Impact of a Computer-Aided Detection (CAD) System Integrated into a Picture Archiving and Communication System (PACS) on Reader Sensitivity and Efficiency for the Detection of Lung Nodules in Thoracic CT Exams

Luca Bogoni · Jane P. Ko · Jeffrey Alpert ·
Vikram Anand · John Fantauzzi · Charles H. Florin ·
Chi Wan Koo · Derek Mason · William Rom ·
Maria Shiao · Marcos Salganicoff · David P. Naidich

Published online: 19 June 2012
© Society for Imaging Informatics in Medicine 2012

Abstract The objective of this study is to assess the impact on nodule detection and efficiency using a computer-aided detection (CAD) device seamlessly integrated into a commercially available picture archiving and communication system (PACS). Forty-eight consecutive low-dose thoracic computed tomography studies were retrospectively included from an ongoing multi-institutional screening study. CAD results were sent to PACS as a separate image series for each study. Five fellowship-trained thoracic radiologists interpreted each case first on contiguous 5 mm sections, then evaluated the CAD output series (with CAD marks on corresponding axial sections). The standard of reference was based on three-reader

agreement with expert adjudication. The time to interpret CAD marking was automatically recorded. A total of 134 true-positive nodules, measuring 3 mm and larger were included in our study; with $85 \geq 4$ and $50 \geq 5$ mm in size. Readers detection improved significantly in each size category when using CAD, respectively, from 44 to 57 % for ≥ 3 mm, 48 to 61 % for ≥ 4 mm, and 44 to 60 % for ≥ 5 mm. CAD stand-alone sensitivity was 65, 68, and 66 % for nodules ≥ 3 , ≥ 4 , and ≥ 5 mm, respectively, with CAD significantly increasing the false positives for two readers only. The average time to interpret and annotate a CAD mark was 15.1 s, after localizing it in the original image series. The integration of CAD into PACS increases reader sensitivity with minimal impact on interpretation time and supports such implementation into daily clinical practice.

This study was supported by the National Institute of Health/National Cancer Institute-UO-1, CA, 86137 and NYU Biomarker, Clinical and Epidemiologic Center—Early Detection Research Network

L. Bogoni (✉) · V. Anand · C. H. Florin · M. Salganicoff
Siemens Healthcare,
Malvern, PA 19355, USA
e-mail: luca.bogoni@siemens.com

J. P. Ko · J. Alpert · J. Fantauzzi · C. W. Koo · D. Mason ·
M. Shiao · D. P. Naidich
Department of Radiology,
New York University-Langone Medical Center, Tisch Hospital,
560 1st Avenue,
New York City, NY 10016, USA

W. Rom
The Division of Pulmonary,
Critical Care and Sleep Medicine, Department of Medicine,
New York University School of Medicine,
560 1st Avenue,
New York City, NY, USA

Keywords Radiographic image interpretation · Computer-assisted · Radiography · Thoracic · PACS reading · Clinical workflow · Lung · Efficiency · Computed tomography · Computer-assisted detection · Chest CT

Introduction

Over the past decade, the potential for computer-aided detection (CAD) systems to augment radiologist's ability to detect lesions [1] and specifically solid lung nodules on multi-detector computer tomography (CT) studies of the thorax has been extensively evaluated [2]. Furthermore, such evaluations have also been performed on databases of cases, such as the Lung Image Database Consortium,

collected for the purpose of assessing CAD impact [3]. To date, nearly all studies have shown a consistent improvement in the identification of potentially significant lung nodules [4–8]. Importantly, this improvement has been shown to be relatively independent of the CAD algorithm employed [9] and CT technique including radiation dose [10], reconstruction algorithm, variations in section thickness [11], as well as differences in the level of reader experience [12]. Additionally, the value of CAD systems is well documented in a variety of clinical environments [13].

Despite the established advantages of CAD systems for improving radiologists' performance, the use of CAD has yet to be widely established as a component of routine clinical practice [14]. While undoubtedly multi-factorial, a major contributing reason is the need to rely on stand-alone workstations rather than being incorporated into picture archive and communications systems (PACS) allowing for routine usage. In fact, there are clear benefits to employing advanced workstations since, in addition to detection, most CAD systems also integrate sophisticated visualization and measurement tools that simultaneously allow readers to perform 3D nodule segmentation, volumetric assessment of size and density, and even the ability to automatically identify nodules on sequential CT examinations [14]. However, there are many clinical scenarios where such advanced functionalities are not required yet patient images must be reviewed for the presence of lung nodules.

The present study avails itself of pre-computed CAD outputs as a mean of integration. Newer directions in workflow management advocate, in fact, the incorporation of content-based image retrieval as a means to support CAD review in PACS [15] and further evolve into fully hospital-integrated solutions [16].

Thus, the inconvenience of needing to employ two separate reading stations (PACS and a CAD workstation) in order to interpret studies using CAD has limited its routine utilization. One obvious solution would be to seamlessly integrate CAD into existing PACS [17, 18]. Anh et al. [19] described a toolkit which implements a solution, similar to that used in this study, for the integration of DICOM captures into the PACS and hence allow the review of CAD findings directly on the PACS. Huang et al. [20] provide a comprehensive view of the various aspects for CAD integration. Unfortunately, at present the integration of a CAD system, employing the entire broad range of CAD interactive functionalities into currently available PACS environments, has proved technically challenging [14]. The available approaches for the incorporation of a CAD system into PACS clinical workflow focus primarily on presenting readers with candidate nodules as DICOM captures; thus, potentially routinely enhance lung nodule detection. This has taken on greater potential importance with the long anticipated report of the National Lung Cancer Screening (NLST) study documenting a 20 % reduction in

disease specific mortality with low-dose CT screening [21]. While, Roos et al. have also assessed the time to evaluate CAD marks as either true (TP) or false positive (FP) [22], to our knowledge the efficacy of CAD as a second reader and the effect on interpretation time have not been evaluated in a PACS settings.

