Investigations and research

Performance of lung nodule computer aided detection software: effect of slice thickness on chest CT

C.S. White  
R.D. Pugatch  
S.D. Teague  
E. Dharaiya  
K. Read

University of Maryland Medical Center, Baltimore, MD, USA.  
Indiana University, Indianapolis, IN, USA.  
Philips Medical Systems, Highland Heights, OH, USA.

Lung cancer is the leading cause of cancer death among men and women in the United States. It is estimated that over 150,000 people die of lung cancer annually in the USA [1], and an estimated 1 million people die of lung cancer worldwide. The overall 5-year survival rate for patients with lung cancer is 15%, a number that has not changed significantly in the last 25 years. If the cancer is detected at early stage, while it is still localized, the survival rate increases to 49% [2] or more. Smaller nodules (< 15 mm in diameter) have a higher probability of being stage I lung cancer [3,4]. It is therefore very important to identify small indeterminate nodules on chest CT scans. These nodules can then be evaluated to determine their etiology by various methods, including measurement of growth on sequential scans.

Nodules that remain stable over a period of two years are typically characterized as benign [5]. In addition to lung cancer, cancers inside and outside the lung can metastasize and manifest as pulmonary nodules [6].

Recent advances in multi-detector CT (MDCT) allow thin section scanning of the chest in a single breath-hold. Identification of indeterminate pulmonary nodules on chest CT scans is common in routine radiology practice. The current generations of MDCT scanners produce huge datasets with a large number of slices. Radiologists typically interpret a high volume of these studies on a daily basis. The vast amount of data, in conjunction with large caseloads, presents a challenge for the practicing radiologist. In one lung cancer screening study, Swensen et al. [7] found that radiologists failed to diagnose nodules in 26% of the patients.

The intrinsic perceptual limitation of human readers has led to newer approaches that can help the radiologists take full advantage of the high spatial resolution offered by MDCT, in order to facilitate efficient and accurate detection of lung nodules. Studies have shown that double reading by two radiologists can increase the efficiency in detecting lung and breast tumors [8]. However, double reading is not an optimized method from an economic standpoint. An alternative approach is now available with the development of computer aided detection (CAD) software as a substitute for a second human reader [8].

The potential value of CAD is demonstrated by at least two recent reports. A study by Gurung et al. [9] demonstrated that the sensitivity of CAD software in detecting a pulmonary nodule less than 10 mm in diameter at 1 mm reconstruction thickness is higher than that of a radiologist [9]. A similarly designed study by Marten et al. [10] showed that, at a reconstruction slice thickness of 0.75 mm, the performance of CAD was better than that of radiologists in detecting nodules less than 10 mm in diameter [10]. These studies suggest that CAD may have incremental value in the assessment of lung nodules on thin slice chest CT scans.

Several studies have shown improvement in the performance of the radiologists using CAD as a second reader. Rubin et al. [8] reported that the mean sensitivity of the radiologists increased from 50% to 63% when CAD was used as a second reader.

A number of studies have been published on CAD software using different slice thickness and different nodule sizes. The stand-alone sensitivity of CAD software in these studies varies from 38% to 91% and the numbers of false positive markings range from 3 to 50 per patient [8]. Investigators have also suggested that performance of the CAD system may improve with a decrease in slice thickness as

Important note:
This article describes ongoing research. The CAD software used in the study is a part of Lung Nodule Application (LNA) on Philips’ Extended Brilliance Workspace (EBW) and is not commercially available in the United States.
well as with an increase in the overlap between slices [11].

CT screening exams for detection of lung nodules are usually low-dose examinations with slice thicknesses in the range of 5 mm to 10 mm [12,13]. The use of thin slice data in screening examinations is limited, mainly due to time and scan length issues. However, with the development of 16, 40 and 64 slice scanners many of these issues can be overcome. Moreover, nodules may be present on examinations that are not obtained in a screening setting and for which thin sections are routinely reconstructed (e.g. to rule out pulmonary embolism).

