Computer-aided of lung nodules on thin collimation MDCT: impact on radiologists’ performance

B Brochu (1), C Beigelman-Aubry (1), J-L Goldmard (2), P Raffy (3), PA Grenier (1) and O Lucidarme (1)

Résumé
Évaluation de l’impact d’un système CAD sur la performance des radiologues pour la détection des nodules pulmonaires sur des examens scanographiques multicoupes du thorax

J Radiol 2007;88:573-8

Objectifs. Évaluer l’amélioration de la détection des nodules pulmonaires apportée par l’utilisation d’un CAD scanographique.

Matériel et méthodes. Trois radiologues (R1, R2, R3), ayant des niveaux d’expérience différents, interprètent de façon indépendante 30 examens scanographiques du thorax de dépistage, sans, puis avec l’aide d’un CAD. La vérité est établie par une séance d’interprétation différée par deux des trois radiologues en consensus, assistés par le CAD.

Résultats. La lecture consensuelle permet d’identifier 133 nodules, dont 61 (46 %) ≥ 4 mm. Les valeurs de sensibilité dans la détection des nodules avant et après l’utilisation du CAD, sont respectivement de 54 % et 80 % (R1), 38 et 71 % (R2), et 70 et 88 % (R3). En ne considérant que les nodules ≥ 4 mm, les valeurs de sensibilité varient avant et après utilisation du CAD de 62 % à 95 % (R1), de 41 % à 84 % (R2), et de 74 % à 92 % (R3). En combinant deux à deux les résultats des trois radiologues obtenus sans CAD, les valeurs de sensibilité sont respectivement de 65, 83 et 77 % pour tous les nodules, et de 70, 85 et 77 % pour les nodules ≥ 4 mm. Le CAD induit au total 105 faux positifs avec une moyenne de 3,5 par examen.

Conclusion. Les nodules pulmonaires manqués par le radiologue peuvent être détectés si le CAD est utilisé comme second lecteur. Le CAD peut être au moins aussi bénéfique que l’utilisation d’un second lecteur indépendant.


Abstract

Objectives. Evaluate the improvement in detecting lung nodules when using multidetector CT (MDCT) computer-assisted diagnosis (CAD).

Material and methods. Three radiologists (R1, R2, R3) with different levels of experience independently interpreted 30 MDCT examinations of the thorax taken for screening purposes, first without and then with CAD. The diagnosis was established by two of the three radiologists interpreting the images together, assisted by the CAD.

Results. The consensus reading identified 133 nodules, 61 (46%) of which were 4 mm or larger. The sensitivity values in the detection of nodules before and after using the CAD were 54% and 80% (R1), 38% and 71% (R2), and 70% and 88% (R3), respectively. When considering only the nodules that were 4 mm or larger, the sensitivity values varied before and after using the CAD, from 62% to 95% (R1), from 41% to 84% (R2), and from 74% to 92% (R3). By combining two by two the three radiologists’ results obtained without the CAD, the sensitivity values were 65%, 83%, and 77%, respectively, for all the nodules, and 70%, 85%, and 77% for the nodules that were 4 mm or larger. The CAD induced a total of 105 false-positive results, with a mean of 3.5 per examination.

Conclusion. The lung nodules missed by the radiologist can be detected if the CAD is used as a second reader. The CAD can be at least as beneficial as the use of a second independent reader.

Key words: Computer-aided diagnosis (CAD). Lung nodules. Multidetector CT (MDCT).


Detection, specific diagnosis, and effective treatment of a large number of lung diseases, including primary or secondary bronchopulmonary cancers, begin by identifying lung nodules. CT sensitivity in detecting lung nodules is much better than chest x-ray sensitivity. In addition, the helical scanner is significantly better than conventional CT (1). Today, with the increasingly widespread use of multidetector CT scanners, the lungs can be acquired in their entirety during a single breath hold with high spatial resolution resulting from thin collimation and overlapping slices in reconstruction. High-resolution acquisition provides anatomic precision of the bronchovascular tree, which results in finer detection of small lung nodules. This detailed anatomy is obtained, however, at the cost of a large number of images to interpret, which makes reading long and laborious, with an increased risk of errors in detecting nodules. Therefore, despite the technological progress made in CT imaging, a high number of lung nodules go unrecognized by radiologists (2-4). To prevent detection errors, several computer-aided

(1) Service de Radiologie, (2) Département de Biostatistique, Hôpital Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris, Université Pierre et Marie Curie, Paris VI, 47-81 bd de l’Hôpital, 75651 Paris Cedex 13.
(3) R2 Technology, Inc., Sunnyvale, CA.
Correspondance : PA Grenier
E-mail : philippe.grenier@psl.aphp.fr
Computer-aided lung nodules on thin collimation MDCT: impact on radiologists’ performance

B Brochu et al.