The present study aims at exploring the effect on reader sensitivity and interpretation time for the detection of solid nodules using a commercially available CAD system that is integrated into the clinical PACS workflow environment; specifically addressing its use in a second-reader paradigm in the interpretation of thoracic CT studies.

Material and Methods

The thoracic screening CT cases, included in this Institutional Review Board approved retrospective study, were collected between January and March 2007 as part of the Early Diagnostic Research Network study. The CT scans, without intravenous contrast administration, were performed beginning at the lung apices and ending in the upper abdomen on a multi-detector CT scanner. The studies were acquired using $2 \times 32 \times 0.6$ detector configuration (CT Sensation Definition, Siemens Healthcare Forchheim, Germany), with gantry rotation time of 0.33 s, tube potential of 120 kV, with angular and longitudinal tube current modulation (CareDose4d) using a reference of 40 mAs. The CT data were reconstructed into axial 1-mm sections at 0.8-mm intervals for the "thin sections" and 5 mm images every 5 mm for the "thick sections", using a high-frequency kernel (B60).

Based on previous evaluations of our CAD algorithm reported in the literature, we determined that a 10 % difference in sensitivity would have been statistically significant [23–25]. Hence, assuming a CAD stand-alone sensitivity of 70 %, and three nodules per patient on average, it was determined that the study would need at least 40 cases to achieve a 95 % confidence interval of ± 10 %. Thus, we included 48 consecutive studies. Five were rejected, following expert review (D.P.N.), due to the finding of diffuse lung disease, while all others (43) met standard criteria for image quality. In order to mirror the normal clinical practice at our institution, only thick sections (5 mm) were assessed by readers for nodules, to determine the impact of an integrated CAD system in a workflow in which thick chest CT axial sections are routinely evaluated on a PACS system. All data used in this study were anonymized.

CAD Server

The CAD server (*syngo* CAD Manager 2008B—not a medical device, Siemens Healthcare, Malvern PA), employed in this study, processed the thin CT images acquired data as

part of a workflow which enabled the results to be viewed directly on the PACS (Fig. 1). The workflow entailed simultaneous transmission of data following completion of a thoracic CT scan to both PACS as well as to a separate computer server for processing. The CAD algorithm (described below) generated findings which were then presented as short series of derived DICOM image objects (Fig. 2) and automatically sent to PACS. The length of this series depended on the number of candidates identified by CAD. Each candidate was displayed as a circle, centered on the finding, on its corresponding axial image. When CAD yielded no candidates for any one case, a message stating “no CAD findings” was overlaid on a single image and output as a series.

CAD Algorithm

CAD algorithms in general and CAD for lung nodules in particular, have been gaining clinical acceptance over the past 15 years. These products are supported by intense research work in the field of Computer Vision and Pattern Recognition [29–32]. The current version of *syngo* LungCAD (FDA clearance K063877) algorithm for nodule detection (Lung CAD VC20A, Siemens Healthcare, Malvern, PA) was included in the *syngo* CAD Manager server, as mentioned above. The algorithm was designed as a multi-step approach with the goal of detecting lung lesions with high sensitivity and specificity focusing on solid lesions greater than 3 mm. The algorithm [33] trained on 330 data sets and tested in over 747 data sets collected from multiple institutions to include various chest CT acquisition protocols and scanners. The algorithm was decomposed into four stages: lung segmentation, candidate generation, feature computation, and classification. The first

three stages combine both image processing and embedded machine learning approaches, while the classification stage (fourth stage) is entirely based on machine learning techniques. The four stages are performed as follows:

1. The individual slices are sorted and transformed to a canonical volume representation. The lungs contours are, then, automatically delineated by first detecting regions in the thorax characterized by air density. The collection of these regions is then processed using a series of morphological operations for filling holes to create a binary mask that covers the entire lung including vascular structures and soft tissues. In order to speed up the CAD processing, and avoid marking nodule-like structures outside the lung parenchyma, CAD processes only that region of the image.
2. Candidate generation aims at achieving high sensitivity while keeping the number of candidates to a manageable number. Initially, a combination of basic Computer Vision techniques such as Gaussian filtering and intensity value thresholding is applied to identify candidate locations. For instance, the diverging gradient field response. Diverging gradient field response (DGFR) algorithm [34] is applied to identify candidate locations. The DGFR algorithm is designed to identify nodular structures surrounded by an area of lower attenuation. Since the computation of the maxima is applied relative to its surrounding, it does not rely on arbitrary Hounsfield Units. After this step as many as 300 candidates per volume are identified. In a second step, a cascading classifier reduces the number of candidates to around 80 per volume using the features already computed by the DGFR algorithm. These include volumetric shape characteristics, intensity distributions, and compactness.
3. Feature extraction gathers image-based discriminative features for each candidate identified in the previous stage. This information is used by the next stage (classification) to label each candidate as either a TP or a FP. The computed features can be categorized into several groups: (a) those indicative of voxel attenuation distributions within the candidate [35], (b) those pertaining to the candidate’s shape and curvature [36], and (c) those that describe the candidate’s edge and margins (computed as part of a volumetric segmentation of the candidate [37, 38]). These features capture properties that can be used to disambiguate true lesions from typical FP. Some of the typical FP, mimicking true nodules, are caused by pleural thickening, connective tissues between vessels, partial volume, etc.
4. Classification, based on data mining approaches, utilizes the multiple instance learning-based classification [39] that performs automatic feature selection and classifier design jointly. During the training stage of the

Fig. 1 Diagram outlining how syngo CAD Manager (CAD server) is integrated into the PACS and Radiology Information Network workflow. Following initial scan acquisition, data are simultaneously channeled to both the PACS and to the CAD server, which in turn outputs findings to the PACS as a separately labeled CAD series. These are immediately available along with routine PACS images as part of the normal clinical workflow

