Published in final edited form as:

Acad Radiol. 2009 April ; 16(4): 464–476. doi:10.1016/j.acra.2008.09.016.

Development of a High-Performance Dual-Energy Chest Imaging System: Initial Investigation of Diagnostic Performance

H. Kashani¹, G.J. Gang², N. A. Shkumat², C. A. Varon¹, J. Yorkston³, R. Van Metter³, N. S. Paul⁴, and J. H. Siewerdsen^{1,2}

¹Ontario Cancer Institute, Princess Margaret Hospital, University Health Network, Toronto ON Canada M5G 2M9

²Department of Medical Biophysics, University of Toronto, Toronto ON Canada M5G 2M9

³Carestream Health Inc., Rochester NY

⁴Department of Medical Imaging, University Health Network and Mt. Sinai Hospital, Toronto ON Canada M5G 2M9

Abstract

Rationale and Objectives—To assess the performance of a newly developed dual-energy (DE) chest radiography system in comparison to digital radiography (DR) in the detection and characterization of lung nodules.

Materials and Methods—An experimental prototype has been developed for high-performance DE chest imaging with total dose equivalent to a single posterior-anterior DR image. Low- and high-kVp projections were used to decompose DE soft-tissue and bone images. A cohort of 55 patients (31 male, 24 female, mean age 65.6 years) was drawn from an ongoing trial involving patients referred for percutaneous CT guided biopsy of suspicious lung nodules. DE and DR images were acquired of each patient prior to biopsy. Image quality was assessed by means of human observer tests involving 5 radiologists independently rating the detection and characterization of lung nodules on a 9-point scale. Results were analyzed in terms of the fraction of cases at or above a given rating, and statistical significance was evaluated from a Wilcoxon signed rank test. Performance was analyzed for all cases pooled as well as by stratification of nodule size, density, lung region, and chest thickness.

Results—The studies demonstrate a significant performance advantage for DE imaging compared to DR ($p < 0.001$) in the detection and characterization of lung nodules. DE imaging improved the detection of both small and large nodules and exhibited the most significant improvement in regions of the upper lobes, where overlying anatomical noise (ribs and clavicles) are believed to reduce nodule conspicuity in DR.

Conclusions—DE imaging outperformed DR overall, particularly in the detection of small, solid nodules. DE imaging also performed better in regions dominated by anatomical noise such as the lung apices. The potential for improved nodule detection and characterization at radiation doses equivalent to DR is encouraging and could augment broader utilization of DE imaging. F studies will

Corresponding Author: Jeffrey H. Siewerdsen, Ph.D., Ontario Cancer Institute, Room 7-502, Princess Margaret Hospital, 610 University Ave., Toronto, ON, Canada M5G 2M9, Phone: 416-946-4501 (x5516), Fax: 416-946-6529, E-mail: jeff.siewerdsen@uhn.on.ca.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

extend the initial cohort and rating scale tests to a larger cohort evaluated by ROC and will evaluate DE imaging in comparison and as an adjuvant to low-dose CT.

Keywords

dual-energy imaging; flat-panel detector; diagnostic radiology; pulmonary nodule; chest radiography; lung cancer

I. INTRODUCTION

As the prognosis for advanced stage lung cancer is extremely poor, early detection is a major priority.^{1,2} X-ray projection imaging [screen-film, computed radiography (CR), or digital radiography (DR)] is still the most common tool in chest radiology, due in part to its low dose, low cost, and high workflow efficiency, but is known to exhibit poor sensitivity in detection of small, subtle pulmonary nodules.^{3,4} A major limiting factor is anatomical clutter superimposed within the projection, which can impede the detection and characterization of subtle lung nodules.^{5,6} To mitigate this limitation, dual-energy (DE) imaging acquires two projections of the patient at different x-ray energies and selectively decomposes the image into soft-tissue and bone components. The former presents soft-tissue structures in a context that is largely free from the main source of anatomical clutter – viz., the ribs and clavicles – thus improving the conspicuity of subtle nodules.⁷ Further, as the presence of calcification is an important indicator of benignancy, DE imaging could help to characterize benign lesions with a higher level of specificity.^{3,8}

Previous studies have investigated single- and dual-exposure DE imaging in comparison to conventional chest radiography,⁹⁻¹² each demonstrating increased detectability of lung nodules in DE images. A variety of DE imaging systems are commercially available and are finding broad clinical utilization, including single-shot CR systems (FCR XU-D1; Fujifilm, Tokyo, Japan) and double-shot FPD systems (XQ/I Revolution, General Electric, Milwaukee, WI) featuring a 200 ms delay between low- and high-energy exposures. Kelcz et al.⁹ established the advantage of DE images in terms of both the detection of nodules and the improved visibility of nodule calcification. More recently, investigators have described the use of a flat-panel detector in DE imaging to improve the detectability of lung nodules even further due to higher detective quantum efficiency (DQE) compared to screen-film and CR systems.¹⁰⁻¹² Higher DQE implies reduced image noise at equivalent dose or, alternatively, equivalent image quality at reduced dose, suggesting the possibility of lower-dose DE imaging without loss of diagnostic information.