Material and methods

Patients and CT protocol

This study was based on the numerical data of multislice CT scans of 30 smokers or ex-smokers (20 men, ten women), with a mean age of 55 years (range, 29-76 years), sent to the radiology department between December 2002 and December 2003 to be included in a randomized multicenter prospective pilot study (Depiscan). This was a feasibility study for a large trial evaluating the benefit of potential systematic thoracic radiography or low-dose lung CT screening of subjects at risk for lung cancer on lung cancer mortality. Written informed consent was obtained beforehand from the patients for the use of their images for research purposes, in accordance with the recommendations of the local ethics committee.

The tests were done with a 16-detector row volume-rendered helical CT (Lightspeed Ultra, GE Medical System, Milwaukee, WI, USA). The image acquisition protocol was based on the following parameters: 110 kV, 50-60 mA, 16×1.25 mm. Images were reconstructed with a 1.25-mm thickness and a 0.6-mm interval using a specific reconstruction filter for the lung. No contrast medium was injected. The lung volume examination was acquired during a single breath hold obtained at the end of a maximal inspiration.

Reading the radiological images

Three radiologists (R1, R2, R3) participated in reading the CTs. Two of them were confirmed thoracic radiologists having 15 years (R1) and 20 years (R2) of experience, and the third was a general radiologist who had recently finished training with only 6 months of experience in interpreting thoracic CT images (R3). The three radiologists interpreted the study’s 30 CT images independently. An Advantage Windows 4.2 work station (GE Medical System, Milwaukee, WI, USA) was used to interpret the images. The axial overlapping images were interpreted in cine mode using the mouse to move from top to bottom through the volume. The interpretation sessions were conducted in the same conditions as the clinical environment, with an interpretation time limited to 15 min per case. The radiologist had to record the presence of a lung nodule, note the anatomic location (lobe and slice number), and measure the largest axial diameter using the work station’s electronic marker. The criteria defining a lung nodule were established beforehand: highly limited opacity with tissue density, spherical or ellipsoidal in shape (length ≤ 3 × width), or a more complex shape with irregular contours. Reticular linear or subpleural opacities with no noddular aspect were excluded, as were bronchial wall thickenings, focal alveolar condensation zones, and zones with substantial movement artifacts. Focal unpolished glass patterns were not taken into account in this study because the CAD software could not detect them.

The criteria defining anatomical location included the slice number, the lobe and structure potentially related to the nodule (bronchus, vessel, pleura), and the axial compartment of the lung in which the nodule was located (internal, middle, or external). These compartments were delimited by two virtual curves from the anterior mediastinal pleura to the posterior mediastinal pleura, 1 and 3 cm from the peripheral pleura, respectively.

CAD system

The CAD used was the ImageChecker V1.0, de R2 Technology (Inc., Sunnyvale, CA, USA). After each of the interpretations by the radiologist, the CAD software was used on the dedicated work station to reassess the presence of nodules. All the nodules detected by the CAD software and missed by the radiologist during the first interpretation were recorded and evaluated in terms of size and anatomical location (fig. 1). All the nodules detected by the CAD software and not considered a nodule by the radiologist (crossing vessel, pleural thickening, reticular or linear pulmonary scars or opacities) were defined as false-positive results on the CAD software.

Consensus reading reference standard

The reference was established by consensus between two of the three radiologists who had already done the independent reading. This reevaluation of the CT images was done on the Advantage Windows 4.2 workstation (GE Medical System, Milwaukee, WI, USA), with no time limit (between 20 and 45 min per case) to establish the presence, size, and anatomic location of all the nodules. In this task, the radiologists were free to use all the analysis processes available on the workstation, particularly the maximum intensity projection technique. The CAD results were also used to determine whether there was a nodule, with the objective of establishing a standard that would be as close to truth as possible.

Statistical analysis

The first part of the analysis included all nodules of all sizes. Given that the CAD system evaluated was particularly optimized to detect nodules measuring more than 4 mm in diameter, the second part of the analysis concentrated on nodules that were 4 mm or less in diameter.