Fig. 2 *Left* CAD server output with candidate marker embedded directly into pixel data of the DICOM-derived image. *Right* detail illustrating one of five markers identified for this case

classifier, medical domain knowledge regarding nodule appearance, shape and sizes is incorporated. It takes into account the information about nodules to construct different type of classifiers with separate decision boundaries that relate to the individual nodule characteristics (e.g., small and large nodules). It uses gating as a means to learn automatically meaningful clusters among the candidates and construct classifiers, one for each cluster. The basic idea of the approach is that of decomposing a complicated task into multiple simple and tractable subtasks. The model consists of several domain experts and a gating network that decides which experts are most trustworthy on any input pattern [40]. In other words, by recursively partitioning the feature space into sub-regions, the gating network probabilistically decides which patterns fall in the domain of expertise of each expert. Thus, a final weight is assigned to any one mark determining those that are presented to the reader for review.

Study Design

All CT studies were evaluated by five fellowship-trained chest radiologists (J.A, J.F, J.K, C.W.K., and D.M.-ranging in experience from one to ten years) in a randomized order on a commercially available PACS system (Siemens Healthcare, Erlangen, Germany) for the study. A training session was performed with each reader by a study coordinator. The session included three additional chest CT cases for reader training and practicing the protocol for interpretation of study CTs, which included the use of a Case Report Form (CRF) software application. The CRF, installed on a computer adjacent to the PACS reading station, was utilized for randomizing case order, instructing readers, collecting interpretation time, and compiling nodule characteristics. The readers first interpreted forty-three contiguous 5 mm studies with the specific intent of identifying all solid, non-calcified lung nodules equal to or greater than 3 mm in diameter. The CRF guided the reader to load a specific case, identify and mark the nodules on the CT

series. Nodule size, whether solid or calcified, and, in the case of CAD, the type of FP, were annotated by the reader using the CRF tool.

Upon completion of reviewing images for each case, each reader examined the CAD series. The series, pre-computed on the CAD server, was available on the PACS (Fig. 1) as an additional series for each study. Each CAD mark was evaluated by the readers following a process which included the following four stages (Fig. 3). First, the readers opened the CAD series to determine whether any marks had been generated by the CAD device (stage 1, “Start-up time/CAD mark present”). Secondly, if any CAD

Fig. 3 CAD marks evaluation process: stages of evaluation

mark was present, the readers located each CAD mark manually on the corresponding contiguous 5 mm CT image data set, previously interpreted by the reader, to better evaluate the finding (stage 2, “Reader Found”). Thirdly, if the CAD mark did not correspond to a structure initially identified, the reader then decided whether it could be easily dismissed as insignificant finding such as a calcified nodule or small finding ≤ 3 mm or an easy-to-dismiss FP, such as a hilar vessel or bony excrescence (stage 3, “Insignificant or Easy FP”). Subsequently, each remaining CAD mark was thus evaluated and assessed by the reader as either a clinically significant nodule (TP) or not a nodule (FP), such as a small vessel or a non-rounded focal abnormality (stage 4, “Significant—TP/FP”). Any CAD mark deemed by a reader to be a new finding was characterized using the same criteria employed for nodules initially identified by the readers alone. The CRF required that all information for any finding be completely entered for both the interpretation of the data alone and also the CAD marks prior to allowing the reader to enter details for the next mark within a case or continue on to the next case. The CRF recorded the timing information at the beginning of interpretation of a CT examination, the end of reader interpretation, and the completion of the evaluation of the CAD marks pertaining to the CT scan.

The marks from all of the readers, alone and with CAD, were consolidated as follows: any marks from two or more readers within a maximum clustering distance were automatically consolidated into one single mark. This clustering distance was defined to be 5 mm or the largest of either nodule’s diameters if any exceeded 5 mm.

Following automated consolidation, a gold standard for all nodules was subsequently established based on majority agreement plus adjudication. All findings independently identified, either by the reader alone or with the aid of CAD, and labeled as nodules by at least three readers were entered as true nodules in the reference standard. All findings identified, either by the reader alone or with the aid of CAD, and labeled as nodules by less than three readers were adjudicated, either as TP or FP, by an expert chest radiologist (D.P.N.) with 30 years of experience. All other CAD findings were assigned as a truth value of FP and confirmed by the expert reader; furthermore, they were labeled according to various FP categories as discussed below. The same rule was also applied to determine nodules’ attenuation, solid or calcified. Nodule’s size was taken as the average of the sizes for each nodule identified by at least three readers, and measured by the expert otherwise.

Statistical Analysis

Reader sensitivity was defined as the number of correctly identified nodules divided by the number of TP lesions included in the gold standard. The per-case FP

rate was defined as the number of reported marks that were not associated with TP nodules divided by the number of CT studies. The difference in sensitivity between the “reader alone” and reader with CAD as a second reader (“reader+CAD”) was tested with the one-tailed McNemar test with p value < 0.05 and with a null hypothesis that the two systems were dependent.

As part of the analysis, both the average time used by readers when reviewing the cases alone and evaluation of CAD marks were analyzed. In addition, the time for each of the stages discussed above, which occurred in the process of evaluating a single CAD mark was studied. However, we observed a large variation in the time to review the CAD marks among some of the readers. This variation was probably caused by two factors. First, the readers did not conduct reading sessions using dedicated workstations, rather the interpretation was performed at workstations utilized for clinical practice, and hence variations from environment cannot be excluded. The second, and main factor, was the use of the CRF for measuring interpretation times which were actively clicked by the reader. Hence, inadvertent failures to begin and stop the CRF could not be avoided. Thus, to address this aspect we made the assumption that any one reader would have been able to review and grade a CAD mark within thirty seconds. This threshold was determined by analyzing the time used to review the CAD findings across all readers. Thus, first the average of per mark review time was computed for each case and each reader as: the time to review all CAD marks in a case by a reader divided by the number of marks in that specific case. Then the median of these values (27 s) was then rounded to 30 s to add additional buffer. Thus, the set of cases for the next stage of the analysis were identified.

