Ultra-Low Dose Chest CT with Denoising for Lung Nodule Detection

Ariel Kerpel MD1,2*, Edith Michelle Marom MD1,2*, Michael Green PhD1,2, Michal Eifer MD1,2, Eli Konen MD1,2, Arnaldo Mayer PhD1,2, and Sonia L. Betancourt Cuellar MD3

1Department of Diagnostic Radiology, Sheba Medical Center, Tel Hashomer, Israel
2Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
3Department of Diagnostic Radiology, Division of Diagnostic Imaging, University of Texas MD Anderson Cancer Center, Houston, Texas, USA

ABSTRACT

Background: Medical imaging and the resultant ionizing radiation exposure is a public concern due to the possible risk of cancer induction.

Objectives: To assess the accuracy of ultra-low-dose (ULD) chest computed tomography (CT) with denoising versus normal dose (ND) chest CT using the Lung CT Screening Reporting and Data System (Lung-RADS).

Methods: This prospective single-arm study comprised 52 patients who underwent both ND and ULD scans. Subsequently AI-based denoising methods were applied to produce a denoised ULD scan. Two chest radiologists independently and blindly assessed all scans. Each scan was assigned a Lung-RADS score and grouped as 1 + 2 and 3 + 4.

Results: The study included 30 men (58%) and 22 women (42%); mean age 69.9 ± 9 years (range 54–88). ULD scan radiation exposure was comparable on average to 3.6–4.8% of the radiation depending on patient BMI. Denoising increased signal-to-noise ratio by 27.7%. We found substantial inter-observer agreement in all scans for Lung-RADS grouping. Denoised scans performed better than ULD scans when negative likelihood ratio (LR-) was calculated (0.04–0.08 vs. 0.08–0.12). Other than radiation changes, diameter measurement differences and part-solid nodules misclassification as a ground-glass nodule caused most Lung-RADS miscategorization.

Conclusions: When assessing asymptomatic patients for pulmonary nodules, finding a negative screen using ULD CT with denoising makes it highly unlikely for a patient to have a pulmonary nodule that requires aggressive investigation. Future studies of this technique should include larger cohorts and be considered for lung cancer screening as radiation exposure is radically reduced.

KEY WORDS: computerized tomography (CT), lung cancer screening, screening chest CT scan, lung imaging reporting and data system (Lung-RADS), ultra-low-dose (ULD) computerized tomography (CT)

Medical imaging and the resultant ionizing radiation exposure is a public concern mainly due to the possible risk of cancer induction. Computed tomography (CT) accounts for approximately half of the rapidly growing medical radiation exposure to patients [1-3] and rapid increases in CT use have heightened concerns about growing potential risks to the population as a whole [4,5] and to individual patients who undergo repetitive imaging studies [5-7]. An extreme form of repetitive exposure is lung cancer screening in which asymptomatic healthy individuals are irradiated [2]. Similarly, healthy patients who are incidentally found to have a pulmonary nodule are subject to numerous follow-up CT scans [8]. In response to these concerns, the radiology community has been striving to develop imaging strategies that deliver the lowest radiation dose necessary to obtain the desired information [9].

One approach for lowering radiation dose is to use CT-based technology for image quality improvement, such as tube current modulation, different iterative reconstruction techniques, or fine focal spot CT. The limitation of these approaches is that they are machine-specific and the images cannot be fixed once they are sent to the picture archiving and communication system (PACS). An alternative approach, known as post-processing, is used to improve or denoise images, following completion of CT acquisition and reconstructing [10].

A powerful method for denoising of ultra-low-dose CT images is by locally enforcing spatial consistency by non-local means [11], thus yielding preservation of fine image details and structures [11]. Another approach for denoising is based on artificial intelligence algorithms that teach deep neuronal networks to denoise low-dose CT [12].

In this study, we assessed the accuracy of ultra-low-dose (ULD) chest CT with and without denoising compared to a normal dose (ND) chest CT in the detection of pulmonary nodules. We also assessed the ability of denoised ULD CT to categorize patients, had they been screened for lung cancer, into those with a negative screening interpretation to those who would require further investigation.
PATIENTS AND METHODS

PARTICIPANTS
Patients selected for this prospective single-arm study were outpatients scheduled for an ambulatory chest CT scan in our department. For inclusion in the study, patients had to be 18 years of age or older, not pregnant, able to understand the study imaging plan, and be able to follow orders and hold their breath for 8 seconds. This study was approved by our institutional review board, and patients were consented in compliance with HIPPA regulations.