The CAD system performance was evaluated for its sensitivity in detecting nodules and in terms of the number of false-positives per examination. Since each nodule seen by the radiologists was detected by the CAD system and clearly positioned on the CT slices, sensitivity was calculated based on nodules rather than patients. On the other hand, specificity could not be calculated because it was impossible to define the number of true negatives. The false-positives were defined as each element detected either by the CAD system or the radiologists that was not confirmed as being a nodule during the consensus reading.

The sensitivity values and the number of false-positives were used to compare the performance of one radiologist alone, without the assistance of the CAD, with the performance of the same radiologist using the CAD, or to compare the performance of each radiologist using the CAD with the results of a double interpretation (R1+R2, R1+R3, R2+R3), as...
**Fig. 1:** Presentation of the ImageChecker CT V1.0 (R2 Technology) system with, in the upper left-hand corner, a coronal view of the 3D reconstruction of the thoracic vascular structures. This phase differentiates nodules from pulmonary vessels. The yellow horizontal line represents the position of the axial slice visualized on the right. A nodule screened using CAD is marked on this axial slice by a green circle. In the lower left-hand corner is a small localized view, a 3D reconstruction of the nodule’s vessels within a VOI defined by the square on the axial image of the right lung. The lung nodule is shown in green.

**Fig. 2:** Two axial slices from the same examination. The CAD detected two lung nodules that were missed by the three radiologists.

- **a** The CAD detected a nodule in the right upper lobe (green circle).
- **b** In the slice 2 mm above, the CAD detected another nodule in the upper left lobe (green circle).

**Fig. 3:** Examples of CAD false-positives

- **a** Wall of emphysema bulla.
- **b** Vascular bifurcation.
- **c** Pulmonary ligament
can be imagined in a screening situation. In this double interpretation situation, a nodule was considered as such in the analysis when one of the two radiologists detected it and indicated it as a potential nodule. Specificity was not evaluated because the CAD’s false-positives were easily and automatically corrected by the radiologist.

The statistical analysis was carried out using McNemar (two-tailed) tests with \( p \)-values under 0.05 for statistical significance. Statistica software V5.5, 1999 (StatSoft, Inc. Tulsa, OK, USA) was used for these analyses.

## Results

The consensus reading of the 30 multislice CT examinations identified 133 nodules, 65 of which (46%) measured 4 mm or more in diameter. Fifty-nine percent of the nodules (79/133) were located in the right lung, and 41% (54/133) were located in the left lung.

The nodules were distributed in relation to the lobe location as follows: 46% (61/133), 9% (12/133), and 45% (60/133) of the nodules were located in the superior, middle, and inferior lobes, respectively.

The majority of the nodules (67%; 89/133) were located in the external third of the lungs, 19% (26/133) in the middle third, and 13% (18/133) in the internal third. The nodules measuring 4 mm or less were mainly located in the inferior lobes: 56% (34/61) versus 35% (21/60) in the superior lobes. Inversely, the nodules measuring 4 mm or more were more frequent in the superior lobes: 55% (40/72) versus 36% (26/72) in the inferior lobes (table I).

### CAD results

Use of the CAD system, before radiologist interpretation, detected 75 of the 133 nodules, giving a 56% raw sensitivity value. The sensitivity for nodules detected measuring 4 mm or more was 79% (48/61) (fig. 2). The CAD was responsible for a total of 105 false-positives observed over the 30 MDCT examinations, with a mean of 3.5 false-positives per examination. The false-positives were mainly the result of vessel crossing, focal thickening of the pleura, pulmonary scars, or areas of linear or reticular opacity (fig. 3).

### Results of the interpretations by radiologists with and without CAD

The sensitivity in detecting nodules by one radiologist not using the CAD, for all nodule sizes, was significantly different for R1, R2, and R3 (\( p \leq 0.01 \)): 54% (72/133), 38% (51/133), and 70% (93/133), respectively. The sensitivity values for detecting nodules 4 mm and larger were higher: 62% (38/61), 41% (25/61), and 74% (45/61), respectively, but here again with a significant difference between the observers (\( p < 0.01 \)). When the results of the CAD system were associated with radiologist performance, sensitivity values increased significantly to 80% for R1 (\( p < 0.0001 \)), 71% for R2 (\( p < 0.0001 \)), and 88% for R3 (\( p < 0.0001 \)). For detecting nodules that were 4 mm or smaller, sensitivity values were even higher: 95% for R1 (\( p < 0.0001 \)), 84% for R2 (\( p < 0.0001 \)), and 92% for R3 (\( p = 0.002 \)). The difference in performance between radiologists remained significant (\( p < 0.01 \)).