Most of the studies evaluating the performance of CAD used thin and thicker reconstruction intervals and included patients with known lung nodules or malignancy. The slice thicknesses used in these studies varied from 0.75 mm to 5 mm. The current study was not specifically targeted towards screening population or patients with pulmonary nodules and therefore was primarily concentrated on thinner reconstruction intervals (<3 mm). The purpose of the study was to evaluate the effect of incremental variation in slice thicknesses on the performance of the CAD software using a range of thin slices. The CAD software was evaluated on datasets with slice thicknesses of 0.9 mm, 1.8 mm and 2.7 mm.

Subjects and methods
The study involved 28 patients scanned between March 2004 and December 2004. Each case was collected from Indiana University Hospital (Indianapolis, Indiana) after approval had been obtained from the US Institutional Review Board (IRB). Because of the retrospective nature of the study, only a verbal informed consent was required. The date and the time of the patient consent was recorded. Although the patients were selected sequentially, only the patients who could be contacted retrospectively and gave their verbal consent were included in the study. Patients with partial lung scans, collapsed lungs and fibrosis or lung surgery were excluded from the study. The patients included in the study were not scanned for lung cancer, and some of them ultimately proved not to have any pulmonary nodules. The 28 patients consisted of 16 women and 12 men, with an average age of 60.2 years. There were 7 smokers, 16 non-smokers and 5 with unknown smoking history enrolled in the study.

Each of the chest CT examinations was performed on a 40-slice MDCT (Brilliance 40, Philips Medical Systems, Cleveland, OH, USA). There were 18 contrast-enhanced scans and 10 non-contrast scans. The original images were reconstructed at a slice thickness of 0.9 mm and a reconstruction interval of 0.45 mm. The original data were stored at the site without any compression. These datasets were then reformatted to obtain two different slice thicknesses of 1.8 mm and 2.7 mm.

Two thoracic radiologists, each with more than ten years of experience, interpreted the studies to provide the ground truth. The two radiologists independently reviewed the datasets to identify nodules. The radiologists identified all nodules greater than or equal to 4 mm and less then 30 mm in maximum diameter. Nodules less than 4 mm in maximum diameter were excluded to avoid partial volume effects due to thick datasets.

The Fleischner society definition [15] of the nodule was used to determine the upper limit on the nodule size for inclusion in the study. The location and the slice number of each nodule were recorded on a case report form and the findings were also saved with the dataset. A second read was performed on cases where the radiologists did not agree. The radiologists were allowed to review the findings from the other radiologist during the second read. The locations of the nodules with slice numbers were also recorded from the second read and the findings were saved again. Only the nodules identified and agreed upon by both the radiologists after the second read were included as ground truth for the study.

All the ground truth nodules were identified on the cases with the reconstruction interval of 0.9 mm. The radiologists also reviewed the 1.8 mm and 2.7 mm thick datasets to verify that the nodules identified as ground truth were also visible on these datasets. The two radiologists identified a total of 33 nodules in 28 datasets. The nodules were between 4 mm and 30 mm in maximum diameter. Nodules less than or equal to 4 mm and greater than or equal to 4 mm and less then 30 mm in maximum diameter. Nodules less than 4 mm in maximum diameter were excluded to avoid partial volume effects due to thick datasets.

The Fleischner society definition [15] of the nodule was used to determine the upper limit on the nodule size for inclusion in the study. The location and the slice number of each nodule were recorded on a case report form and the findings were also saved with the dataset. A second read was performed on cases where the radiologists did not agree. The radiologists were allowed to review the findings from the other radiologist during the second read. The locations of the nodules with slice numbers were also recorded from the second read and the findings were saved again. Only the nodules identified and agreed upon by both the radiologists after the second read were included as ground truth for the study.

Subjects and methods
The study involved 28 patients scanned between March 2004 and December 2004. Each case was collected from Indiana University Hospital (Indianapolis, Indiana) after approval had been obtained from the US Institutional Review Board (IRB). Because of the retrospective nature of the study, only a verbal informed consent was required. The date and the time of the patient consent was recorded. Although the patients were selected sequentially, only the patients who could be contacted retrospectively and gave their verbal consent were included in the study. Patients with partial lung scans, collapsed lungs and fibrosis or lung surgery were excluded from the study. The patients included in the study were not scanned for lung cancer, and some of them ultimately proved not to have any pulmonary nodules. The 28 patients consisted of 16 women and 12 men, with an average age of 60.2 years. There were 7 smokers, 16 non-smokers and 5 with unknown smoking history enrolled in the study.

