × 6 repeated scans × 3 reconstruction algorithms = 216 nodules) were measured, and the measurements of six repeated scans were averaged at each reconstruction algorithm. The volumetric measurements for the artificial pure GGNs were performed independently at the same lung window setting (window level, −600 HU; window width, 1500 HU) using a semi-automated method on CT workstation (Synapse Vincent; Fujifilm Medical, Tokyo, Japan).

In the volumetric measurement, a level of the maximum diameter of an artificial pure GGN was selected and the GGN volume measurement was performed semi-automatically by the CT workstation by drawing its major axis. In clinical practice, the growth rate of lung nodules is managed based on the nodules' volume doubling time, and we therefore defined the absolute percent volumetric error (APEvol) of the artificial pure GGN by the following equation.

$$APEvol = |Vm − Vt| / Vt × 100 \ [\%]$$

where Vm is the measured nodule volume and Vt is the true nodule volume.

The volumes of the artificial pure GGNs once again with the same set 1 month later for the assessment of intraobserver variability.

Figure 2. The Yellow circular regions of interest on 10-mm-diameter GGNs at the middle lung of the chest phantom (a), and the enlarged image (b).
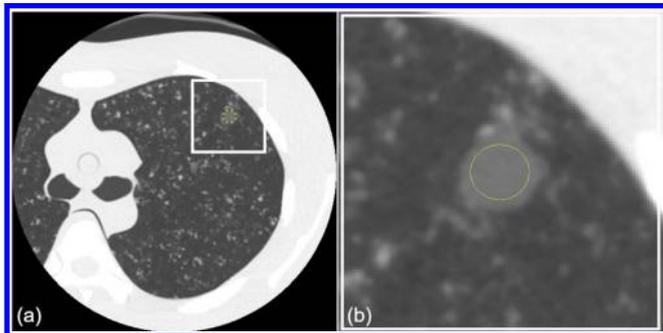

Image noise

The image noise for HIR, MBIR, and DLR at each dose setting were assessed by measuring the standard deviation of the CT values for each corresponding pixel between six consecutive scans. Circular regions of interest (20-pixel-diameter) were placed in the 10-mm-diameter middle lung GGN at the same location in all of the image sets (Figure 2). The average standard deviation value was measured as an estimate of the image noise.

Statistical analyses

An one-way analysis of variance (ANOVA) was used to test the APEvol and image noise data among the three reconstruction methods, followed by post hoc pair-wise comparisons based on the Tukey-Kramer method to account for multiple comparisons. Dunnett's test was performed to assess the differences in the APEvol between MBIR with the low-dose-setting (which is similar to the exposure dose in the NLST) and all reconstruction methods with the ultra-low-dose setting. Statistical significance was accepted at p-values < 0.05.

The inter- and intraobserver agreements for the APEvol data were evaluated using Bland-Altman analyses and intraclass correlation coefficients (ICCs). The interobserver variability was assessed by using the mean results from each observer. In the Bland-Altman analysis, the 95% confidence interval for the limits of agreement was determined.[17] Each ICC was calculated by using two-way random single measures with an absolute agreement condition (ICC[1,2]) for interobserver agreement and one-way random single measures (ICC[1,1]) for intraobserver agreement. Inter- and intraobserver agreement was interpreted in the following way: ICC <0.20 = slight agreement, ICC 0.21–0.40 = fair agreement,

ICC 0.41–0.60 = moderate agreement, ICC 0.61–0.80 = substantial agreement, ICC 0.81–1.0 = perfect agreement. All statistical analyses were performed with the JMP Pro 15.0.0 program (SAS Institute, Cary, NC, USA).

RESULTS

Radiation doses

Table 1 presents the results with the CTDIvol, DLP, ED, and percentage compared with NLST mean ED at each dose settings. The relative percentage doses to the low-dose screening of NLST in United States were 212.8%, 104.5%, and 14.9% at 100, 50, and 6.3 mAs setting, respectively. Kim et al set the percentage of dose reduction compared to the radiation doses of 6.74 mGy at 120 kV, 100 mAs setting.[8] In addition, a radiation dose was compared to the average effective dose of 1.5 mSv in a low-dose scan of NLST.[8,12] We therefore defined the dose of 5.7 mGy at 100 mAs as the standard-dose, 2.8 mGy at 50 mAs as the low-dose, and 0.4 mGy at 6.3 mAs as the ultra-low-dose settings.

Comparisons of APEvol

Images of the artificial pure GGNs scanned at each dose setting are provided as Figure 3. The results of the comparison of APEvol values obtained with the HIR, MBIR, and DLR methods are summarized in Table 2 and illustrated in Figure 4. In all reconstruction methods, the APEvol was larger at low-dose CT compared to standard-dose CT. For the 10-mm artificial pure GGN, DLR showed a smaller APEvol than the other two reconstruction methods in the low-dose scanning.

At 6.3 mAs ultra-low-dose setting, there were significant differences in the APEvol for the 10-mm-diameter middle-lung artificial pure GGN (ANOVA, p = 0.0014) among each reconstructed image. The post hoc analysis showed that the 6.3 mAs ultra-low-dose setting images with DLR (1.4 ± 1.0) had a significantly lower APEvol for the 10-mm-diameter middle-lung GGN compared to those obtained with HIR (8.0 ± 3.8, p = 0.0022) and MBIR (7.3 ± 2.8, p = 0.0053).