Table II compares the sensitivity values for detecting nodules 4 mm and larger when the results of two radiologists were combined (the equivalent of a double reading, R1+R2, R1+R3, R2+R3), with sensitivity values obtained by a single radiologist associated with using the CAD.

Table III reports the number of false-positive results obtained for each radiologist, with and without the CAD. The majority of the false-positives induced by the CAD were rapidly and easily identified by the radiologist. Nevertheless, the CAD led to a decrease in the radiologist’s reading specificity by inducing an increase in the false-positive rate.

### Table I

| Number of lung modules detected and lobe location. |
|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| Superior lobe                  | Middle lobe                    | Inferior lobe                   | Total                           |
| Right lung                     | Left lung                      | Right lung                      | Left lung                       |
| 12                             | 9                              | 36                              | 24                              |
| 6                              | –                              | 12                              | –                               |
| 17                             | 17                             | 32                              | 29                              |
| 35                             | 26                             | 80                              | 53                              |

### Table II

| Sensitivity values (%) of interpretations radiologists using CAD to detect nodules \( \geq 4 \) mm. |
|-------------------------------------------------|-------------------------------------------------|---------------------------------|
| Double independent interpretation               | Radiologists + CAD                              | McNemar test                    |
| R1+R2=70                                       | R1+CAD=95                                      | \( 0.0001^* \)                  |
| R1+R2=70                                       | R2+CAD=84                                      | 0.09                             |
| R1+R3=77                                       | R1+CAD=95                                      | 0.08                             |
| R1+R3=77                                       | R3+CAD=92                                      | 0.28                             |
| R2+R3=85                                       | R2+CAD=84                                      | 0.31                             |
| R2+R3=85                                       | R3+CAD=92                                      | \( 0.006^* \)                   |

* NB: sensitivity is expressed in %

* Statistically significant results.

### Table III

| Number of false-positive results in radiologists’ interpretations in single or double reading without CAD and after using CAD. |
|---------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|--|-------------------|
| Radiologists or CAD | Double independent interpretation | Radiologists + CAD |
| R1=25 | / | / | R1+CAD=25 |
| R2=4 | R1+R2=25 | R1+R3=28 | R2+CAD=15 |
| R3=6 | R2+R3=6 | R3+CAD=10 |

B Brochu et al.
Discussion

Despite the clearly documented improvement in the detection of lung nodules using thoracic CT compared to plain radiograph, misjudging a nodule on a CT examination remains a relatively frequent occurrence (2-4, 6, 7). In evaluating annual CT examinations within a systematic screening program for lung cancer on an at-risk population, Swensen et al. demonstrated that one or several nodules were missed on the first CT image in 26% of the subjects examined (4). Setting up an independent or consensus double interpretation could reduce the number of these interpretation errors, but this remains impractical in daily clinical practice, apart from a mass screening program. All CT thoracic examinations, no matter what the indication, are challenged by correctly screening the lung nodules actually present.

Therefore, the detection of lung nodules is one of the most difficult tasks in interpreting thoracic CT examinations. In accordance with Awai et al., the radiologist’s performance in detecting lung nodules probably depends more on how attentive the radiologist is when reading the images than on experience and knowledge as a radiologist (8). In this study, a significant difference was noted between the performance of the three radiologists, with or without CAD, which seems independent of their experience, since the lowest performance levels were obtained by the most senior radiologist. The best performance was obtained by the recently certified radiologist with the least experience in thoracic CT. This study confirms that the objective of a CAD is indeed to improve the performance of radiologists in detecting lung nodules, however knowledgeable and experienced they are in thoracic MDCT (9). One must remain aware, however, that the quality of the final result in a combined radiologist-CAD system also depends on the radiologist’s detection performance during the first reading (9).

Many studies have demonstrated high variability in sensitivity values for CAD in detecting lung nodules. These values vary between 38% and 95% depending on the CAD system evaluated and the variations in the nodule characteristics used in the study and the acquisition criteria (8, 10-18). Studies assessing CAD systems developed for 5- to 10-mm CT slice thicknesses on lesions measuring more than 5 mm in diameter have reported sensitivity values varying from 38% to 86%, with a false-positive rate of approximately five per examination (8, 10, 12, 15, 17-20). Armato et al. reported their experience with this type of CAD system, which detected 84% of bronchopulmonary cancers missed by radiologists in a database of low-dose thoracic CT within a screening program for lung cancer (12).