The time for each four stages for the evaluation of CAD mark was estimated by using a robust, least-square linear fitting algorithm (RANSAC, MatLab V. 7.8 Mathworks) and solving an over-constrained set of linear equations. Hence from the original number of 215 CT evaluation (43 CT studies each evaluated by five readers) 41 % (88/215) were not used in the estimation of the timing calculations. Of these, 12 % (26/215) rejected because violating the first assumption and additional 29 % (62/215) following the second step above. Thus, the resulting in 127 CT studies were utilized to compute (a) the average time for the evaluation of the cases by the “reader alone” and “reader+CAD” as well as (b) the average time for each stage taken by a reader to review a CAD mark. The results thus obtained represent an estimate of the time taken by the readers overall as well as for each of these separate steps when evaluating a CAD mark. Sensitivities and timing results are presented using percentages and 95 % confidence intervals (CI).

Results

Nodule Detection Sensitivity

A total of 134 TP nodules, measuring 3 mm and larger ($49 \geq 3$ and <4 , $35 \geq 4$ and <5 , and $50 \geq 5$ mm; mean, 4.9 mm) were included in our study (accounting for 39 of the 43 chest CTs). Four of the 43 chest CTs had no solid nodules ≥ 3 mm. CAD significantly improved detection of solid nodules regardless of nodule size ($p < 0.05$) (Table 1). Reader sensitivity improved from: 44 % (296/670; 95 % CI, ± 8 %) without to 57 % (381/670; 95 % CI, ± 8 %) with CAD for ≥ 3 mm nodules ($n = 134$ nodules $\times 5$ readers; $p < 0.001$), 48 % (203/425; 95 % CI, ± 10 %) without to 61 % (258/425; 95 % CI, ± 10 %) with CAD for ≥ 4 mm nodules ($n = 85 \times 5$ readers; $p < 0.01$), and 44 % (110/250; 95 % CI, ± 13 %) without to 60 % (149/250; 95 % CI, ± 13 %) with CAD ≥ 5 mm nodules ($n = 50 \times 5$ readers; $p < 0.03$), respectively. On average, 1.44 additional nodules per case were identified and accepted by the readers after reviewing the CAD marks (range, 0–4). CAD stand-alone sensitivity was 65 (87/134), 68 (58/85), and 66 % (33/50) for nodules ≥ 3 , ≥ 4 , and ≥ 5 mm, respectively (Table 1; Fig. 4).

The number of identified nodules, without regard to adjudication, decreased as a function of the number readers agreeing. Namely, when considering nodules ≥ 3 mm, for example, this number decreased from 142, when defined by only one reader to 28 when all five readers had to agree (Table 2).

The impact of CAD was more pronounced with respect to cases which were initially identified as negative (without nodules) by the reader prior to the use of CAD. In fact, each of the five readers re-interpreted the CT studies as positive following the evaluation of CAD marks: five studies both for readers 1 and 2, one case for reader 3, and two cases for readers 4 and 5, respectively. The average diameter for the additional nodules was 4.6 mm with six nodules ranging in size between 3 and 4 mm, four between 4 and 5 mm, and eight between 5 and 6.5 mm.

False Positives

The evaluation of CAD findings significantly increased the readers FP interpretations on average per/case ($p < 0.001$) from 0.07 (95 % CI, ± 0.04 (range, 0.02–0.14)) to 0.15 (95 % CI, ± 0.08 (range, 0.05–0.28)) after using CAD. For all the cases, the average number of FPs increased from 2.8 (95 % CI, ± 1.5 (range, 1–6)) to a total of 6.4 (95 % CI, ± 2.5 (range, 2–12)) when using CAD. The total number of FPs increased significantly after reviewing CAD marks, for two of readers, from 1 to 8 and from 3 to 12; for one reader, it increased from 2 to 4, whereas the other two accepted no additional FPs (at 2 and 6, respectively). Importantly, none of the patients, who otherwise would have been interpreted as having a normal study, were incorrectly classified as having a potentially significant nodule following the review of CAD proposed findings.

The CAD algorithm generated a total of 206 marks. Following the review by the expert chest radiologist, 42 % (87/206) were included in the reference standard as TP ≥ 3 mm, 23 % (48/206) were calcified, 10 % (20/206) were true nodules below 3 mm in diameter, and 25 % (51/206) were other types of FPs (at average of 1.19 per patient). The 25 % other FPs were further characterized in the following four groups: pulmonary vessels, 29 % (15/51), scars, 33 % (17/51), peribronchial and mucoid impactions, 10 % (5/51), and extra-pulmonary—chest wall, sub-pleura, mediastinum, heart, etc., 27 % (14/51) (Fig. 5). Scars were defined as predominantly linear opacities instead of true lung nodules. The majority of the accepted FPs 67 % (8/12) proved to be pulmonary vessels, followed by peribronchial findings, 17 % (2/12), scars, 8 % (1/12), and extra-pulmonary, 8 % (1/12). Finally, it is worth emphasizing that of the 51 FP CAD marks proposed to the readers, 39 (76 %) were either easily labeled as FP, representing either pulmonary vessels perifissural or chest wall scars, or were easily dismissed as either lying outside the lungs or superimposed over the trachea.