IMAGING METHOD
All patients were scanned on a dedicated CT machine, General Electric Revolution (GE, Milwaukee, WI, USA) for both the routine normal dose (ND) diagnostic scan and the ultra-low-dose (ULD) experimental scan. As the diagnostic scan was performed as routine clinical practice, intravenous contrast was used when clinically indicated. Immediately following the patient's diagnostic chest CT scan, the ULD chest CT scan was obtained. The ND CT scans were performed using our routine protocol consisting of 120 kVp and automatic current modulation. The additional ULD CT scan consisted of 120 kVp and fixed current at 10 mA for patients with a BMI < 29 kg/m² and 20 mA for a BMI ≥ 29 kg/m². For estimation of the effective radiation dose, the automatically recorded dose length product was multiplied with the conversion coefficient k of 0.017 mSv/mGycm [13].

Figure 1. Mismeasurement of a solid lung nodule (arrowhead) with differing techniques. The mean diameter measurements (average of maximal long-axis and perpendicular maximal short-axis measurement in the same plane) were 6 and 8 mm by reader 1 and 2, respectively for the ND CT [A], only 5 mm and 6 mm by the ULD CT [B], and 7 mm and 8 mm by the dULD scan [C]. Despite a 1 mm difference, this result affected Lung-RADS categorization, as the measurement of 5mm with the ULD would lead to a classification of a negative screening exam, whereas a 6 mm nodule would be considered Lung-RADS 3, which would require closer follow-up.

CT = computed tomography, dULD = denoised CT scans images, Lung-RADS = Lung CT Screening Reporting and Data System, ND = normal dose, ULD = ultra-low-dose
A fully convolutional denoising network was implemented, containing 8 convolution layers trained on the patch dataset (described above) by back-propagation using an Adam optimizer and a perceptual loss function [14]. Batch normalization was performed during training, a convergence was observed after 50 epochs. Consecutively, the ULD CTs of the evaluation set were processed by the trained network to generate the dULD CT scans.

**IMAGE NOISE ASSESSMENT**

Image noise was assessed in ND, ULD, and dULD CT scans using a cylindric ROI placed in the ascending aorta as previously described [16]. The ROI axial cross-section measured 100 mm² and extended along five consecutive axial slices. The average standard deviation (SD) and average signal-to-noise ratio (SNR), defined as the mean value-to-standard deviation ratio, were computed for all ND, ULD, and dULD CT scans.

**IMAGE EVALUATION**

The three sets of CT scans were de-identified, mixed, and distributed into batches by a non-radiologist (MG), so that each batch contained the same patient only once. The batches were sent one at a time to the radiologists who were collecting the data. Once the radiologist completed data collection from one batch of scans, a second batch was sent to the radiologist, ensuring that at least one week passed between the completion of a batch and the initiation of the subsequent batch, to prevent hindsight bias from recalling the salient findings.

Two experienced chest radiologists with 23- and 20-years of experience (radiologist 1 and 2) independently collected data from these three sets of chest CT scans while blinded to the clinical history. The data collected included: the presence of nodules and their type (solid, part-solid, or pure ground glass) and nodule diameter. Nodules were measured on lung windows, recording maximal long-axis and perpendicular maximal short-axis diameter in the same plane. An average was calculated, rounded to the nearest whole number. For the purpose of our study, nodules measuring < 4mm in diameter were disregarded; and each scan was assigned the Lung CT Screening Reporting and Data System (Lung-RADS) score (using Lung-RADS version 1.0). When there was more than one nodule type, the readers were instructed to document the size of the largest of them. The ND scan was used as the reference truth for the presence or absence of the collected imaging findings.

**STATISTICAL ANALYSIS**

Categorical variables were described as frequency and percentage. Continuous variables were evaluated for normal distribution using histograms and reported as median and interquartile range. Kappa statistic was used to evaluate the agreement in part-solid nodule types (i.e., solid, sub-solid, and pure ground glass) in each CT imaging method (ND, ULD, and dULD). Sensitivity, specificity, positive and negative predictive values, and likelihood ratios and accuracy were reported. Mann–Whitney Test was used to compare effective dose between obese (BMI ≥ 29) and non-obese (BMI < 29 kg/m²) patients. All statistical tests were two sided and the P value < 0.05 was used for statistical differences. Statistical analyses were performed using IBM Statistical Package for the Social Sciences statistics software, version 24 (SPSS, IBM Corp, Armonk, NY, USA).