As described below, a DE imaging prototype has been developed with performance characteristics permitting high-quality DE images to be acquired at a total dose equivalent to that of a conventional PA chest radiograph.^{13, 14} Previous work involved characterization of physical performance metrics, such as contrast-to-noise ratio, noise-power spectrum (NPS), and noise-equivalent quanta (NEQ),¹⁵⁻¹⁷ and identified optimal system configuration and imaging techniques, including x-ray filtration, kVp selection, and allocation of dose between low- and high-kVp images.^{13,14} The work has yielded a system suitable for an early clinical trial designed to evaluate the diagnostic performance of DE imaging. This paper draws from the first 55 patients of that trial to evaluate the diagnostic performance of DE imaging in comparison to DR. Performance was evaluated in terms of satisfaction in the detection and characterization of lung nodules. While the approach does not provide a full characterization of sensitivity and specificity (as would be yielded from ROC tests), the work is valuable and distinct from previous studies in several respects. First, the DE imaging prototype offers potentially improved imaging performance by virtue of physical implementation (e.g., a cardiac gating system and a filter wheel for differential filtration between low- and high-kVp

projections) and optimized imaging techniques (e.g., optimal low-/high-kVp pair and dose allocation). Second, the data and methodology support diagnostic performance evaluation in terms of characteristics previously not investigated (e.g., stratification of data according to nodule size, density, location, and patient size). Finally, the ability of the prototype to acquire DE images at total radiation dose equivalent to that of the DR images suggests a potential 'new normal' in x-ray radiographic examination, wherein a composite radiograph roughly equivalent to DR may always be decomposed, but material-specific "soft-tissue" and "bone-only" decompositions offer improved conspicuity and characterization of subtle abnormalities.

II. MATERIALS AND METHODS

A. Volunteer Patient Cohort

Patients were accrued under informed consent in a prospective, non-randomized trial with approval from the institutional Research Ethics Board and with Health Canada Investigational Testing Authorization. Total accrual for the trial is approximately 200 patients, each drawn from the patient population referred for a percutaneous CT guided biopsy of suspicious lung nodules. Patients were accrued into 5 arms differing in DE imaging technique to evaluate different DE imaging parameters, including differential x-ray spectra, cardiac gating, dose allocation, and total dose. All images used in the study reported herein were taken from Group 1 ("optimal" DE imaging technique, described below), giving 55 cases from the 90 patients accrued at the time of this study, including 31 male and 24 female subjects. Mean age was 65.6 yr (ranging 26 – 90 yr). The following data were gathered for each patient in the trial: standard-of-care image data (diagnostic CT, ultra low-dose CT acquired just prior to lung biopsy, and a post-biopsy CR image); experimental protocol image data (a DR image and a DE image, each described below); and biopsy data. Percutaneous CT guided transthoracic biopsies were performed immediately following the DE/DR imaging exam to provide a definitive diagnosis of the lung lesion. Biopsies were performed either by fine-needle aspiration and cytologic examination or by core biopsy and histologic examination.¹⁸

B. DE Imaging System

The DE imaging system was developed in collaboration with Carestream Health Inc. (Rochester NY). The basic platform for the prototype was a Modified Kodak RVG 5100 digital radiography chest stand (Carestream Health Inc., Rochester, NY).¹³ An acquisition workstation controls generator technique setting, filter selection, detector acquisition parameters, and data transfer. Modifications include a high performance FPD (Trixell Pixium-4600, Moirans, France) with a CsI:Tl scintillator and 143 μm pixel pitch. A system for cardiac-gated image acquisition was implemented, using a fingertip pulse oximeter to trigger x-ray exposure coincident with diastole, thus reducing cardiac motion artifacts. DE images were acquired according to optimal techniques identified in previous work.^{13,14} For example, for the "average" patient size (24 cm chest thickness): $\text{kVp}^{\text{Low}} = 60 \text{ kVp}$ (2.5 mm Al total filtration); $\text{kVp}^{\text{High}} = 120 \text{ kVp}$ (4.5 mm Al + 0.6 mm Ag total filtration). The high-energy filter was selected to "harden" the beam, reduce spectral overlap between the low- and high-energy projections, and thereby improve contrast in the resulting DE image. Such was the subject of considerable investigation in previous work.^{13,14} Added filtration for the DR image was typical of that in conventional digital chest radiography. Three different filters were therefore required for the low-energy, high-energy, and DR exposures and were implemented using a computer-controlled, multi-position filter wheel within the collimator. The dose (i.e., imparted energy¹⁹) was computed separately for the low-energy projection (ϵ_{Low}), the high-energy projection (ϵ_{High}), and the DR projection (ϵ_{DR}) as in Reference 13. Note that each of these was independent and was computed separately based on the kVp, filtration, mAs, and patient thickness specific to each. The technique chart of Reference 13 was interpolated across patient thickness ranging 18 – 28 cm in increments of 2 cm, with total dose dependent on patient

thickness (e.g., 0.11 mGy for 24 cm chest thickness). For the DE image, the total dose was $\epsilon_{DE} = \epsilon_{Low} + \epsilon_{High}$, and the “dose allocation” was such that the fraction of dose from the low-kVp projection was ~ 0.33 of the total (i.e., $\epsilon_{Low} / \epsilon_{DE} = 0.33$), previously determined to be optimal for soft-tissue image quality.¹³ The DE image dose was equivalent to that of a conventional PA DR image (i.e., $\epsilon_{DE} = \epsilon_{Low} + \epsilon_{High} = \epsilon_{DR}$). Such was accomplished by: 1.) fixing ϵ_{DR} to that of current clinical technique (1 mR to the detector etc., as described above); and 2.) adjusting mAs_{Low} and mAs_{High} such that the allocation was ~ 0.33 while ensuring $\epsilon_{DE} = \epsilon_{DR}$. DE images were processed and decomposed as illustrated in Fig. 1. Offset and gain corrections were based upon 25 averaged dark-fields and 25 averaged flood-fields (acquired at $\sim 50\%$ sensor saturation), respectively. Prior to DE decomposition, the high-energy image was automatically registered to the low-energy image by means of deformable registration based on mutual-information maximization and a morphological pyramid.²⁰ The average vector magnitude displacement of high-energy projections (averaged over all 55 cases) was ~ 3.1 mm, corrected using the deformable registration technique. A complete description and quantitative characterization of the registration technique is in progress. The low-energy image and the (registered) high-energy image were decomposed into soft-tissue and bone images by log-weighted subtraction:

$$\ln(I_{soft}^{DE}) = \ln(I^H) - w_s \ln(I^L) \quad (1a)$$

$$\ln(I_{bone}^{DE}) = -\ln(I^H) + w_b \ln(I^L) \quad (1b)$$

$$\ln(I_{comp}^{DE}) = \ln(I^H) + w_c \ln(I^L) \quad (1c)$$

Where the tissue cancellation parameters (w_s , w_b and w_c) were selected according to previous work¹⁵ that identified optimal settings (dependent on kVp selection) and were qualitatively validated during image decomposition by an expert observer (who was not among the observers in the performance evaluation below). For this initial study, values of $w_s = 0.20$, $w_b = 0.57$ and $w_c = 0.9$ were identified as optimal and were judged to give acceptable tissue cancellation across all 55 patients. Future work will include more advanced means of patient-specific, automatically selected, spatially varying parameter selection and will consider incorporation of w_s and w_b as parameters that may be freely varied by the radiologist in a manner analogous to window/level. The composite image [Eq. (1c)] is intended to be nearly identical to a conventional radiograph, and future clinical implementations could present this image in combination with the soft-tissue and bone images, potentially boosting the performance of the DE imaging further. For purposes of the current study, only the soft-tissue and bone images were included in the DE image set in the observer tests.

Soft-tissue and bone images were transformed to log-exposure space, and window-level/width settings were adjusted qualitatively by the same expert observer to yield display-ready images for each case. Example DE images are shown in Fig. 2(a,b). For the observer studies described below, the DE images were presented as a “two-slice volume”. The observers were able to scroll easily (mouse-wheel) between the soft-tissue and bone images, a feature that allowed readers to quickly evaluate corresponding locations in the images (as opposed to, for example, side-by-side display). The composite image was not used, as the objective of the current study

was to evaluate the diagnostic performance associated with the DE decompositions (rather than the combined information of composite and decomposed images).

DR image processing was consistent with that in clinical DR systems based on the same imaging platform (Kodak RVG-5100 DR chest stand; Carestream Health Inc. Rochester NY), with tissue equalization, edge enhancement, and other post-processing parameters defined in the associated WIISE™ and Eclipse™ image processing software (Carestream Health Inc., Rochester NY), with parameters therein proprietary to the manufacturer. DR images were acquired on the same imaging prototype immediately (within a few seconds) following the DE image acquisition. DR image acquisition technique was 120 kVp (1 mm Al + 0.2 mm Cu), with mAs interpolated from the technique chart of Reference 13 such that exposure to the detector was 1 mR [e.g., ~1.6 mAs for an average (24 cm thick) patient]. For each patient, the total dose for the DE image and DR image were equivalent (e.g., 0.11 mGy for 24 cm chest thickness). Offset and gain corrections, transformation to log-exposure space, and window / level adjustment for the DR images were the same as for DE images. Both DE and DR images were therefore acquired at optimal techniques and at equivalent dose, allowing reasonably fair comparison between the two modalities. As mentioned above, image processing and display were similarly consistent between the two. Such included image processing filters and tone-scaling that may have favored the DR images, as such techniques have been thoroughly investigated over the last decade for DR but are yet to be fully optimized for DE images (the subject of ongoing work). The specific image processing filters and tone-scaling techniques were proprietary to the manufacturer (Kodak Health Imaging Systems, now Carestream Health Inc.). They are standard to commercially available DR image reading stations, but additional technical details are proprietary and not available. An example DR image for the same patient as in Fig. 2(a,b) is shown in Fig. 2(c).

C. Observer Study

Because the initial dataset was not well suited to ROC analysis (specifically due to a lack of true-negative cases in the patient cohort), a simple test based on diagnostic satisfaction was adopted as an initial evaluation of diagnostic performance. While such does not provide a measure of diagnostic sensitivity and specificity, such tests have shown to provide objective differentiation of performance associated with perception of subtle image quality factors.²¹ Such tests were deemed sufficient to establish whether is a performance advantage for DE or DR imaging based on the initial dataset. Future evaluation based upon the completed trial will include sensitivity and specificity (ROC analysis). In the event that the full patient cohort similarly lacks true Negative cases, we hope to overcome such limitation by performing an ROC study using images of the left or right half of the chest (since patients typically exhibited nodules in one lung or the other, but not both). Each reader scored each image on a 9-point satisfaction scale as shown in Table I. Scales with fewer (5) and more (up to 100) levels of rating were considered in preliminary studies, with 9 found to be tolerated well by observers and consistent with reasonably fine levels of image quality discrimination, as described by Van Metter and Foos.²¹

Five expert observers (3 radiology fellows, 2 radiology staff, each a specialist in thoracic imaging) were independently presented with the images. The reading order was randomized, with one image presented at a time [either a DR image or a DE image (two-slice volume, described above)]. The observer was asked to rate each image in terms of his/her ability to detect and characterize abnormalities according to the 9-point satisfaction scale shown in Table. I. The scale was visible to observers at all times on a second monitor (illustrated in Fig. 1). The following methods were employed to minimize observer bias in the image evaluation. The study was conducted in a clinical radiology reporting room with subdued lighting on diagnostic-quality, monochrome LCD monitors (AM-QX21-A9300, National Display, San

Jose, CA). The monitors were adjusted to meet the DICOM gray scale standard. To reduce inter-reader variability and simplify image controls, window width /level and magnification were fixed. In this initial study, given that all cases were positive (and often conspicuous), such display controls would likely not have affected the results. Future tests using the completed trial data will allow individual readers to freely vary window width / level to better accommodate variability in reader preferences, particularly with respect to a broader spectrum of cases.