Table 1 Average readers' sensitivities without and with CAD, and stand-alone CAD, are presented

Size (# nodules)	Sensitivity (detected nodules; ranges; <i>p</i> values)		
	CAD	Readers	Reader+CAD
≥ 3 mm (134)	67 % (87)	44 % (296/670; 31–60 %)	57 % (381/670; 43–72 %; $p < 0.001$)
≥ 4 mm (85)	68 % (58)	48 % (203/425; 36–67 %)	61 % (258/425; 52–75 %; $p < 0.01$)
≥ 5 mm (50)	66 % (33)	44 % (110/250; 36–62 %)	60 % (149/250; 50–72 %; $p < 0.03$)

Range shows the average sensitivity for the readers; *p* values reflect if a significant increase occurred with addition of CAD. Denominator=the total number of possible evaluations (number of nodules $\times 5$ readers); numerator=sum of detections by the five readers

Fig. 4 Sensitivities for CAD alone, readers alone, and readers using CAD, expressed according to the number of nodules with respect to increasing size. The two numbers (N_1 || N_2) adjacent to each of the reader, e.g., Reader5 (80 || 97) in the “Overall ≥ 3 mm” category, indicates the number of nodules found by the reader without CAD ($N_1=80$) and with CAD ($N_2=97$) for a sensitivity of 60 (80/134) and 72 % (97/134), respectively. Thus, in the specific example for Reader 5, 17 additional nodules were found with the aid of CAD. In the CAD-alone rows CAD (N), the number in parentheses indicates the number of nodule found by CAD in the respective category

Evaluation Time

The total time for readers to interpret each CT scan and evaluate CAD marks displayed in the CAD manager series averaged 3 m and 45 s (95 % CI, ± 51 s (range, 1 m and 4 s–10 m and 3 s)) with a median time of 3 m and 6 s. The time for reader interpretation of the CT scans alone averaged 2 m and 23 s (95 % CI, ± 27 s (53 s–5 m and 31 s); median, 2 m and 16 s). Thus, an additional average of 1 m and 22 s (95 % CI, ± 34 s (3 s–6 m and 32 s), 57 s) was required per CT

Table 2 Number of nodules in the ground truth as a function of reader agreement further classified by size

No. of readers agreeing	No. of nodules		
	≥ 3 mm	≥ 4 mm	≥ 5 mm
1	142	91	52
2	101	64	35
3	71	50	28
4	54	40	23
5	28	25	15

Fig. 5 Distribution of FP CAD markings by location: extra-pulmonary (including: chest wall, mediastinum, and other), mucoid impaction/peribronchial, pulmonary vessels, and scars (including: scar, perifissural, and subpleural)

study by readers to assess CAD marks (Fig. 6). The average time across all the CAD marks needed by the readers to evaluate each CAD mark was 24.2 s.

With respect to the various phases, when readers evaluated the CAD marks, an average “start-up time” of 8.1 s (95 % CI, ± 3.15 s (range, 3–14 s)) began when readers looked for and opened the CAD-generated series in the chest CT study and ended after they determined whether or not there were any CAD marks (stage 1). When CAD marks were present on this series, an average 18.0 s (95 % CI, ± 4.49 s (range, 9–44 s)) was spent by the readers to locate the finding that corresponded to the CAD mark in the complete chest CT series and to determine whether it had been already detected by the reader (stage 2). An additional average 3.6 s (cumulative 21.6 s; 95 % CI, ± 4.10 s (10–46 s)) were utilized for the reader to decide whether the mark warranted further review (stage 3). Then, an additional

Fig. 6 Average time taken by readers to evaluate CAD findings (see also Fig. 3 for evaluation workflow). Stage 1, “Start-up time/CAD mark present” indicates the time (8.1 s; 95 % CI, ± 3.15 s) taken to open the CAD generated series and determine whether it contained any CAD marks; stage 2, “Reader Found” (18.0 s; 95 % CI, ± 4.49 s) taken to search for a specific CAD mark (new or also found by a reader); stage 3, “Insignificant or Easy FP” (21.6 s; 95 % CI, ± 4.10 s) to reject the mark as a FP; and finally stage 4, “Significant–TP/FP” (33.1 s; 95 % CI, ± 5.70 s) to assess whether a CAD mark is a nodule

11.5 s (33.1 s; 95 % CI, ± 5.70 s (15–41 s)) were spent by the readers to assess the proposed CAD mark as either a FP or TP (stage 4). Thus, the time needed to evaluate an individual CAD mark, stages 3 and 4, averaged of 15.1 s.

Discussion

Despite being advocated by numerous studies, to date, the routine application of CAD for detecting lung nodules [4–12] has yet to gain acceptance. In addition, while the integration of CAD into a routine clinical PACS environment is gaining momentum [25–28], the interactive functionality for reviewing CAD marks within PACS is currently limited. The main purpose of the present report, therefore, has been to assess the impact on reader's identification of solid lung nodules employing a CAD server integrated into a PACS environment as a potential component of routine clinical CT interpretation. In particular, the evaluation of consecutive low-dose screening CT studies, as those included in this study, is receiving additional attention in light of the recent observation that low-dose CT screening may decrease disease specific mortality by as much as 20 % [12].

The availability of CAD results efficiently accessible as DICOM series, alongside standard clinically reconstructed images, improved sensitivity for detecting solid lung nodules while adding only minimally to the number of FP interpretations without impact at the per-patient level. In fact, the sensitivity of lung nodule detection significantly increased for all readers with the aid of CAD in our investigation. Furthermore, all these nodules, except one, proved to be greater than 5 mm in size, consistent with current guidelines for the monitoring of lung nodules recommended by the Fleischner Society [43].