At the 50 mAs low-dose setting, there were significant differences in the APEvol values for the 6-mm-diameter middle lung-GGN (ANOVA, p = 0.0106) among each reconstructed image. The post hoc analysis showed that the 50 mAs low-dose setting images with DLR (3.4 ± 1.9) had a significantly lower APEvol for the 6-mm-diameter middle-lung GGN compared to that obtained with HIR (16.3 ± 8.3, p = 0.0079).

Table 1. Descriptive statistics for the radiation dose protocols.

Setting	Tube current-time product (mAs)	CTDIvol (mGy)	DLP (mGy · cm)	ED (mSv)	Percentage compared with NLST mean ED (%)
ultra-low	6.3	0.4	16	0.2	14.9
low	50	2.8	112	1.6	104.5
standard	100	5.7	228	3.2	212.8

CTDIvol, volume CT dose index; DLP, dose-length product; ED, effective dose; NLST, National Lung Screening Trial.
Radiation dose of each CT scan was compared to the average ED of NLST, which was 1.5 mSv.

Figure 3. CT images of the scanned chest phantom and artificial pure ground-glass nodules (GGNs). The lung apex of the chest phantom at the standard dose of 100 mAs with HIR (a) and at the ultra-low-dose of 6.3 mAs with HIR (b), MBIR (c), and DLR (d). The middle lung of the chest phantom at the standard dose of 100 mAs with HIR (e) and at the ultra-low-dose of 6.3 mAs with HIR (f), MBIR (g), and DLR (h). The 10-mm-diameter GGNs at the middle lung of the chest phantom at the standard dose of 100 mAs with HIR (i) and at the ultra low dose of 6.3 mAs with HIR (j), MBIR (k), and DLR (l). The overlay images for volume extraction in semi-automatic measurement of the 10-mm-diameter GGNs at the middle lung of the chest phantom at the standard dose of 100 mAs with HIR (m) and at the ultra-low-dose of 6.3 mAs with HIR (n), MBIR (o), and DLR (p).
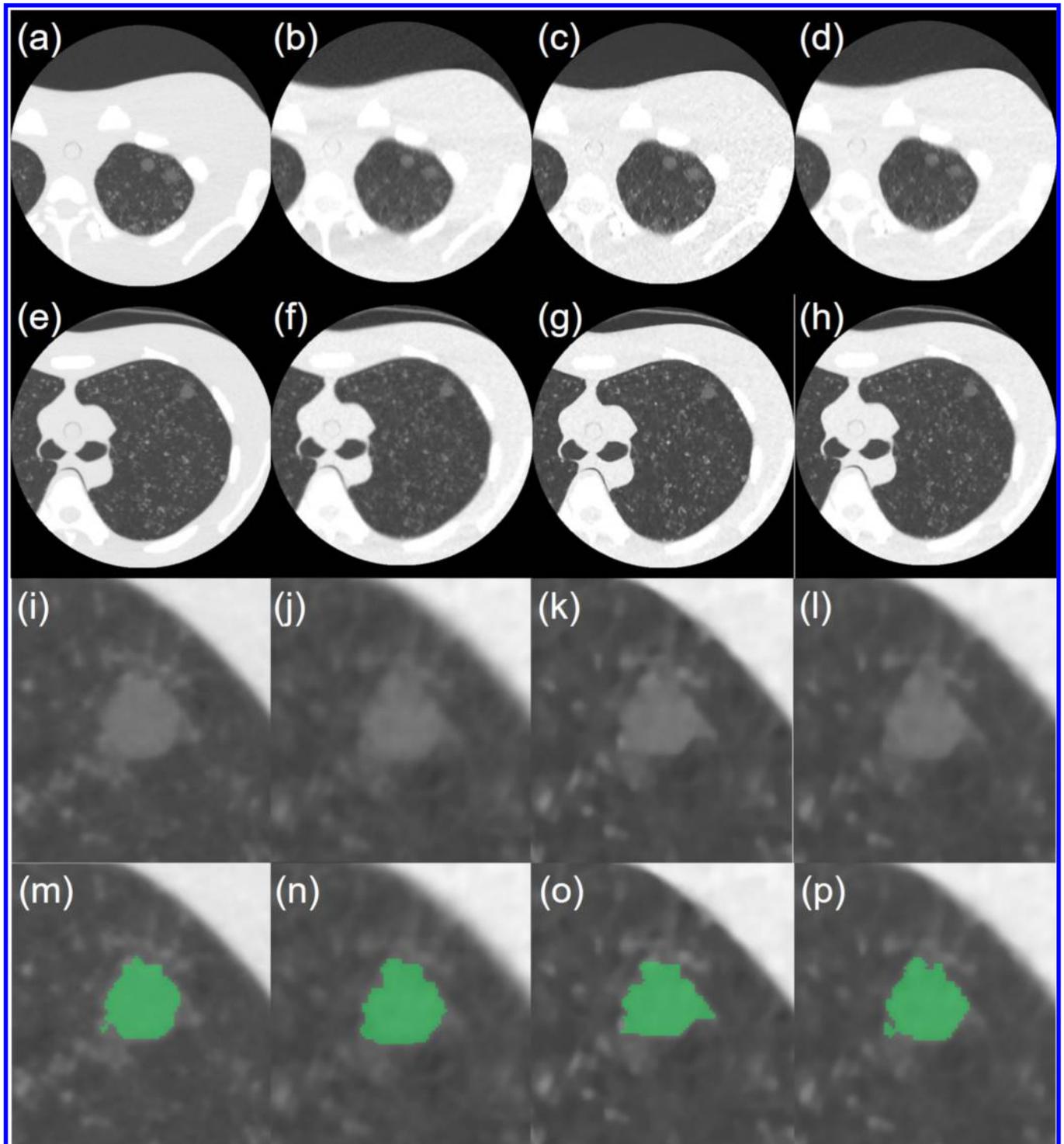
Table 2. Comparison of APEvol values among HIR, MBIR, and DLR at each dose setting for 10 and 6-mm-diameter artificial pure GGNs.