Preceding each observer test, a training session involving 8 DE and 7 DR images was conducted to familiarize the observers with the software and standardize their understanding of the rating scale. The training images were drawn from the pool of cases available at the time of the study and did not overlap with those used as test images. Given the differences between the two modalities (DE and DR) and the fact that the observers were specialist thoracic radiologists, it was obvious whether a given image was a DE image or a DR image. While such represents a potential source of bias, readers were asked to respond strictly with respect to their satisfaction as described by the rating scale, irrespective of the image modality. We used this approach together with randomization of the case reading order in order to minimize observer bias for or against a given modality. To examine intra-reader consistency during the actual test, the first 7 images displayed in each test were displayed again at the end of the test (without informing the observers that the images were repeated), with differences examined in terms of the Wilcoxon signed rank p-value, as described below. The first 7 images were rejected from the study, except for purposes of intra-reader variability; therefore, the analysis pertains to each DE and DR image shown once (no repeats). Repeat readings were found to be highly reproducible for expert observers in preliminary studies and would not add to the statistical power of the study.

D. Statistical Analysis

1.) Fraction of responses at or above a given rating—The rating scale responses constitute ordinal data. The fraction of observer responses (F) at or above a given rating (R) was plotted versus rating scale, giving curves that range 0 to 1 on the vertical axis plotted versus the ordinal rating scale (1 to 9) on the horizontal axis – essentially a cumulative histogram of responses. As each observer rated the images based on a 9-point rating scale, we were able to calculate the fraction of responses at or above a given value of R . For example, the percentage rated $R=1$ or higher was always 100%, the percentage rated $R=2$ or higher was a bit less, and so forth. The plots summarize the results in a simple form, with “higher curves” corresponding to superior performance. The analysis is similar to that of Van Metter and Foos²¹ who showed the technique capable of differentiating fairly subtle differences in image quality. Moreover, computing fractions at or above a given rating is an appropriate way to handle qualitative, ordinal data (as opposed to computing the mean value of R).

Error bars on such plots reflect a two-sided 95% confidence interval computed according to a binomial distribution²² as described below. The corresponding error bars are asymmetric and appropriately bounded between 0 and 1. Each image was scored as either (i) at or above a certain rating, or (ii) below that rating, giving two mutually exclusive outcomes such that F (the ‘fraction at or above a given rating’) follows a binomial distribution. The upper and lower bounds of the confidence intervals were calculated as:

$$\text{Upper bound: } \sum_{\kappa=0}^{N_d} \binom{N}{\kappa} P_U^\kappa (1 - P_U)^{N-\kappa} = \frac{\alpha}{2} \quad (2a)$$

$$\text{Lower bound: } \sum_{\kappa=0}^{N_d-1} \binom{N}{\kappa} P_L^\kappa (1 - P_L)^{N-\kappa} = \frac{1 - \alpha}{2} \quad (2b)$$

where N is the total sample size, N_d is the number of samples with the outcome of interest, p is the proportion of samples with the outcome of interest, p_U is the upper bound of p , and p_L is the lower bound of p , and α (commonly fixed at 5%) is the Type I error (i.e., the probability of rejecting the null hypothesis when, in fact, the null hypothesis is correct). The results are similarly interpreted in terms of the p-value (i.e., the probability of obtaining the measured results if the null hypothesis were correct, also commonly taken as a statistically significant difference for p-value < 5%). The data are complemented by the 95% confidence interval as in Eqs. (2) and plotted as error bars to indicate the limits within which the true difference between groups is expected to lie. The upper and lower bounds were calculated using Newton-Raphson method in Matlab (The Mathworks, Natick MA) to a precision of 1×10^{-6} .

2.) Statistical significance: P-value from Wilcoxon signed rank test—The statistical significance in differences observed between DE and DR scores was evaluated in terms of the p-value at a 95% level of significance from a Wilcoxon signed rank test – a non-parametric test suitable to paired ordinal data, assuming all observations within a given modality are independent.²³ The p-value was calculated using the Matlab function ‘signrank’ accounting for both the sign and magnitude of the difference in ratings. A one-sided hypothesis test was used (with the alternative hypothesis that ‘DR scores are significantly higher than DE’) and the p-values were corrected by a factor of 0.5 from that of the two-sided test.

To examine intra-reader consistency, the Wilcoxon signed rank test was also used in relation to the first and last 7 images in the reading study, repeated at the beginning and end of the test as described above. In this case, the alternative hypothesis was that ‘the two sets of scores are not equivalent;’ therefore, a two-sided p-value was calculated directly from ‘signrank’. Across 5 observers, all p-values assessing intra-reader consistency were greater than 0.05 (specifically, p-value = 0.125, 1, 0.0625, 0.5313, and 1 for observers A-E, respectively), suggesting that there was no significant difference in observer readings at the beginning and end of the test – i.e., that training appears to have been sufficient, and there was no evidence of observer fatigue.

3.) Stratification of the results—Performance was analyzed for all cases pooled, as well as by post-hoc stratification of the data according to lesion size, lesion density, chest thickness, gender, and location of the lesion. Lesion size was characterized as the greatest linear dimension as measured on CT, and the results were stratified as “nodule” (<3 cm) or “mass” (≥3 cm), consistent with typical clinical terminology. Lesion density was measured using the attenuation measurement tool on a PACS workstation (Fusion E-film 2.1, Merge Healthcare, Milwaukee, WI), and the results were stratified as “Solid” (≥20 HU) or “Non-solid” (<20 HU). Chest thickness was characterized as the anterior-posterior distance measured from the xiphoid process to the T9 thoracic vertebra taken from the axial CT image at this level, and the results were stratified as “Average” (≤26 cm) or “Thick” (>26 cm). Lesion location was determined according to the anatomical position with respect to lung and lobe (or mediastinum), and the results were stratified as “Right-Upper,” “Left-Upper,” “Right-Middle,” “Left-Middle,” “Right-Lower,” “Left-Lower,” and “Mediastinum.” The number of cases overall and within each stratum are summarized in Table II.