Although obtaining agreement regarding the definition of a true lung nodule remains elusive [31], a clear definition as to how the reference standard is defined allows a proper assessment of performance by readers as well as CAD's impact [42]. In our study, we noted a large variation of agreement, even for non-diminutive nodules. The number of nodules ≥ 3 mm in this study, for example, varied from 142 when defined by one reader only to 28 when requiring definition by all five readers stands testament to a remarkable lack of agreement considering that all five readers were dedicated, fellowship-trained chest radiologists. In these conditions, the adjudicators' role was to mitigate the bias introduced by the decision to consider a consensus level of three readers out of five for the reference standard. Furthermore, since the objective of this study was to evaluate a CAD system in a clinical workflow of 5 mm sections, the reference standard was also established on the 5 mm reconstructions and any CAD findings not visible was deemed a FP. It is acknowledged that small nodules that perhaps were TP may not have been included in the reference

standard, however, these nodules were likely to be very small and of low clinical significance.

Similarly to prior studies [5–12], the number of FP interpretations significantly increased with CAD on average from 0.07 to 0.15 findings per case. However, since a majority of the FP CAD marks, 76 %, were easily labeled as FPs, not all FPs proved equally problematic and hence the reader did not need to spend similar amount of time in analyzing all proposed findings.

In our investigation, the CAD manager introduced an additional average of 1 m and 22 s for reader review of the CAD proposed findings per CT study, with an average of 24.2 s/mark. An average of 2 m and 23 s was utilized for the review of CT study by the readers without CAD mark evaluation. When analyzing the individual marks, the majority of the time, an average 18.0 s (74 %), was taken for localizing the proposed CAD marks on the original images. Hence, with an automated localization mechanism in PACS, this time could be minimized, thus eliminating a large source of inefficiency. The decision time, in terms of whether the CAD mark was a nodule entailed a relatively smaller time, averaging 15.1 s (stages 3 and 4), with 3.6 s utilized to easily eliminate obvious FP (stage 3) and an additional 11.4 s (stage 4) to interpret the finding as a FP or TP. Therefore, the time for interpretation of CAD marks will decrease in the era of PACS systems with localizer tools, that enable localizing the finding in one series, the CAD mark series, within the larger series of CT data, which were not utilized in this investigation. In practice, we suspect that some of the findings were easily dismissed by the readers in the CAD series even without localization on the 5 mm data. This dismissal of FP's, without localization, may also account for some of the overlapping of the time confidence intervals for localization (stage 2) and easy dismissal (stage 3), as shown in Fig. 6.

In a previous study, Roos et al. [22] designed a protocol to analyze, amongst other aspects, the time required by readers to evaluate CAD marks. Three radiologists reviewed 20 cases using a dedicated standalone workstation. Timestamps for every individual mouse-click by the reader could be automatically recorded and associated with a specific evaluation activity. In their study, consistent with our results, true-negative CAD findings were easily dismissed (on average, 4.7–in our study, 3.6 s) whereas some other CAD marks, eventually determined to be FPs, took up to three times longer (on average, 14.4 s–in our study, 11.5 s). Hence, as shown in our study, while the integration of CAD on the PACS enables the review of CAD findings, the lack of interactive capabilities is a limitation most noticeable when reviewing less obvious CAD findings. Thus, some of the standard functionalities, normally available in standalone workstation (e.g., 3D rendering, sliding this

slab MIPS, cartwheel views, etc), will also need to be incorporated into PACS.

The time analysis contributes a quantitative understanding of the steps and efforts involved in reviewing CAD marks when differentiating between obvious and non-obvious findings, allowing focusing on the actual time to evaluate a CAD mark. However, a more rigorous study involving more readers designed specifically to explore the timing aspects for the evaluation of CAD findings as well as more sophisticated modeling and computation of the fitting, may yield better insight in reader behavior and understanding as to how different types of proposed CAD findings may impact workflow. Furthermore, the use of CRF, as a means of annotation and recording time, artificially increases the time use by readers to evaluate a CAD mark since, as it is practice when reporting on a PACS system, a voice dictation system is normally employed.

A number of additional limitations need to be noted regarding this study. Firstly, evaluations were limited to the interpretation of solid nodules only: pure ground-glass lesions were purposefully excluded. Although a number of recent reports indicate that CAD devices are becoming considerably more adept at identifying these clinically important but frequently difficult to identify lesions, it may be anticipated that inclusion of these will substantially increase both the number of TP and FP nodules identified [44–47]. It is also noteworthy that the present study focused exclusively on low-dose CT screening studies. The decision to restrict the study in this manner reflects our impression that the best use of CAD is likely for evaluating patients for whom the primary indication is identification of lung nodules with a low pre-test probability with a high standard for defining which nodules are worth further evaluation, an approach supported by the recent results of the NLST study documenting a 20 % decrease in disease specific mortality with low-dose lung cancer screening. In this regard, the ability of CAD systems to work effectively using low dose data sets is well established [9, 10, 48]. The use of CAD in patients with diffuse lung disease, and specifically those with diffuse nodular lung diseases, or patients with easily identified multiple lung metastases, is less promising. Hence, while this study showed benefit in the use of CAD when applied to screening cases, the optimal clinical scenarios for which CAD is of use remains to be determined [14].

Another limitation in our investigation is that receiver operating characteristic analysis was not performed. However, the impact on reader evaluation with the integrated CAD results was of primary objective; therefore grading in terms of confidence was not asked of the readers to minimize this effect on timing evaluation.

It is also noted that this study employed dedicated chest radiologists instead of a more diverse reader population. This limitation always pertains to studies relying on

specialists [2, 13]. Extrapolation to a more general reader population is clearly indicated for further evaluation. In this regard, it is noted that compared with numerous prior studies employing CAD, the overall sensitivity of the five readers in this study proved to be unusually low despite familiarity of reviewing the series generated by the CAD server directly on the PACS. Several possible explanations may account for this result including, in particular, the fact that this study focused on low-dose screening studies—a population for which readers are prone to limit identification of potential nodules in order to avoid over-diagnosis. Paradoxically, once potentially significant nodules are identified in this population, follow-up studies are mandated, hence decreasing the need to identify all additional nodules. Additionally, the knowledge by readers being timed might also have affected their willingness to take extra care to identify all potential lesions. Regardless of the explanation, consistent with previously reported studies assessing the utility of CAD as a second reader, CAD significantly improved the detection of potentially significant nodules for all five readers.