Size	Dose setting	Level of GGN	HIR	MBIR	DLR	One-way ANOVA	Tukey-Kramer test		
			Mean ± SD				HIR *vs* MBIR	HIR *vs* DLR	MBIR *vs* DLR
						p-value	*p*-value	*p*-value	*p*-value
10 mm	Ultra-low-dose 6.3 mAs	Lung apex	9.4 ± 5.9	6.3 ± 2.2	3.7 ± 2.7	0.0755	0.3964	0.0819	0.5929
		Middle lung	8.0 ± 3.8	7.3 ± 2.8	1.4 ± 1.0	0.0014	0.8968	0.0022	0.0053
	Low-dose 50 mAs	Lung apex	3.1 ± 3.2	3.0 ± 2.6	2.9 ± 1.6	0.9822	0.9978	0.981	0.9917
		Middle lung	2.5 ± 2.2	1.9 ± 1.5	1.3 ± 0.8	0.4319	0.7476	0.4008	0.824
	Standard-dose 100 mAs	Lung apex	1.5 ± 1.0	1.4 ± 0.4	2.4 ± 1.1	0.1404	0.9783	0.2278	0.1655
		Middle lung	1.8 ± 1.5	1.1 ± 0.9	1.3 ± 0.7	0.5111	0.4854	0.7544	0.894
6 mm	Ultra-low-dose 6.3 mAs	Lung apex	17.2 ± 3.6	20.3 ± 12.8	12.6 ± 9.1	0.9703	0.8343	0.6795	0.3533
		Middle lung	23.4 ± 11.4	17.3 ± 12.3	9.3 ± 7.1	0.4329	0.5872	0.0838	0.4076
	Low-dose 50 mAs	Lung apex	8.2 ± 5.4	6.8 ± 4.6	6.7 ± 3.7	0.8098	0.8515	0.8289	0.9989
		Middle lung	16.3 ± 8.3	9.5 ± 6.9	3.4 ± 1.9	0.0106	0.1864	0.0079	0.2447
	Standard-dose 100 mAs	Lung apex	7.4 ± 5.7	6.6 ± 6.5	6.9 ± 3.2	0.1651	0.9677	0.9887	0.9944
		Middle lung	5.6 ± 2.3	4.2 ± 3.1	3.7 ± 2.1	0.3814	0.6333	0.4214	0.9292

ANOVA, analysis of variance; DLR, deep learning reconstruction; HIR, Hybrid iterative reconstruction; MBIR, Model-based iterative reconstruction.

Figure 4. The APEvol for the 10 mm-diameter (a), and 6 mm diameter (b) GGNs placed at the lung apex and the middle-lung field level. *$p < 0.05$, Tukey-Kramer method; **$p < 0.05$, Dunnett's test.

(a) Nodule size: 10 mm   (b) Nodule size: 6 mm
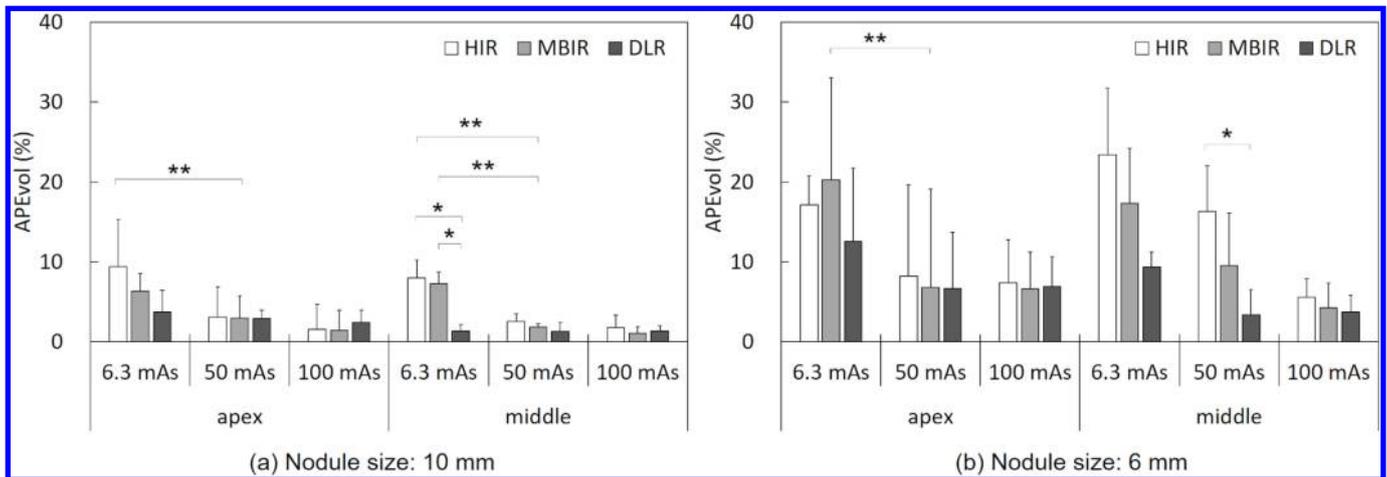
At the 100 mAs standard-dose setting, there was no significant difference in the APEvol for the lung-apex or middle-lung GGNs with 10-and 6 mm diameters among the reconstructed images.