III. RESULTS

A. All Cases Pooled

The results for the 275 total ratings (5 radiologists * 55 cases) for each of the two modalities (DE and DR) are summarized in Fig. 3. Individual case-by-case comparison of DE and DR for each patient is evaluated in Fig. 3(a). In 41.5% (114/275) of cases, the DE image was rated superior to DR by at least a difference of $R=1$. In 38.9% (107/275) of cases, the DE and DR images were rated equal. In 19.6% (54/275) of cases, the DR image was rated superior. Further to this case-by-case examination, the proportion of cases for which one modality was superior (or equal) to the other as judged by 3 or more observers (out of 5) is plotted in Fig. 3(b). Such analysis yielded 55 responses (out of 275) in which 3 or more observers agreed on the comparison between the two modalities. Of these: 36.4% (20/55) scored DE superior to DR; 36.4% (20/55) rated DE and DR equivalent; and 5.5% (3/55) rated DR superior to DE. In the remaining 12 cases (21.8%), a majority could not be reached regarding the superiority / equality / inferiority between the two modalities. The fraction of images rated at or above a given rating score [Fig. 3(c)] shows that DE rated consistently higher than DR (p -value < 0.001) in the detection and characterization of lung nodules.

B. Stratification by Lesion Size

The data were subsequently analyzed in terms of cases for which the lesion size was ≤ 3 cm and > 3 cm (36 and 19 cases, respectively, as shown in Table II). The rationale for the size demarcation is consistent with clinical convention for which lesions smaller than 30 mm are termed “nodules,” whereas lesions 30 mm or larger are termed “masses.” Results are shown in Fig. 4. A statistically significant improvement in satisfaction was observed for DE imaging in each case [p -value < 0.001 for nodules (lesion size < 3 cm), and p -value = 0.0264 for masses (lesion size ≥ 3 cm)]. The advantage of DE is more pronounced for nodules, as seen from the distinctly separated curves in Fig. 4(a) and the correspondingly smaller p -value.

C. Stratification by Lesion Density

Cases were alternatively stratified according to lesion density as solid (≥ 20 HU) and non-solid (< 20 HU), with 46 and 9 cases, respectively (Table II). A statistically significant improvement in diagnostic performance was observed for DE imaging of solid lesions (p -value < 0.001). For non-solid lesions, DE and DR scores were not significantly different overall (p -value = 0.0968). As mentioned above, we classified nodules as “solid” for a CT attenuation ≥ 20 HU, corresponding to typical clinical terminology and representing the prevalent phenotype for the most common histological subtypes of lung cancer (adenocarcinoma, squamous cell, and small cell lung cancer). Non-solid lesions, on the other hand, such as bronchioalveolar carcinoma (a subtype of adenocarcinoma) characteristically presents as ground glass opacities that typically take several years to grow. Such lesions are difficult to detect in chest radiography and are typically followed by annual (low-dose) chest CT. The superior detection of solid nodules observed for DE in comparison to DR is important in the detection of primary or metastatic disease of the thorax. The results show no statistical difference in the detection of non-solid nodules between DE and DR ($p=0.0968$), and the difficulty in detecting ground glass opacities (for both DE and DR) is an issue that will be investigated further using the completed trial data and modified DE imaging techniques (e.g., noise reduction algorithms).

D. Stratification by Chest Thickness

Cases were also stratified by PA chest thickness as measured with chest calipers, with 47 cases measuring ≤ 26 cm and 8 cases measuring > 26 cm (Table II). Results grouped according to PA chest thickness are shown in Fig. 6. A statistically significant boost was evident for DE imaging in the detection and characterization of lesions in both categories (viz., p -value < 0.0001 for

“average” and p -value <0.002 for “thick”). The curves indicate a fairly uniform improvement in diagnostic satisfaction for cases of “average” thickness (i.e., a uniform boost across all ratings), whereas for the “thick” cases, the curves appear to suggest a boost at the higher ratings (i.e., more conspicuous lesions). This observation, along with the increased p -value, suggests that the benefit of DE images may be somewhat less in “thick” than in “average” cases, presumably because images for the former are limited, at least in part, by quantum noise, contrast, x-ray scatter, etc., rather than anatomical clutter.

E. Stratification by Gender

Cases were further grouped based on gender – a total of 31 male and 24 female. Results in Fig. 7 suggest a significant improvement in diagnostic performance for DE in each case (p -value <0.001 for male and p -value = 0.013 for females). Such is likely consistent with the trend for improved performance overall for DE (Fig. 3). The smaller level of improvement suggested for the female sub-cohort (while still a statistically significant improvement) is possibly related to that observed for larger chest thickness (in this case, breast tissue), which correlated with a smaller improvement in diagnostic performance.

F. Stratification by Region

Cases stratified according to 7 regions of the chest are listed in Table II. Results grouped according to 7 regions of the chest (left and right apex, left and right middle, left and right lower, and mediastinal regions) are shown in Fig. 8. A significant improvement in diagnostic performance is observed for DE imaging in the left apex, right apex, left-middle, and left-lower regions. The results for the left and right apex regions are consistent with the hypothesis that DE imaging improves diagnostic quality by removing anatomical noise – in this case, the clavicles and 1st and 2nd ribs, which pose complex anatomical clutter and can significantly diminish conspicuity. The significant improvement observed for DE imaging in the left-lower region is somewhat surprising, given that this region is challenged by a preponderance of soft-tissue structures (the heart) and is most susceptible to cardiac motion artifacts.