**RESULTS**

**PARTICIPANTS**

Fifty-seven patients were enrolled in the study between August 2017 and January 2018. After excluding the 5 patients used for the denoising algorithm training set, there were 52 patients in our study cohort with a mean age of 69.9 ± 9 (range 54–88) years and included 30 (58%) men and 22 (42%) women. The patients' scan indication was: oncological follow-up in 41 (79%), pulmonary inflammation follow-up in six (11%) evaluation of an incidental abnormality detected by chest radiography in two (4%), respiratory symptoms in two (4%) and preoperative evaluation before cardiac surgery in one (2%). Forty-one patients (79%) had a BMI < 29 kg/m² (mean 24.3 ± 2.9 kg/m²; range 18.3–24.3 kg/m²); 11 patients (21%) had a BMI ≥ 29 kg/m² (mean 32.7 ± 3.3 kg/m², range 29.7–38.7 kg/m²).

**IMAGING METHOD**

Intravenous contrast (Omnipaque 350, GE Healthcare) was used in 20 (38%) patients. The ND average effective dose for patients with a BMI < 29 was 6.39 mSv, SD 2.5, range 1.03-10.92 as compared to an average effective dose of 0.23, SD 0.08, range 0.19-0.66 mSv using the ULD technique (an average radiation dose reduction of 95.1%). The ND average effective dose for patients with a BMI ≥ 29 kg/m² was 8.76 ± 2.70, range 1.10–11.52 mSv as compared to an average effective dose of 0.42 ± 0.03, range 0.39–0.49 mSv using the ULD technique (an average radiation dose reduction of 92.4%).

**IMAGE NOISE ASSESSMENT**

For ND CT, the computed average standard deviation (STD) and signal to noise ratio (SNR) where STD = 9.5 H.U. ± 2.2 and SNR = 117.1 ± 27.9, while for ULD CT the values degraded to STD =33.9 H.U. ± 4.2 and SNR=31.8 ± 4.4. Following, denoising by our AI-based method (dULD CT images), STD and SNR recovered to 26.3 H.U. ± 3.6 and 40.6 ± 6.4, respectively, corresponding to a decrease in STD of 22.4 %, and an increase in SNR of 27.7%.

**IMAGE EVALUATION**

The distribution of solid nodules ≥ 4 mm or masses, part solid nodules and pure-ground glass nodules by the reference ND
CT scan as compared to the ULD and dULD CT scans is depicted in Table 1. The number of patients who had part solid (n=3–8) and pure ground-glass nodules (n=5–8) was significantly smaller (P < 0.001) than patients with solid nodules (n=32–38) by all three scanning methods.

The distribution of Lung-RADS categories for both readers is presented in the supplemental table.

Inter-observer reliability for Lung-RADS categories, when grouped into those with a negative screening exam (Lung-RADS 1 + 2) and to those requiring further investigation (Lung-RADS 3 + 4), was tested by Cohen's kappa coefficient, and a substantial (k = 0.650) agreement for the ND scans was recorded. Substantial agreement for the ULD (k = 0.615) and moderate agreement for dULD (k = 0.532) was achieved. Because radiation changes were present in our enriched patient cohort, assigning them a Lung-RADS category resulted in considerable inter-observer and intra-observer variability. When patients with pulmonary radiation changes were excluded, the inter-observer agreement improved and was substantial in all scans; ND (k = 0.724), ULD (k = 0.729), and dULD (0.683).

Intra-observer variability for Lung-RADS categories among the different scans is manifested in the following likelihood ratios that were calculated after excluding patients with radiation changes; ULD negative likelihood ratio (LR-) was 0.08–0.12. dULD performed better with a negative likelihood ratio of 0.04–0.08 [Table 2].

There were several reasons for inter-observer and intra-observer discrepancies in Lung-RADS categorization among the ND and both dULD and ULD scans. Diameter measurement differences caused the miscategorization of patients from a negative scan to that requiring further investigation [Figure 1]. Both ULD and dULD had 5/52 inter-observer discrepancies, whereas concordance was better with the ND with only 3/52 discrepancies; dULD scans had less intra-observer measurement discordance, 1-2/52 versus 2-3/52.

Part-solid nodules misclassification as a ground-glass nodule was found by both readers in one (1/52) patient with part-solid nodules when assessed by the ULD CT scans and only one reader misclassified this single case (1/52) by dULD [Figure 2]. One nodule in one patient (1/52) was missed by one reader on both the ULD CT scans the dULD scan.