Dunnett's test showed that the ultra-low-dose setting with MBIR or HIR resulted in significantly higher APEvol values for the 10-mm-diameter GGN compared to the low-dose setting with MBIR ($p < 0.05$). For the 6 mm GGN, the ultra-low-dose setting with MBIR provided a significantly higher APEvol compared to the low-dose setting with MBIR ($p < 0.05$). However, there was no significant difference in the APEvol for the 10- or 6-mm GGNs between the ultra-low-dose setting with DLR and the low-dose setting with MBIR.

Variabilities of volumetric measurements

The Bland-Altman plots for the inter- and intraobserver agreement at ultra-low-dose setting regarding volumetric measurements of the 10 mm nodule are shown in Figures 5 and 6, and those for the 6 mm nodule volumetric measurements are shown in Figures 7 and 8. Table 3 presents the results with mean differences and limits

of agreement in the variabilities of the volumetric measurements. Both the intra- and interobserver volumetric measurement agreement results for DLR were better than those for MBIR and HIR, and the corresponding Bland-Altman plots indicated no bias. The mean differences in the inter- and intraobserver agreement ranged from 1.8 to 5.8% and from −0.6 to 2.5% for the 10 mm nodule, and from 3.0 to 9.8% and from 0.2 to 2.3% for the 6 mm nodule, respectively. The mean differences in the intra- and intraobserver agreement results for DLR were the minimum values compared to MBIR and HIR for both nodule sizes. The inter- and intraobserver agreements for the volumetric measurements by DLR showed narrower limits of agreement compared to MBIR and HIR (Figures 5–8). The ICCs for the inter- and intraobserver agreement ranged from 0.03 to 0.63 and from 0.18 to 0.51 for the 10 mm nodule and from 0.12 to 0.54 and from 0.39 to 0.80 for the 6 mm, respectively. The DLR method thus showed a higher ICC for inter- and intraobserver agreement compared to HIR and MBIR.

Image noise

The image noise was compared among the three reconstruction algorithms, and the results are summarized in Table 4. At the

Figure 5. Interobserver variability in the measured volume of the 10 mm nodule by ultra-low-dose CT at 6.3 mAs with HIR (a), MBIR (b), and DLR (c). The variability of DLR was less one-half of that of MBIR. The mean difference is shown by a continuous line; the upper and lower limits of agreement are shown by dashed lines.
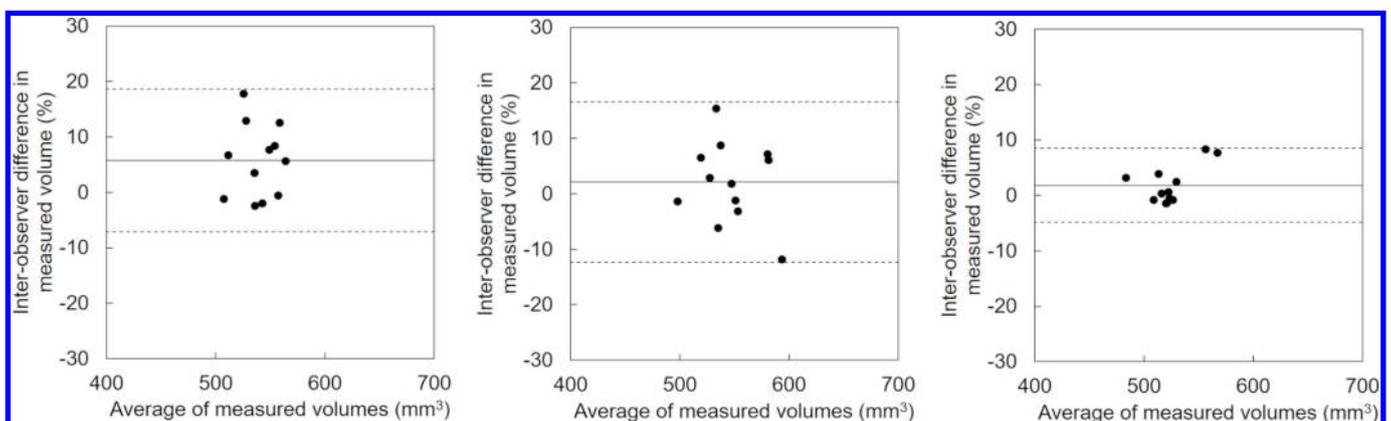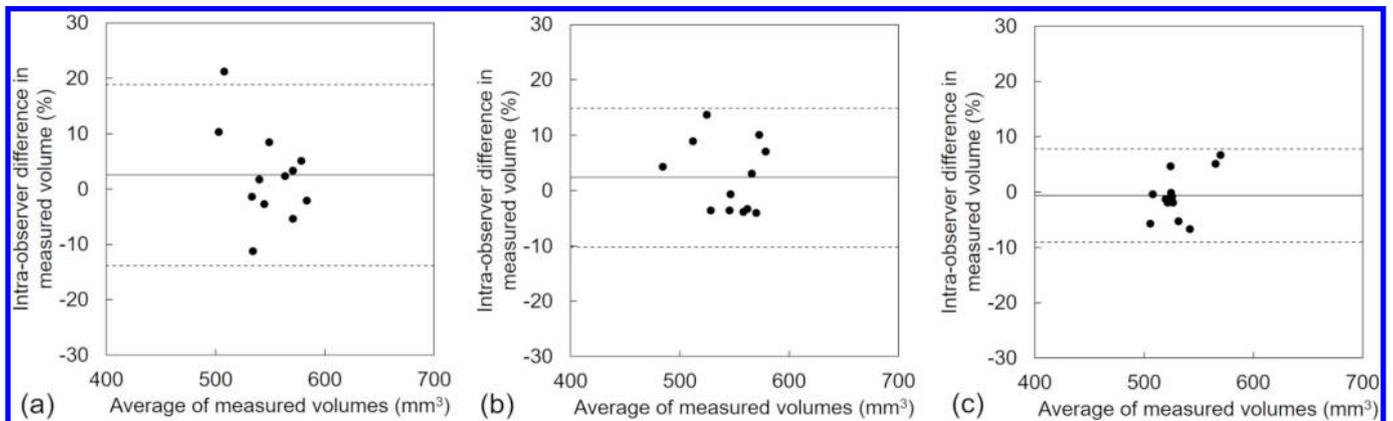
Figure 6. Intraobserver variability in the measured volume of the 10 mm nodule by ultra-low-dose CT at 6.3 mAs with HIR (a), MBIR (b), and DLR (c). The mean difference is shown by a continuous line; the upper and lower limits of agreement are shown by dashed lines.