IV. DISCUSSION AND CONCLUSIONS

DR imaging of the chest represents a cost-effective, widely available, low-dose modality employed for a broad spectrum of applications, ranging from bedside exams to initial examination and diagnosis of lung disease. Still, it is known to suffer in sensitivity for detection of subtle lesions, limited primarily by a lack of conspicuity caused by superimposed anatomical structures. DE imaging, which reduces anatomical clutter by selectively removing material-specific components from the image, showed a significant improvement in the satisfaction associated with the detection and characterization of pulmonary nodules. The results were based on an initial evaluation in 55 patients drawn from an ongoing trial, providing initial investigation of diagnostic performance and supporting the hypothesis that DE imaging boosts lesion conspicuity. That this performance improvement is achieved without increase in radiation dose is particularly encouraging (i.e., the total dose for the DE image is equivalent to that of a single PA DR image). We attribute the boost in diagnostic satisfaction primarily to the nature of DE imaging itself (i.e., reduction of anatomical noise) but also to a system design in the experimental prototype based on first principles of performance metrology (contrast, noise, noise-equivalent quanta, and task-dependent detectability) and system optimization (e.g., differential added filtration, selection of kVp pair, and optimal allocation of dose between low- and high-energy projections).¹³⁻¹⁷ Furthermore, the availability of a composite radiograph might boost the overall performance of a DE imaging system. Because the composite image presents information comparable to a conventional radiograph, the DE imaging system could present the radiologist with the familiar context of the (composite)

radiograph, within which s/he could examine tissue-specific characteristics via the soft-tissue and/or bone images.

The differences in satisfaction between DE and DR imaging are shown in Figs. 3-8, with qualitative differences illustrated in the patient images of Figs. 9-11. For example, the difference observed for small, solid nodules (<3 cm diameter, ≥ 20 HU density) in Figs. 4-5 is illustrated qualitatively in Fig. 9. The improved visualization of small, solid nodules by virtue of rib cancellation (particularly rib crossings) is clear. Figures 10-11 similarly illustrate the improved visibility in the lung apex, where the clavicles and first and second ribs present confounding clutter that is significantly reduced in the DE soft-tissue image. That the difference in performance was greatest for small, solid nodules located in the apex is particularly valuable, since these characteristics represent precisely the most challenging (i.e., small), important (i.e., solid, more likely malignant), and frequent (i.e., location in lung apex) cases and the areas most in need of improvement in chest imaging.

The initial results are encouraging, although the study is not without its limitations. First, the number of cases is low – particularly for certain stratifications (e.g., anatomical region) – as these early data constitute an initial study from an ongoing trial. Furthermore, the stratifications within the data are *post-hoc*, and while the overall study was prospectively designed to investigate the difference in performance between DE and DR imaging, the data pertaining to individual strata (i.e., nodule size, nodule density, chest thickness, gender, and nodule location) should be considered ‘hypothesis-generating’ in the sense of retrospective analysis. Furthermore, the strata are distinct and independent (e.g., grouping “small” or “solid” or “apical” nodules, but not “small, solid, apical” nodules), and joint grouping was not examined.

Second, the satisfaction tests based on a 9-point scale is a fairly coarse method. For example, one may justifiably wonder whether such a test comparing DR to screen-film would be sufficiently sensitive to demonstrate a statistically significant difference. That a performance improvement was, in fact, observed for DE imaging in this study suggests a considerable improvement in conspicuity. Statistical significance was evaluated in terms of p-values obtained from the Wilcoxon signed rank test. A possible limitation associated with clustering effects within ratings for the same patient across observers will be investigated with the benefit of more cases and could be addressed using a modified version of the Wilcoxon signed rank test for clustered data in future work.²³ A more comprehensive characterization of diagnostic sensitivity and specificity for DE imaging would include evaluation of the receiver-operating characteristic (ROC) curve, to be investigated in future work.

A third limitation of the study involves the DE image processing and decomposition. While the DR images were post-processed according to techniques established for clinical DR imaging that were optimal, the DE image post-processing was fairly simple. The DE images used in the observer study used the simple image processing and decomposition techniques illustrated in Fig. 1 – specifically, single-point offset and gain corrections, non-optimized registration of low- and high-energy projections, and simple log-weighted subtraction according to a scalar tissue cancellation parameter. Future work will improve upon each of these as well as optimal rendering, tissue equalization, and edge enhancement and presumably further increase the performance of DE imaging. In addition, noise reduction algorithms (e.g., anti-correlated noise reduction, smoothing of the high-energy image, etc.) will be implemented in the decomposition of images for tests using the full patient cohort.

The clinical role of DE imaging of the chest is yet to be fully determined. The studies herein relate only to a PA exam. Whether clinical implementation would utilize a lateral DE or DR image is under investigation. Furthermore, the role of DE with respect to volumetric imaging modalities – such as tomosynthesis and CT – are subject to future analysis of diagnostic

sensitivity, specificity, workflow, cost effectiveness, and clinical implementation. Also the effect of utilizing DE as an adjunct to DR will be of clinical interest, and evaluated in the future study.

Although DE imaging requires assessment of a greater number of images compared to DR imaging, this will be compensated for by an increased level of diagnostic confidence in lesion detection which, in turn, should translate into earlier detection of disease. In addition, the characteristics of DE imaging offer promise in other areas of thoracic disease such as earlier detection of airspace disease (pneumonia in patients with fever of unknown origin), improved demonstration of airways disease (bronchiectasis in patients with chronic productive cough) and improved visualization of catheters, tubes and pneumothoraces in the ICU patient.²⁴

DE imaging – with diagnostic performance exceeding that of DR, yet at equivalent radiation dose – could represent a new normal means of chest projection imaging, since an image equivalent to DR may always be decomposed, but more importantly, the soft-tissue image presents subtle lesions more conspicuously by virtue of reduced anatomical clutter. While the observer response (rating) pertained to the combined DE image set (i.e., the soft-tissue and bone image considered together, rather than each rated individually), it was clear that the soft-tissue image was the more important in nodule detection, while the bone image presented complementary information regarding characterization (e.g., calcification). For this patient cohort in particular (drawn from a clinical patient population referred for a lung nodule biopsy), there were few cases exhibiting calcified nodules; therefore, the bone image was likely utilized less than in a general screening population (in which the frequency of calcified nodules would presumably be greater). Furthermore, the bone image could provide diagnostic value regarding bone differentiation of rib metastasis from fracture, and improved visualization of fine bony detail to exclude cortical invasion. Future studies will also examine the use of DE imaging as an adjuvant to low-dose CT in lung nodule evaluation, maintaining the high-sensitivity associated with low-dose CT, but possibly improving upon specificity through better quantification of subtle nodule calcification using DE imaging.