**DISCUSSION**

The main finding of our study is that our denoising method for ULD CT, significantly improves SNR of ULD CT while still maintaining interpretive accuracy of the ULD CT as a screening tool. Applying this denoising to ULD CT improves nodule type classification and measurement, which improved performance in the identification of patients with a negative screening for lung cancer examination who do not require further investigation. Although we compared the performance of dULD CT using Lung-RADS criteria (classification that was devised to standardize lung cancer screening), guidelines for incidental pulmonary nodules use similar size criteria to differentiate patients who require more aggressive evaluation to those who do not [8]. Both scenarios are similar. In both, healthy individuals are subject to repetitive CT scans, a great majority of whom will be found to not have a malignancy.

In our study we assessed our denoising technique on real patients by performing two separate scans, rather than relying on artificial noise to develop our denoising method. Another advan-

### Table 1. Nodule type and size distribution for both reads in the three chest CT techniques

<table>
<thead>
<tr>
<th></th>
<th>ND Reader 1</th>
<th>ND Reader 2</th>
<th>ULD Reader 1</th>
<th>ULD Reader 2</th>
<th>dULD Reader 1</th>
<th>dULD Reader 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Solid</strong></td>
<td></td>
<td></td>
<td>--------------</td>
<td>--------------</td>
<td>---------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Number*</td>
<td>32 (62%)</td>
<td>36 (69%)</td>
<td>34 (65%)</td>
<td>37 (71%)</td>
<td>37 (71%)</td>
<td>38 (75%)</td>
</tr>
<tr>
<td>Diameter**</td>
<td>12.8 (4.0–73.5)</td>
<td>12.0 (4.0–55.0)</td>
<td>11.5 (3.0–73.5)</td>
<td>21.75 (4.0–60.0)</td>
<td>12.4 (3.5–69.0)</td>
<td>11.5 (4.0–60.0)</td>
</tr>
<tr>
<td><strong>Part solid</strong></td>
<td></td>
<td></td>
<td>--------------</td>
<td>--------------</td>
<td>---------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Number*</td>
<td>7 (13%)</td>
<td>4 (8%)</td>
<td>8 (15%)</td>
<td>5 (10%)</td>
<td>8 (15%)</td>
<td>3 (6%)</td>
</tr>
<tr>
<td>GGO size**</td>
<td>8.5 (3.5–36.0)</td>
<td>28.0 (8.0–38.0)</td>
<td>12.8 (5.0–36.5)</td>
<td>27.3 (9.0–37.0)</td>
<td>12.8 (5.0–36.5)</td>
<td>31.5 (23.5–38.0)</td>
</tr>
<tr>
<td>Solid size**</td>
<td>7.7 (1.0–21.0)</td>
<td>15.0 (4.0–20.0)</td>
<td>10.7 (1.0–31.0)</td>
<td>17.5 (15.0–20.0)</td>
<td>11.9 (1.0–20.0)</td>
<td>21.7 (15.0–30.0)</td>
</tr>
<tr>
<td><strong>Pure ground glass</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number*</td>
<td>7 (14%)</td>
<td>5 (10%)</td>
<td>7 (13%)</td>
<td>8 (15%)</td>
<td>6 (12%)</td>
<td>6 (12%)</td>
</tr>
<tr>
<td>Size**</td>
<td>10.0 (5.5–26.5)</td>
<td>15.5 (7.0–22.0)</td>
<td>9.0 (4.5–21.5)</td>
<td>12.3 (6.5–21.5)</td>
<td>11.5 (5.5–23.5)</td>
<td>9.5 (2.5–21.5)</td>
</tr>
</tbody>
</table>

*Number of nodules and percentage in the study population in parenthesis

Mean nodule size in millimeters, range in parenthesis
dULD = denoised ultra-low dose, ND = normal dose; ULD = ultra low dose
tage to our denoising method is that it uses AI and deep-learning techniques to denoise ULD CT studies and is thus applicable to any CT machine without the addition of any filters or hardware. Our denoising method has been shown to outperform state-of-the-art denoising neural network that are trained on synthetic noise models for practical reasons [14].

Lung-RADS categorization discrepancies occurred in the interpretation of all scan types when comparing both readers (inter-observer) as well as when comparing each reader's categorization using the different scan methods. It is known that visual assessment of pulmonary nodules with regard to type and size is susceptible to observer variability [17,18], which may lead to Lung-RADS assignment variability between readers [19]. Nodule mismeasurement was the most common cause of disconcertedness other than radiation changes. The dULD scans had less measurement variability compared to ULD scans.