Each of the chest CT examinations was performed on a 40-slice MDCT (Brilliance 40, Philips Medical Systems, Cleveland, OH, USA). There were 18 contrast-enhanced scans and 10 non-contrast scans. The original images were reconstructed at a slice thickness of 0.9 mm and a reconstruction interval of 0.45 mm. The original data were stored at the site without any compression. These datasets were then reformatted to obtain two different slice thicknesses of 1.8 mm and 2.7 mm.

Two thoracic radiologists, each with more than ten years of experience, interpreted the studies to provide the ground truth. The two radiologists independently reviewed the datasets to identify nodules. The radiologists identified all nodules greater than or equal to 4 mm and less then 30 mm in maximum diameter. Nodules less than 4 mm in maximum diameter were excluded to avoid partial volume effects due to thick datasets.

The Fleischner society definition [15] of the nodule was used to determine the upper limit on the nodule size for inclusion in the study. The location and the slice number of each nodule were recorded on a case report form and the findings were also saved with the dataset. A second read was performed on cases where the radiologists did not agree. The radiologists were allowed to review the findings from the other radiologist during the second read. The locations of the nodules with slice numbers were also recorded from the second read and the findings were saved again. Only the nodules identified and agreed upon by both the radiologists after the second read were included as ground truth for the study.

All the ground truth nodules were identified on the cases with the reconstruction interval of 0.9 mm. The radiologists also reviewed the 1.8 mm and 2.7 mm thick datasets to verify that the nodules identified as ground truth were also visible on these datasets. The two radiologists identified a total of 33 nodules in 28 datasets. The nodules were between 4 mm and 30 mm in the long axis direction. The radiologists were able to identify all the nodules on 0.9 mm, 1.8 mm and 2.7 mm thick datasets. There were 11 studies with no nodules identified by the two radiologists. All 28 cases were then evaluated on CAD software to identify the sensitivity and false positives of the software on datasets with different reconstruction thickness.

Computer Aided Lung Nodule Detection
The CAD software used in the study is a part of Lung Nodule Application (LNA) on Philips’ Extended Brilliance Workspace (EBW). It is not commercially available in the United States. The lung nodule CAD software is engineered to detect nodules greater than 4 mm and less
The results of CAD software were displayed as square markers around the nodule for visualization purposes. The automated detection process takes about 30 seconds to 1 minute depending on the size and quality of the dataset. Once the case is loaded into the application the user can continue to work on the application and identify nodules, while CAD is working in the background. CAD generated candidates are displayed to the user as a list and also on a reference lung image (Figure 1).

If the CAD finds a nodule that the user has identified, it is placed next to the user-identified nodules in the list. The software allows a single click acceptance or rejection of any candidate. Once the user accepts the candidate the volume and area of the nodule are displayed to the user as shown in the Figure 2.

In our study a CAD finding was considered to be true positive (Figure 3) if it matched with the ground truth as identified by the radiologists, and false positive if it did not match. A false negative result was defined as a nodule missed by CAD but identified on ground truth (Figure 4).

**Statistical Analysis**

The CAD performance was evaluated by comparing all CAD identified nodules with the ground truth. Since ground truth consisted of the location and the slice number of the nodules and also the findings saved as DICOM, a straightforward comparison could be made to classify a CAD mark as a true or a false positive. The evaluation procedure was repeated for the 1.8 mm and 2.7 mm datasets. The sensitivity and the false positive rates were calculated for the CAD software on these datasets. The results were then evaluated for statistical significance using Student’s t-test.

**Results**

A total of 33 nodules in the 28 cases were identified in the study as ground truth within the range of 4 mm to 30 mm. There were 19 nodules less than 10 mm (4 mm-10 mm) in maximum diameter and 14 nodules greater than 10 mm in maximum diameter. These nodules were identified on 0.9 mm, 1.8 mm and 2.7 mm thick datasets. The CAD software detected 25, 23 and 13 out of 33 nodules on 0.9 mm, 1.8 mm and 2.7 mm reconstruction interval datasets respectively.