Consistent with previously reported CAD studies, a lack of histological verification is an important limitation. Typically in clinical practice small nodules are not histologically confirmed, given their low likelihood to represent malignancy, rather they are followed up with chest CT [43]. In this regard, however, reliance on using a variable standard for determining a gold standard by use of the extent of reader agreement to identify TP nodules in our judgment represents a reasonable compromise [49]. Additionally, the readers were limited to reviewing the 5 mm sections, whereas CAD processed the 1 mm sections. The reading protocol was designed purposefully so as to remain as close as possible to the standard clinical practice at our institution, and because of technical limitations from the CAD tool. Furthermore, to prevent any bias, the reference standard was established on the 5 mm sections, so that potential small nodules, not visible on the 5 mm sections and marked by CAD, were rejected by the readers. Finally, this study focused exclusively in the use of CAD as a second read. Other reading paradigms, such as concurrent read, may prove of better clinical value. However, the effective benefit still remains to be assessed in further studies.

In conclusion, the result of this study indicates that the use of a CAD device seamlessly incorporated in a PACS environment for efficient use represents a practical application for this otherwise underutilized technology. Similarly to prior studies, in nearly all cases CAD improved the sensitivity of reader's detection of solid lung nodules without an undue number of FP findings. Measurement of the time taken to clinically utilize CAD proved acceptable, despite the need for readers to independently document all findings. While this study supports the use of CAD when restricted to

studies in patients in whom identification of nodules is critical with a low pre-test probability, use of CAD integrated with PACS clearly requires still further evaluation in more diverse clinical settings.

References

- Summers RM: Road maps for advancement of radiologic computer-aided detection in the 21st century. *Radiology* 229 (1):11–13, 2003
- Brown MS, Goldin JG, Rogers S, et al: Computer-aided lung nodule detection in CT: results of large-scale observer test. *Acad Radiol* 12:681–686, 2005
- McNitt-Gray MF, Armato SG, Meyer CR, Reeves AP, McLennan G, Pais RC, Freymann J, Brown MS, Engelmann RM, Bland PH, Laderach GE, Piker C, Guo J, Towfic Z, Qing DPY, Yankelevitz DF, Aberle DR, van Beek EJ, MacMahon H, Kazerooni EA, Croft BY, Clarke LP: The lung image database consortium (LIDC) data collection process for nodule detection and annotation. *Acad Radiol* 14(12):1464–1474, 2007
- Suzuki K, Li F, Sone S, Doi K: Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network. *IEEE Trans Med Imaging* 24(9):1138–1150, 2005
- Goldin JG, Brown MS, Petkovska I: Computer-aided diagnosis in lung nodule assessment. *J Thorac Imaging* 23:97–104, 2008
- Marten K, Engelke C, Seyfarth T, et al: Computer-aided detection of pulmonary nodules: influence of nodule characteristics on detection performance. *Clin Radiol* 60:196–206, 2005
- Way T, Chan HP, Hadjiiski L, et al: Computer-aided diagnosis of lung nodules on CT scans: ROC study on its effect on radiologists' performance. *Acad Radiol* 17:323–332, 2010
- White CS, Pugatch R, Koonce T, et al: Lung nodule CAD software as a second reader: a multicenter study. *Acad Radiol* 15:326–333, 2008
- Hein PA, Rogalla P, Klessen C, et al: Computer-aided pulmonary nodule detection—performance of two CAD systems at different CT dose levels. *Rofo* 181:1056–1064, 2009
- Das M, Muhlenbruch G, Heinen S, et al: Performance evaluation of a computer-aided detection algorithm for solid pulmonary nodules in low-dose and standard-dose MDCT chest examinations and its influence on radiologists. *Br J Radiol* 81:841–847, 2008
- Kim JS, Kim JH, Cho GS, et al: Automated detection of pulmonary nodules on CT images: effect of section thickness and reconstruction interval—initial results. *Radiology* 236:295–299, 2005
- Teague SD, Trilakis G, Dharaiya E: Lung nodule computer-aided detection as a second reader: influence on radiology residents. *J Comput Assist Tomogr* 34:35–39, 2010
- Doi K: Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput Med Imaging Graph* 31:198–211, 2007
- van Ginneken B, Schaefer-Prokop CM, Prokop M: Computer-aided diagnosis: how to move from the laboratory to the clinic. *Radiology* 261(3):719–732, 2011 Dec
- Welter P, Hocken C, Deserno TM, Grouls C, Günther RW: Workflow management of content-based image retrieval for CAD support in PACS environments based on IHE. *Int J Comput Assist Radiol Surg* 5(4):393–400, 2010
- Faggioni L, Neri E, Castellana C, Caramella D, Bartolozzi C: The future of PACS in healthcare enterprises. *Eur J Radiol* 78(2):253–258, 2011
- Erickson BJ, Bartholmai B: Computer-aided diagnosis at the start of the Third Millennium. *J Digit Imag* 15(2):59–68, June 2002
- Boone JM: Radiological interpretation 2020: toward quantitative image assessment. *Med Phys* 34(11):4173–4179, 2007
- Anh H, Le T, Liu B, Huang K: Integration of a computer-aided Diagnosis/Detection (CAD) results in a PACS environment using CAD-PACS toolkit and DICOM SR. *Int J Comput Assist Radiol Surg* 4(4):317–329, 2007
- Huang K, Liu BJ, Anh H et al, Chapter 18: PACS-based computer aided detection and diagnosis. *Biomedical Image Processing (Biological and Medical Physics, Biomedical Engineering)*, 455–470, DOI: 10.1007/978-3-642-15816-2_18
- National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al: Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 365:395–409, 2011
- Roos JE, Paik D, Olsen D, et al: Computer-aided detection (CAD) of lung nodules in CT scans. *Eur Radiol* 10:549–557, 2010
- Naidich DP, Ko JP, Stockel J, et al: Computer aided diagnosis: impact on nodule detection among community level radiologists, a multi-reader study. *Int J Comput Assist Radiol Surg* 1268:902–907, 2004
- Godoy M, Kim TJ, Ko J, Florin CH, et al, Computer-aided detection of pulmonary nodules on CT: evaluation of a new prototype for detection of ground-glass and part-solid nodules, SSK04-07 RSNA 2008, p.517.
- Das M, Honnef, D, O'Dell D et al, Prospective Evaluation of a CAD Server for Computer-aided Detection in Daily Routine Chest CT Examination: Evaluation of 234 Patients, SSK-08 RSNA 2008, p. 517.
- Sakai S, Sod Y, Takahashi N, et al: Computer-aided nodule detection on digital chest radiography: validation test on consecutive T1 cases of resectable lung cancer. *J Digit Imag* 19(4):376–382, Dec. 2006
- Pietka E, Pospiech-Kurkowska S, Gertych A: Integration of computer assisted bone age assessment with clinical PACS. *Comp Med Img Graph* 27(2):217–228, Mar. 2003
- Sakai S, Yabuuchi H, Matsuo Y, et al: Integration of temporal subtraction and nodule detection system for digital chest radiographs into picture archiving and communication system (PACS): four-year experience. *J Digit Imag* 21(1):91–98, Mar 2008
- Ko JP, Betke M: Chest CT: automated nodule detection and assessment of change over time—preliminary experience. *Radiology* 218(1):267–273, 2001
- Tam M, Deklerck R, Jansen B, et al: A novel computer-aided lung nodule detection system for CT images. *Med Phys* 38(10):5630–5645, 2011
- Armato III, S, Giger M, Moran C, Blackburn J, Doi K, MacMahon H: Computerized detection of pulmonary nodules on CT scans. *Radiographics* 19:1303–1311, 1999
- Lee Y, Hara T, Fujita H, Itoh S, Ishigaki T: Automated detection of pulmonary nodules in helical CT images based on an improved template-matching algorithm technique. *IEEE Trans Med Imaging* 20(7):595–604, 2001
- Bogoni L, Bi J, Florin C, et al: Lung nodule detection. In: Müller H, Clough P, Deselaers T, Caputo B Eds. *ImageCLEF—experimental evaluation in visual information retrieval series: the information retrieval series*. Springer, Berlin, 2010, pp 415–451
- Periaswamy S, and Bogoni L, System and method for filtering and automatic detection of candidate anatomical structures in medical images. US Patent 7,912,292.
- Liang J and Bogoni L, Toboggan-based shape characterization. US Patent 7,480,412.
- Jerebko A, Bogoni L, Lakare S, Segmentation of structures based on curvature slope. US Patent 7,634,133.
- Okada K, Comaniciu D, Krishnan A: Robust anisotropic Gaussian fitting for volumetric characterization of pulmonary nodules in multislice CT. *IEEE Trans Med Imaging* 24(3):409–423, Mar 2005