When comparing Lung-RADS categorization to those with a negative screening exam (Lung-RADS 1 + 2) to those requiring further investigation (Lung-RADS 3 + 4), the main cause for variability between ULD and dULD was from radiation changes. Unlike a lung cancer screening population, we used an enriched study population with sickest-of-the-sick patients, which included patients treated with lung irradiation. These radiation changes were recognized as such on all scans; however, as our readers were forced to use Lung-RADS categorizations on this enriched population, at times these radiation changes were categorized as a nodule and at times disregarded as radiation changes. It is because of this issue that we focused on patients without radiation changes to evaluate the performance of Lung-RADS in our study population. Otherwise, variability chiefly resulted from nodule measurement differences. It is possible that some of the discrepancies we saw with 1 mm differences in diameter measurements causing Lung-RADS category discrepancies were due to slice selection rather than an erroneous measurement on the ULD or dULD technique. Therefore, it could be that measurements with ULD or dULD were actually more accurate than displayed in this study.

### Table 2. Comparison of Lung-RADS assignment to categories between denoised ULD and dULD chest CT when using normal dose chest CT as the standard of reference

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>LR+</th>
<th>LR-</th>
</tr>
</thead>
<tbody>
<tr>
<td>ULD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reader 1</td>
<td>0.92</td>
<td>0.95</td>
<td>18.04</td>
<td>0.08</td>
</tr>
<tr>
<td>Reader 2</td>
<td>0.90</td>
<td>0.88</td>
<td>7.17</td>
<td>0.12</td>
</tr>
<tr>
<td>dULD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reader 1</td>
<td>0.96</td>
<td>1.00</td>
<td>∞</td>
<td>0.04</td>
</tr>
<tr>
<td>Reader 2</td>
<td>0.93</td>
<td>0.88</td>
<td>7.45</td>
<td>0.08</td>
</tr>
</tbody>
</table>

For statistical evaluation, Lung-RADS categories were simplified into 1 + 2 vs. 3 + 4 groups

After excluding seven patients who had radiation changes (n=45) equilibrium

CT = computed tomography, dULD = denoised CT scans images, Lung-RADS = Lung CT Screening Reporting and Data System, ND = normal dose, ULD = ultra-low-dose

Figure 2. Misclassification of nodule density

A nodule in the left upper lobe was classified as part-solid [A] on the ND scan, as a pure ground glass nodule by both readers on the ULD scan [B], whereas with the dULD scan, nodule density was identified correctly as a part-solid nodule by one reader and misclassified as a pure ground glass nodule by the second reader [C]

CT = computed tomography, dULD = denoised CT scans images, Lung-RADS = Lung CT Screening Reporting and Data System, ND = normal dose, ULD = ultra-low-dose
LIMITATIONS
Because drastic radiation reductions are a concern for repetitive imaging of healthy individuals, as is screening, we tried to assess our method in the screening realm. There are two reasons why we did not conduct the study using a screening cohort. First, at the time we prepared the manuscript, screening for lung cancer was not performed in Israel. Second, because more than 90% of screening exams are negative, assessing nodule detection accuracy on a screened population cohort would have required a very large patient cohort. Therefore we chose to conduct our study on an enriched population in which more nodules were expected to test our method. Although a newer Lung-RADS version was released in 2019, after we completed data collection, both Lung-RADS version 1.0 and version 1.1 maintain the same cutoff value for a negative screen, of a nodule smaller than 6 mm. Thus the results of our study are applicable to the current Lung-RADS version. Another limitation to our study is that despite being a prospective study, the ULD CT and the ND CT were two separate acquisitions, two separate breath-holds. An ideal scenario would have been to conduct both methods on the same breath-hold to eliminate any slice selection inaccuracies in the measurement of nodules and in nodule subtype interpretation. However, technically this latter scenario is impossible.

CONCLUSIONS
When assessing asymptomatic patients for pulmonary nodules using ULD CT with denoising it is unlikely for a patient to have a pulmonary nodule that requires aggressive investigation. This result pertains to whether the assessment was for screening or following a pulmonary nodule. This technique should be further studied on a larger population cohort and considered for lung cancer screening as patients are exposed on average to 3.6–4.8% of standard chest CT radiation dose depending on patient BMI.

Correspondence
Dr. A. Kerpel
Dept. of Diagnostic Radiology, Sheba Medical Center, Tel Hashomer 52621, Israel
Fax: (972-3) 530-2200
e-mail: arikerpel@gmail.com

Reference