REFERENCE LIST

1. Kido S, Kuriyama K, Kuroda C, Nakamura H, Ito W, Shimura K, Kato H. Detection of simulated pulmonary nodules by single-exposure dual-energy computed radiography of the chest: Effect of a computer-aided diagnosis system. *Eur J Radiol* 2002;44(3):205–209. [PubMed: 12468069]
2. Niklason L, Hickey N, Chakraborty D, Sabbagh E, Yester M, Fraser R, Barnes G. Simulated pulmonary nodules: Detection with dual-energy digital versus conventional radiography. *Radiology* 1986;160(3):589–593. [PubMed: 3526398]
3. Fraser R, Barnes G, Hickey N, Luna R, Katzenstein A, Alexander B, McElvein R, Zorn G, Sabbagh E, Robinson C. Potential value of digital radiography: Preliminary observations on the use of dual-energy subtraction in the evaluation of pulmonary nodules. *Chest* 1986;89(4 Suppl):249S–252S. [PubMed: 3514175]
4. Tagashira H, Arakawa K, Yoshimoto M, Mochizuki T, Murase K. Detectability of lung nodules using flat panel dual energy subtraction by two shot method: Evaluation by ROC method. *Eur J Radiol* 2007;64(2):279–284. [PubMed: 17386991]
5. Kundel H, Revesz G. Lesion conspicuity, structured noise, and film reader error. *AJR Am J Roentgenol* 1976;126(6):1233–8. [PubMed: 179387]
6. Samei E, Flynn MJ, Eyler WR. Detection of subtle lung nodules: Relative influence of quantum and anatomic noise on chest radiographs. *Radiology* 1999;213(3):727–734. [PubMed: 10580946]
7. Brody WR, Butt G, Hall A, Macovski A. A method for selective tissue and bone visualization using dual energy scanned projection radiography. *Med Phys* 1981;8(3):353–357. [PubMed: 7033756]
8. Fraser R, Hickey N, Niklason L, Sabbagh E, Luna R, Alexander C, Robinson C, Katzenstein A, Barnes G. Calcification in pulmonary nodules: Detection with dual-energy digital radiography. *Radiology* 1986;160(3):595–601. [PubMed: 3526399]

9. Kelcz F, Zink F, Peppler W, Kruger D, Ergun D, Mistretta C. Conventional chest radiography vs. Dual-energy computed radiography in the detection and characterization of pulmonary nodules. *AJR Am J Roentgenol* 1994;162(2):271–278. [PubMed: 8310908]
10. Ricke J, Fischbach F, Freund T, Teichgraber U, Hanninen E, Rottgen R, Engert U, Eichstadt H, Felix R. Clinical result of CsI-detector-based Dual-exposure Dual-energy subtraction chest radiography. *Euro Radiol* 2003;13:2577–2582.
11. Ide K, Mogami H, Murakami T, Yasuhara Y, Miyagawa M, Mochizuki T. Detection of lung cancer using single-exposure Dual-energy subtraction chest radiography. *Radiat Med* 2007;25:195–201. [PubMed: 17581707]
12. Uemura M, Miyagawa M, Yasuhara Y, Murakami T, Ikura H, Sakamoto K, Tagashira H, Arakawa K, Mochizuki T. Clinical evaluation of pulmonary nodules with Dual-exposure Dual-energy subtraction chest radiography. *Radiat Med* 2005;23(6):391–397. [PubMed: 16389980]
13. Shkumat NA, Siewerdsen JH, Dhanantwari AC, Williams DB, Richard S, Paul NS, Yorkston J, Van Metter R. Optimization of image acquisition techniques for dual-energy imaging of the chest. *MedPhys* 2007;34(10):3904–15.
14. Shkumat NA, Siewerdsen JH, Richard S, Paul NS, Yorkston J, Van Metter R. Dual-energy imaging of the chest: Optimization of image acquisition techniques for the ‘bone-only’ image. *Med Phys* 2008;35(2):629–32. [PubMed: 18383684]
15. Richard S, Siewerdsen JH. Optimization of dual-energy imaging systems using generalized NEQ and imaging task. *Med Phys* 2007;34(1):127–139. [PubMed: 17278498]
16. Richard S, Siewerdsen JH, Jaffray DA, Moseley DJ, Bakhtiar B. Generalized DQE analysis of radiographic and dual-energy imaging using flat-panel detectors. *Med Phys* 2005;32(5):1397–1413. [PubMed: 15984691]
17. Richard S, Siewerdsen JH. Cascaded systems analysis of noise reduction algorithms in dual-energy imaging. *Med Phys*. 2007 Accepted for publication; in press
18. Gohari A, Haramati LB. Complication of CT scan-guided lung biopsy. *Chest* 2004;126:666–668. [PubMed: 15364737]
19. Boone JM, Shaber GS, Tecotzky M. Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician* 1998;52(2):119–126.
20. Dhanantwari A, Siewerdsen J, Shkumat N, Williams D, Richard S, Moseley D, Paul N, Yorkston J, Van Metter R. Multi-Resolution, multi-scale mutual information technique for registration of a high- and low-kVp projections in dual-energy imaging. *Med Phys* 2007;34(6):2517.
21. Van Metter R, Foos D. Enhanced latitude for digital projection radiography. *Proc SPIE Conf Image Display* 1999;3658:468–483.
22. Agresti A, Coull BA. Dual-energy mammography: A detector analysis. *Med Phys* 1990;17:665–675. [PubMed: 2215412]
23. Rosner B, Glynn R, Lee MT. The Wilcoxon signed rank test for paired comparisons of clustered data. *Biometrics* 2006;62:185–192. [PubMed: 16542245]
24. Kuhlman JE, Collins J, Brooks GN, Yandow DR, Broderick LS. Dual-energy subtraction chest radiography: what to look for beyond calcified nodules. *Radiographics* 2006;26(1):79–92. [PubMed: 16418245]

Figure 1.