6.3 mAs ultra-low-dose setting, there was no significant difference in the image noise among the three reconstructed sets of data (ANOVA, $p = 0.1811$). A significant difference was observed at the 50 mAs low-dose setting (ANOVA, $p < 0.0001$); the post hoc analysis showed that the 50 mAs low-dose setting images with HIR (11.1 ± 3.6) had significantly greater image noise compared to those with MBIR (9.2 ± 3.3, $p < 0.0001$) and those with DLR (8.9 ± 3.2, $p < 0.0001$). At the 100 mAs standard-dose setting, another significant difference was observed (ANOVA, $p < 0.0001$); the post hoc analysis showed that the 100 mAs standard setting images with HIR (9.2 ± 2.9) had significantly greater image noise compared to those with MBIR (5.5 ± 1.8, $p < 0.0001$) and DLR (6.9 ± 2.4, $p < 0.0001$). DLR had significantly greater image noise compared to those with MBIR ($p < 0.0001$).

DISCUSSION

Our present findings suggest that unlike other IR algorithms, DLR could be applied to ultra-low-dose CT for the volumetric measurement of lung nodules without compromising measurement reliability. We observed that the semi-automated lung nodule volumetric measurements were significantly affected by the reconstruction algorithms from the aspect of measurement accuracy and reproducibility, especially at low-dose settings.

Our results demonstrated that at ultra-low-dose levels, significantly lower lung nodule volumetric measurement errors could be obtained by DLR compared to MBIR and HIR. At the low- and ultra-low-doses of 50 and 6.3 mAs, DLR provided a lower APEvol compared to MBIR and HIR in artificial pure GGNs with diameters 10 mm and 6 mm. There were significant differences in the APEvol for the 10-mm-diameter nodule at the middle lung between DLR and MBIR (1.4 and 7.3%, $p = 0.0053$) at the ultra-low-dose setting of 6.3 mAs. This may be because MBIR tends to be remarkably degraded due to low-frequency noise, particularly at low radiation dose settings,[18] and because DLR could improve the image contrast between artificial pure GGNs and lung parenchyma due to its superior noise reduction without a loss of spatial resolution.[13,19] We observed that at the standard dose setting, the image noise was lowest on MBIR images, whereas at ultra-low-dose settings, DLR provided the lowest image noise because the low-frequency noise component was effectively reduced.[13] Regarding CT radiation doses, most studies have reported that

Figure 7. Interobserver variability in the measured volume of the 6 mm nodule by ultra-low-dose CT at 6.3 mAs with HIR (a), MBIR (b), and DLR (c). The variability of DLR was less one-half of that of HIR. The mean difference is shown by a continuous line; the upper and lower limits of agreement are shown by dashed lines.
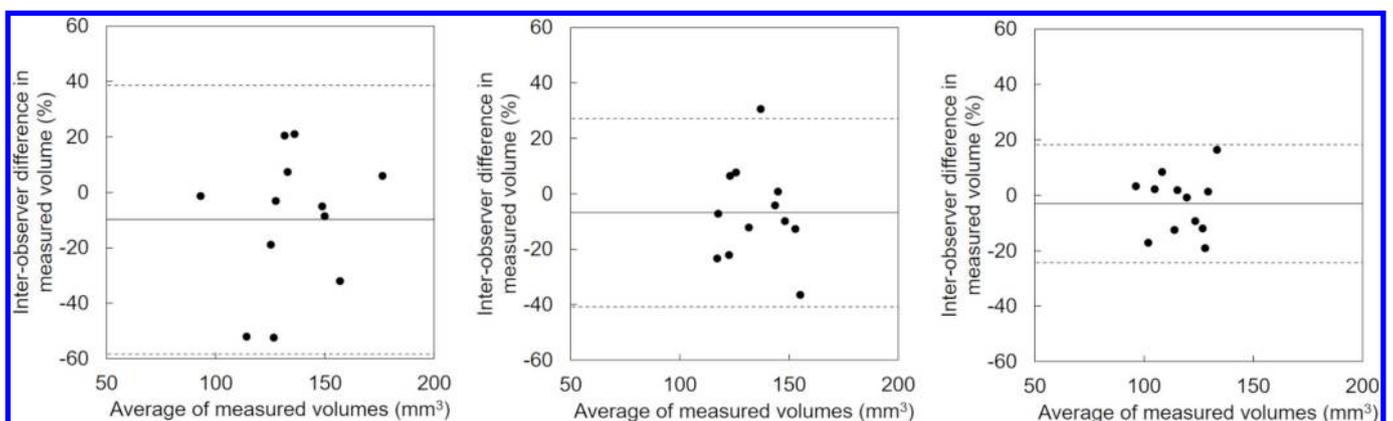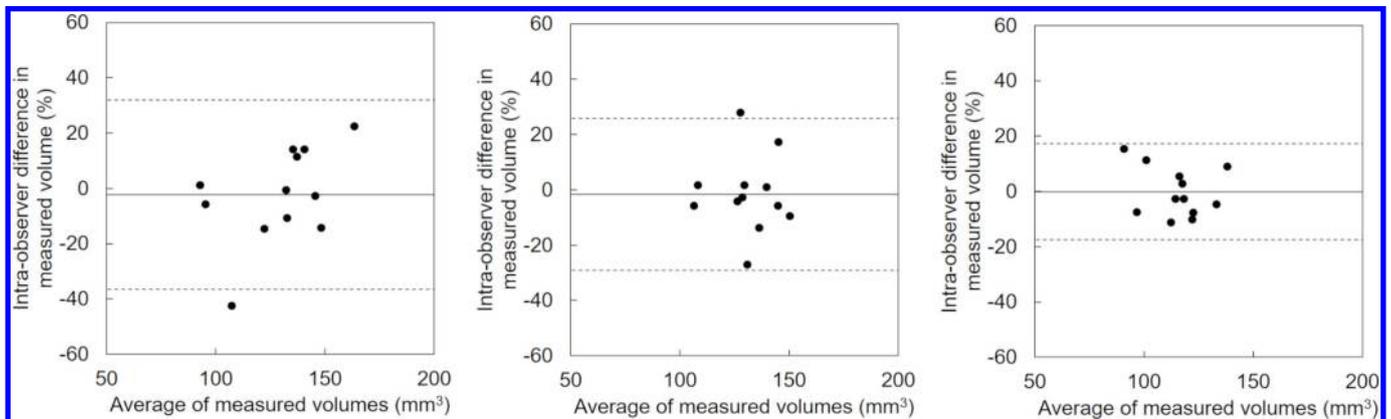
Figure 8. Intraobserver variability in the measured volume of the 6 mm nodule by ultra-low-dose CT at 6.3 mAs with HIR (a), MBIR (b), and DLR (c). The mean difference is shown by a continuous line; the upper and lower limits of agreement are shown by dashed lines.