Although most CAD systems were developed to detect lung nodules measuring more than 5 mm and to be used on relatively thick CT slices, others were designed for thinner slices (1-3 mm) and to detect nodules of any size (21-23). Brown et al. reported their experience with a CAD tool developed to detect nodules measuring less than 3 mm on thin slices and applied to thoracic CT the images of 15 subjects (13). The CAD system was capable of detecting the 22 nodules measuring more than 3 mm, and 40 of the 57 (70%) nodules measuring less than 3 mm, with a mean of 15 false-positives per examination, for a total of 219 false-positives. The benefit contributed by the CAD in the detection of nodules after radiologists’ reading was 4% for nodules 3 mm and larger and 23% for nodules measuring less than 3 mm. Recently, Yuan et al. reported the experience of R2 Technology’s Image-Checker CAD on 150 low-dose CT examinations within a lung cancer screening program. The CAD detected 456 (73%) of the 628 true nodules, with a mean false-positive rate of 3.2 per patient (23).

In our study, the CAD tool was optimized to detect nodules measuring 4 mm or greater, with a 79% sensitivity for these nodules. The gain in sensitivity, after a first reading by the radiologists, was a mean 31% for nodules that were 4 mm or larger for the three radiologists and a mean 21% for nodules under 4 mm. This substantial gain can be explained in several ways. The first is the time limited to 15 min for the radiologist’s reading so as to be as close as possible to realistic interpretation conditions. This restriction imposed on the radiologists was warranted by the study’s secondary objective of measuring the effect of the CAD on the reading of CT examinations within a mass lung cancer screening program (standardization and cost reduction). Particularly on thin-slice examinations, the number of lesions detected, proportional to the reading time required to interpret the examination on a workstation, can be determined. The second may be related to radiologists not using the maximum intensity projection technique during their first reading, even though it is known that this technique significantly improves nodule detection, particularly those nodules projecting in the parahilar regions (24). Systematic use of the maximum intensity projection by the radiologists participating in this study would probably have reduced the benefit provided by the CAD.

Used as a second reader, the CAD provides sensitivity values in detecting lung nodules that are higher than those of two independent readers, although this difference in our study was only significant in two of six possible combinations (table II). Similar results were reported by Wermanns et al. Comparisons of the performance on single readings by three independent radiologists, a double reading, and a reading combining radiologists and CAD pleads in favor of the latter, with a mean sensitivity value in detecting nodules of 54%, 67%, and 79%, respectively (25). Similarly, in another study on a population of 195 noncalcified nodules measuring 3 mm or greater, detected on 20 thoracic CT examinations, Rubin et al. demonstrated that CAD improved radiologists’ performance and that the CAD-radiologist combined reading provided higher sensitivity rates than a double reading by radiologists (22).

One of the limits of CAD in the detection of lung nodules stems from the fact that excellent sensitivity comes at the cost of an unavoidable number of false-positive results, as is the case for mammmary nodules (26). A few false-positives can be attributed to cardiac movement artifacts, giving a nodular or fragmented appearance to vessels. Other anomalies of the parenchyma or the pleura that are detected correspond in fact to nonspecific opacities or normal anatomical structures such as pulmonary ligaments, pulmonary veins, the esophagus, or pleural plaques. Additional morphological analysis tools integrated into the more recent versions of CAD systems, such as maximum intensity and neuron projection, will prevent confusion between vessels and nodules, but these improvements will not successfully resolve all the problems. Fortunately, most of the false-positives with CAD are easily detected visually by the radiologist who can differentiate them from true nodules.
Nevertheless, the radiologist’s complementary analysis lengthens the interpretation time, emphasizing the need for as small a number of false-positives as possible.

Finally, the objective of lung CAD is not limited to the simple detection of the nodule but can also measure the nodule’s volume and evaluate its growth on successive examinations. The international recommendations advance monitoring the size of the nodule on successive CT examinations as the best diagnostic strategy for undetermined nodules measuring 8 mm or less (27). CAD performance in this phase of describing the nodule during monitoring will be evaluated at a later date.

References


B Brochu et al.