Illustration of Dual Energy (DE) image processing, registration, decomposition, and display. Notation includes: the low-energy projection (LE); the high-energy projection (HE); the soft-tissue decomposition (Soft); the bone decomposition (Bone); and the composite equivalent radiograph (Comp). Processing steps include: offset / gain / defect correction; deformable registration (Reg); weighted log-subtraction decomposition (Decomp); and transformation of pixel values to log-exposure space for display. The photograph at right shows the experimental setup for observer studies, with images displayed on the left and the rating scale at the right.

Figure 2. Example (a) Dual Energy soft-tissue image, (b) Dual Energy bone image, and (c) Digital Radiography image. The arrow marks the location of a nodule (2.9 cm, 35.2 HU), confirmed as squamous cell carcinoma by needle biopsy. The white streak around the heart in the bone image is an example of cardiac motion artifact.

Figure 3.

Diagnostic satisfaction in DE (DualEnergy) and DR (Digital radiography) image readings (all cases pooled). **(a)** Percentage of image pairs for which the DE image was rated superior, equal, or inferior to the DR image. **(b)** Percentage of image pairs for which the DE image was rated superior, equal, or inferior to the DR image as agreed by 3 or more observers. **(c)** Fraction of observer responses at or above a given rating score. The error bars represent a 95% confidence interval.

Figure 4. Diagnostic satisfaction in DE (Dual energy) and DR (Digital radiography) image readings (stratified by lesion size). The curves show the fraction of responses at or above a given rating score for (a) lesion size ≤ 3 cm and (b) lesion size > 3cm. The error bars reflect a 95% confidence interval.

Figure 5. Diagnostic satisfaction in DE (Dual energy) and DR (Digital radiography) image readings (stratified by lesion density). The curves show the fraction of responses at or above a given rating score for (a) non-solid lesions ($<20\text{HU}$) and (b) solid lesions ($\geq 20\text{HU}$). The error bars reflect a 95% confidence interval.

Figure 6. Diagnostic satisfaction in DE (Dual energy) and DR (Digital radiography) image readings (stratified by chest thickness). The curves show the fraction of responses at or above a given rating score for (a) thickness ≤ 26 cm and (b) thickness > 26 cm. The error bars reflect a 95% confidence interval.

Figure 7. Diagnostic satisfaction in DE (Dual energy) and DR (Digital radiography) image readings (stratified by patient gender). The curves show the fraction of responses at or above a given rating score for (a) male and (b) female. The error bars reflect a 95% confidence interval.

Figure 8.

Diagnostic satisfaction in DE (Dual energy) and DR (Digital radiography) image readings (stratified by lesion location). The curves show the fraction of responses at or above a given rating score for lesion located in (a) Right-Upper, (b) Left-Upper, (c) Right-Middle, (d) Left-Middle, (e) Right-Lower, (f) Left-Lower, and (g) Mediastinal regions. The error bars reflect a 95% confidence interval.

Figure 9.

Illustration of (a) Dual energy soft-tissue, (b) Dual Energy bone, and (c) DR (Digital radiography) images for a case exhibiting a small, solid nodule (0.8 cm, 58.4 HU, Benign Lymphoid tissue) marked by the arrow. The bone image exhibits some residual soft-tissue owing to incomplete cancellation as well as motion artifact, which is more obvious in this case because of the cardiac motion artifact.

Figure 10.

Illustration of (a) Dual Energy soft-tissue, (b) Dual Energy bone, and (c) DR (Digital radiography) images for a case exhibiting a nodule (2 cm, 67.5 HU, *Adenocarcinoma*) in the right upper lobe marked by the arrow.

Figure 11. Illustration of (a) Dual Energy soft-tissue, (b) Dual Energy bone, and (c) DR (Digital Radiography) images for a case exhibiting a nodule (2.8 cm, 35.1 HU, *Adenocarcinoma*) in the left apex marked by the arrow.

Table 1

Satisfaction rating scale and descriptions. The Table was displayed to observers on a second monitor during all tests as shown in Fig. 1.

Score	Rating	Description
9	Very Satisfied	The abnormality is perfectly obvious and easily characterized.
8		
7	Satisfied	The abnormality is visible and can be well characterized.
6		
5	Neither Satisfied nor Dissatisfied	The abnormality is reasonably well seen and characterized.
4		
3	Dissatisfied	The abnormality is visible, but detection and characterization of subtle features are a bit challenging .
2		
1	Very Dissatisfied	The abnormality could be overlooked or mischaracterized.

Table 2

Summary of cases and stratifications of the data. The p-values relate to the statistical significance in the difference between DE and DR satisfaction, regarded as significant for p-value < 0.05.

	Description:	N	N readings	p-Value
All pooled		55	275	<0.001
Lesion Size	≤3 cm	36	180	<0.001
	>3 cm	19	95	=0.0264
Lesion Density	<20 HU	9	45	=0.0968
	≥20 HU	46	230	<0.001
Chest Thickness	≤26 cm	47	235	<0.001
	>26 cm	8	40	=0.0015
Gender	Male	31	155	<0.001
	Female	24	120	=0.0137
Region	Right-Upper	20	100	=0.0012
	Left-Upper	5	25	=0.0067
	Right-Middle	4	20	=0.1800
	Left-Middle	2	10	=0.0313
	Right-Lower	8	40	=0.3625
	Left-Lower	9	45	=0.0010
	Mediastinum	7	35	=0.2455