the accuracy of lung nodule volumetric measurements would decrease in low-dose CT.[8,9,20–22] The excessive noise on a low-dose chest CT image can cause inaccurate results for lung nodule volumetric measurement.[23] DLR is thus a suitable reconstruction method for lung nodule follow-up CT that requires reduction of the cumulative dose due to repeated radiation exposure.

Our present findings demonstrate that DLR contributes to measurement reliability, especially in ultra-low-dose follow-up CT. In our Bland-Altman analysis of the data obtained at the ultra-low-dose setting, DLR showed a mean difference of volumetric measurement <3% in the inter- and intraobserver agreements, and these were smaller compared to those provided by HIR and MBIR. Similarly, the limits of agreement with DLR was less than ±25% and narrower compared to those of MBIR and HIR in the inter- and intraobserver agreements. The lung cancer screening trials commonly use a follow-up CT schedule of third

to fourth/year to evaluate the growth of nodules, based on a minimum increase of 25%.[1,24,25] The DLR algorithm thus enables ultra-low-dose CT screening to confirm lung nodule growth. In addition, the ICCs of DLR showed moderate or substantial agreement in this study, whereas there was only slight inter- and intraobserver agreements for HIR and MBIR. Padole et al noted that the blurring of lung structures would be due to an increased or pronounced blotchy and pixelated appearance related to noise reduction processing on MBIR, and that lung abnormalities would thus be missed on MBIR images at submillisievert dose levels.[26] In addition, the increased low-frequency noise in low-dose images reconstructed by MBIR or HIR induces a degradation of image texture (such as blurry or plastic-like images).[27,28] In semi-automatic volumetric measurements drawing both ends of the lung nodule, the blurring of the contour by MBIR or HIR might induce larger measurement error between observers. Conversely, Higaki et al reported that the noise power spectral

Table 3. Inter- and intra observer agreement of the GGNs' volumes on ultra-low dose CT image reconstructed with each algorithm.

Nodule size	Type of agreement	Type of agreement	Mean difference[a] (limits of agreement, %)	ICC[b]
10 mm	Inter observer	HIR	5.8 (-7.1, 18.7)	0.0307
		MBIR	2.1 (-12.3, 16.6)	0.2999
		DLR	1.8 (-4.9, 8.5)	0.6354
	intra observer	HIR	2.5 (-13.9, 18.9)	0.1834
		MBIR	2.4 (-10.2, 14.9)	0.4436
		DLR	−0.6 (-9.0, 7.8)	0.5143
6 mm	Inter observer	HIR	9.8 (-38.7, 58.3)	0.2763
		MBIR	6.8 (-27.2, 40.7)	0.1241
		DLR	3.0 (-18.3, 24.3)	0.5417
	intra observer	HIR	2.3 (-31.9, 36.5)	0.6049
		MBIR	1.5 (-25.6, 29.0)	0.3935
		DLR	0.2 (-17.2, 17.5)	0.7967

HIR, Hybrid iterative reconstruction; MBIR, Model-based iterative reconstruction; DLR, deep learning reconstruction; ICC, intraclass confidence coefficient.
[a]Bland-Altman analysis.
[b]random single measures with an absolute agreement condition (ICC [1,1], [2, 1]).

Table 4. The image noise provided by HIR, MBIR, and DLR at each dose setting.