38. Kubota T, Estimation of solitary pulmonary nodule diameters with reaction-diffusion segmentation. US Patent 7,720,271.
39. V, Krshmapuram B, Bi J, et al. Bayesian multiple instance learning: automatic feature selection and inductive transfer. In: Proc. 25th Intr Conf Mach. Learning, 2008, pp 808–815.
40. Raykar VC, Yu S, Zhao LH, Hermosillo G, Florin CH, Bogoni L, Moy L: Learning from crowds. *J Mach Learn Res* 11:1297–1322, 2010
41. Armato SG, Roberts RY, Kocherginsky M, et al: Assessment of radiologist performance in the detection of lung nodules: dependence on the definition of “truth”. *Acad Radiol* 16:28–38, 2009
42. Ochs RA, Kim HJ, Angel E, et al. Forming a reference standard from LIDC data: impact of LIDC reader agreement on the reference dataset and reported CAD performance. In: Proc. SPIE, 30 Mar 2007, vol. 6514, p 82, DOI: [10.1117/12.707916](https://doi.org/10.1117/12.707916)
43. MacMahon H, Austin JH, Gamsu G, et al: Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology* 237:395–400, 2005
44. Beigelman-Aubry C, Hill C, Boulanger X, et al: Evaluation of a computer aided detection system for lung nodules with groundglass opacity component on multidetector-row CT. *J Radiol* 90:1843–1849, 2009
45. Kim KG, Goo JM, Kim JH, et al: Computer-aided diagnosis of localized ground-glass opacity in the lung at CT: initial experience. *Radiology* 237:657–661, 2005
46. Lee JW, Jeong JW, Lee S, et al: The GGO lesions detected by computer-aided detection system on chest MDCT images. *Conf Proc IEEE Eng Med Biol Soc* 1:1983–1985, 2006
47. Okada T, Iwano S, Ishigaki T, et al: Computer-aided diagnosis of lung cancer: definition and detection of ground-glass opacity type of nodules by high-resolution computed tomography. *Jpn J Radiol* 27:91–99, 2009
48. Hein PA, Romano VC, Rogalla P, et al: Variability of semiautomated lung nodule volumetry on ultralow-dose CT: comparison with nodule volumetry on standard-dose CT. *J Digit Imag* 23:8–17, 2009
49. Park EA, Goo JM, Lee JW, et al: Efficacy of computer-aided detection system and thin-slab maximum intensity projection technique in the detection of pulmonary nodules in patients with resected metastases. *Invest Radiol* 44:105–113, 2009