Dose setting	HIR	MBIR	DLR	One-way ANOVA	Tukey-Kramer test		
					HIR *vs* MBIR	HIR *vs* DLR	MBIR *vs* DLR
	Mean ± SD			*p*-value	*p*-value	*p*-value	*p*-value
Ultra-low-dose	16.3 ± 5.5	17.1 ± 6.2	16.3 ± 5.8	0.1811	-	-	-
Low-dose	11.1 ± 3.6	9.2 ± 3.3	8.9 ± 3.2	<0.0001	<0.0001	<0.0001	0.7274
Standard-dose	9.2 ± 2.9	5.5 ± 1.8	6.9 ± 2.4	<0.0001	<0.0001	<0.0001	<0.0001

ANOVA, analysis of variance; DLR, deep learning reconstruction; HIR, Hybrid iterative reconstruction; MBIR, Model-based iterative reconstruction.

curve of DLR images was superior especially for low-frequency components because the teaching images were scanned at relatively high radiation doses and reconstructed with MBIR whose parameters were adjusted to obtain the best image quality. We thus speculate that the deterioration of the ultra-low-dose image quality on MBIR and HIR reduced the inter- and intraobserver agreement compared to that on DLR in our present investigation.

DLR has the potential for an 85% reduction in exposure compared to the effective dose used in the NLST. Our present results demonstrate that compared to other reconstruction algorithms, DLR can improve the accuracy of volumetric measurements of pure GGNs and reduce volumetric measurement variability, especially at the ultra-low-dose setting. There was no significant difference between the APEvol of MBIR with the low-dose-setting (which is similar to the exposure dose in the NLST) and that of DLR with the ultra-low-dose setting. Dose reduction is especially important for the follow-up of pure GGNs, because these nodules require repeated CT examinations due to their slower growth compared to solid or part-solid nodules. A further exposure dose reduction would thus be achievable for the volumetric measurement of pure GGNs while maintaining diagnostic accuracy by using the DLR algorithm.

This study has several limitations. It was a phantom study, and the shape and CT value of the targeted artificial pure GGNs was limited to spherical and −660 HU, respectively. These values differ from those of pure GGNs examined *in vivo*. In addition, the number of nodules was limited. We believe that a larger number of nodules with various diameters, attenuations, and margin characteristics are required for a more precise evaluation of measurement feasibility. Second, multiple parameters (noise reduction level) for each reconstruction algorithm were not evaluated. In this study, only one set of parameters for each algorithm was evaluated. We focused on the effect elicited by the difference in the reconstruction algorithm rather than the noise reduction level in the reconstruction parameter. However, it is possible that adjusting the parameters to match the noise and spatial resolution of the three algorithms may provide results that differ from those of our present study. A validation study is necessary to determine whether the noise reduction level for each reconstruction algorithm affects the volumetric measurement of GGNs. Finally, only one software package was used for the semi-automated method, and only one CT scanner of a single vendor was used. Further studies comparing various software packages are needed, and the influence of different CT scanners should be identified.

CONCLUSION

DLR was superior to both MBIR and HIR for the volumetric measurement of pure GGNs by ultra-low-dose CT. Substantial dose reductions might be achieved in lung CT screening for the volumetric measurement of pure GGNs by adapting DLR.

COMPETING INTERESTS

The authors have no conflicts of interest directly relevant to the content of this article.

FUNDING

REFERENCES

1. 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* 3, 2009; **361**: 2221–29. https://doi.org/10.1056/NEJMoa0906085

2. de Margerie-Mellon C, Ngo LH, Gill RR, Monteiro Filho AC, Heidinger BH, Onken A, et al. The growth rate of subsolid lung adenocarcinoma nodules at chest ct. *Radiology* October 2020; **297**: 189–98. https://doi.org/10.1148/radiol.2020192322

3. Park S, Lee SM, Kim S, Lee J-G, Choi S, Do K-H, et al. Volume doubling times of lung adenocarcinomas: correlation with predominant histologic subtypes and prognosis. *Radiology* June 2020; **295**: 703–12. https://doi.org/10.1148/radiol.2020191835

4. de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, et al. Reduced lung-cancer mortality with volume ct screening in a randomized trial. *N Engl J Med* February 6, 2020; **382**: 503–13. https://doi.org/10.1056/NEJMoa1911793

5. Silva M, Milanese G, Sestini S, Sabia F, Jacobs C, van Ginneken B, et al. Lung cancer screening by nodule volume in lung-rads v1.1: negative baseline ct yields potential for increased screening interval. *Eur Radiol* 2021; **31**: 1956–68. https://doi.org/10.1007/s00330-020-07275-w

6. Ohno Y, Yaguchi A, Okazaki T, Aoyagi K, Yamagata H, Sugihara N, et al. Comparative evaluation of newly developed model-based and commercially available hybrid-type iterative reconstruction methods and filter back projection method in terms of accuracy of computer-aided volumetry (cadv) for low-dose ct protocols in phantom study. *Eur J Radiol* 2016; **85**: 1375–82. https://doi.org/10.1016/j.ejrad.2016.05.001

7. Katsura M, Matsuda I, Akahane M, Yasaka K, Hanaoka S, Akai H, et al. Model-based iterative reconstruction technique for ultralow-dose chest ct: comparison of pulmonary nodule detectability with the adaptive statistical iterative reconstruction technique. *Invest Radiol* 2013; **48**: 206–12. https://doi.org/10.1097/RLI.0b013e31827efc3a

8. Kim H, Park CM, Song YS, Lee SM, Goo JM. Influence of radiation dose and iterative reconstruction algorithms for measurement accuracy and reproducibility of pulmonary nodule volumetry: a phantom study. *Eur J Radiol* 2014; **83**: 848–57. https://doi.org/10.1016/j.ejrad.2014.01.025

9. Sakai N, Yabuuchi H, Kondo M, Kojima T, Nagatomo K, Kawanami S, et al. Volumetric measurement of artificial pure ground-glass nodules at low-dose ct: comparisons between hybrid iterative reconstruction and filtered back projection. *Eur J Radiol* 2015; **84**: 2654–62. https://doi.org/10.1016/j.ejrad.2015.08.018

10. Millon D, Vlassenbroek A, Van Maanen AG, Cambier SE, Coche EE. Low contrast detectability and spatial resolution with model-based iterative reconstructions of mdct images: a phantom and cadaveric study. *Eur Radiol* 2017; **27**: 927–37. https://doi.org/10.1007/s00330-016-4444-x

11. Tatsugami F, Higaki T, Nakamura Y, Yu Z, Zhou J, Lu Y, et al. Deep learning-based image restoration algorithm for coronary ct angiography. *Eur Radiol* 2019; **29**: 5322–29. https://doi.org/10.1007/s00330-019-06183-y

12. Kang E, Min J, Ye JC. A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction. *Med Phys* 2017; **44**: e360–75. https://doi.org/10.1002/mp.12344

13. Higaki T, Nakamura Y, Zhou J, Yu Z, Nemoto T, Tatsugami F, et al. Deep learning reconstruction at ct: phantom study of the image characteristics. *Acad Radiol* 2020; **27**: 82–87. https://doi.org/10.1016/j.acra.2019.09.008

14. Muramatsu Y, Tsuda Y, Nakamura Y, Kubo M, Takayama T, Hanai K. The development and use of a chest phantom for optimizing scanning techniques on a variety of low-dose helical computed tomography devices. *J Comput Assist Tomogr* 2003; **27**: 364–74. https://doi.org/10.1097/00004728-200305000-00012

15. Wang Y, de Bock GH, van Klaveren RJ, van Ooyen P, Tukker W, Zhao Y, et al. Volumetric measurement of pulmonary nodules at low-dose chest ct: effect of reconstruction setting on measurement variability. *Eur Radiol* 2010; **20**: 1180–87. https://doi.org/10.1007/s00330-009-1634-9

16. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 4, 2011; **365**: 395–409. https://doi.org/10.1056/NEJMoa1102873

17. Martin Bland J, Altman D. STATISTICAL methods for assessing agreement between two methods of clinical measurement. *The Lancet* 1986; **327**: 307–10. https://doi.org/10.1016/S0140-6736(86)90837-8

18. Racine D, Ba AH, Ott JG, Bochud FO, Verdun FR. Objective assessment of low contrast detectability in computed tomography with channelized hotelling observer. *Phys Med* 2016; **32**: 76–83. https://doi.org/10.1016/j.ejmp.2015.09.011

19. Akagi M, Nakamura Y, Higaki T, Narita K, Honda Y, Zhou J, et al. Deep learning reconstruction improves image quality of abdominal ultra-high-resolution ct. *Eur Radiol* 2019; **29**: 6163–71. https://doi.org/10.1007/s00330-019-06170-3

20. Xu Y, He W, Chen H, Hu Z, Li J, Zhang T. Impact of the adaptive statistical iterative reconstruction technique on image quality in ultra-low-dose ct. *Clin Radiol* 2013; **68**: 902–8. https://doi.org/10.1016/j.crad.2013.03.024

21. Chen B, Barnhart H, Richard S, Robins M, Colsher J, Samei E. Volumetric quantification of lung nodules in ct with iterative reconstruction (asir and mbir). *Med Phys* 2013; **40**(11): 111902. https://doi.org/10.1118/1.4823463

22. 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**: 864–70. https://doi.org/10.1148/radiol.2283020059

23. Scholten ET, Jacobs C, van Ginneken B, Willemink MJ, Kuhnigk J-M, van Ooijen PMA, et al. Computer-aided segmentation and volumetry of artificial ground-glass nodules at chest ct. *AJR Am J Roentgenol* 2013; **201**: 295–300. https://doi.org/10.2214/AJR.12.9640

24. Wang Y, van Klaveren RJ, van der Zaag-Loonen HJ, de Bock GH, Gietema HA, Xu DM, et al. Effect of nodule characteristics on variability of semiautomated volume measurements in pulmonary nodules detected in a lung cancer screening program. *Radiology* 2008; **248**: 625–31. https://doi.org/10.1148/radiol.2482070957

25. Rampinelli C, De Fiori E, Raimondi S, Veronesi G, Bellomi M. In vivo repeatability of automated volume calculations of small pulmonary nodules with ct. *AJR Am J Roentgenol* 2009; **192**: 1657–61. https://doi.org/10.2214/AJR.08.1825

26. Padole A, Singh S, Ackman JB, Wu C, Do S, Pourjabbar S, et al. Submillisievert chest ct with filtered back projection and iterative reconstruction techniques. *AJR Am J Roentgenol* 2014; **203**: 772–81. https://doi.org/10.2214/AJR.13.12312

27. Ehman EC, Yu L, Manduca A, Hara AK, Shiung MM, Jondal D, et al. Methods for clinical evaluation of noise reduction techniques in abdominopelvic ct. *Radiographics* 2014; **34**: 849–62. https://doi.org/10.1148/rg.344135128

28. Geyer LL, Schoepf UJ, Meinel FG, Nance JW Jr, Bastarrika G, Leipsic JA, et al. State of the art: iterative ct reconstruction techniques. *Radiology* 2015; **276**: 339–57. https://doi.org/10.1148/radiol.2015132766