The average sensitivity of CAD was 75.8% on 0.9 mm data, 69.7% on 1.8 mm data and 39.4% on 2.7 mm data. The median number of false positive results per patient on the 0.9 mm, 1.8 mm and 2.7 mm data was 8.5, 9.5 and 10 respectively. It was not possible to calculate specificity because the number of true negatives was not identified for the study. The results of the study are summarized in Table 1.

Overall, the CAD missed a total of 38 nodules in all three datasets. There were 6 nodules that were missed by CAD on all the slice thicknesses. There was one nodule that the CAD missed on 0.9 mm but found on 1.8 mm and 2.7 mm. The CAD software missed 8 nodules on 0.9 mm thick datasets, 5 of those were less than 10 mm in maximum diameter and 3 nodules were greater than 10 mm in maximum diameter. Of the 10 nodules that CAD missed on 1.8 mm
datasets, 6 were less than 10 mm in maximum diameter and 4 were greater than 10 mm in maximum diameter. There were 12 nodules less than 10 mm and 8 nodules greater than 10 mm in maximum diameter in 20 nodules than CAD missed on 2.7 mm datasets. In all three datasets, the CAD software missed more nodules of less than 10 mm in diameter than nodules greater than 10 mm in diameter.

Although the CAD software missed nodules in all the different parts of the lungs, the most nodules missed by CAD were in left lower lung for all three slice thicknesses. CAD missed more nodules that were either attached to the pleura or very close to the pleura as compared with central nodules. The false positives generated by the CAD systems were almost equally distributed in the right, left, upper and lower lungs.

Statistical analysis with Student’s t-test showed that there was no significant difference (p> 0.05) in the sensitivity of the CAD software between the 0.9 mm and 1.8 mm thick datasets. The difference in the sensitivity of the CAD software was statistically significant (p< 0.05) for 0.9 mm and 2.7 mm thick data and 1.8 mm and 2.7 mm thick datasets. However, there was no statistical difference (p> 0.05) between the false positive results on the three different datasets.

<table>
<thead>
<tr>
<th>Slice Thickness</th>
<th>Number of nodules</th>
<th>Sensitivity (%)</th>
<th>False Positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9 mm</td>
<td>25</td>
<td>75.8</td>
<td>8.5</td>
</tr>
<tr>
<td>1.8 mm</td>
<td>23</td>
<td>69.7</td>
<td>9.5</td>
</tr>
<tr>
<td>2.7 mm</td>
<td>13</td>
<td>39.4 (p&lt;0.05)</td>
<td>10</td>
</tr>
</tbody>
</table>

Discussion

The accuracy of the CAD systems improves significantly for thin slice data as compared to thick slice data. A recent study [11] showed that the sensitivity of CAD on 1 mm data was 95.2% as compared to 88.6% on 5 mm data, and the number of false positives increased from 5.4 to 23.6 per study. Our results are supportive of this study, showing that sensitivity of CAD decreases from 76% to 40% and the number of false positives increase from 8.5 to 10 per case with a smaller increase in slice thickness. The reconstruction intervals used in the study were less than 3 mm because we evaluated routine chest CT cases and not screening studies. The slice thickness range we used in the study is the typical reconstruction interval increasingly used in current practice. Aside from nodules, the vessels and lung parenchyma were also sharper and thus discernable to CAD on the thin slice data as compared to thick slice data.

Ideally, the performance of the CAD system should be independent of slice thickness and other reconstruction parameters. In addition to nodules, other structures such as vessels and scars in the lung parenchyma are also affected by the partial volume effect, increasing the number of false positives in the CAD results. The common
false positive markings for a CAD system are cardiac and respiratory motion artifacts, vessel bifurcations, sharp bends in the vessels, scars, presence of diffuse lung disease and vessels in the mediastinal region. Investigators have established that if thick slice data are used in the evaluation of CAD software, overlapping reconstruction increments may help increase the efficiency of the CAD software as compared to contiguous data [11]. The CAD software missed more nodules less than 10 mm in maximum diameter as compared with nodules greater than 10 mm in diameter. The smaller nodules are highly influenced by the partial volume effect, making it more difficult for CAD to detect those.

When single-slice scanners were common there were concerns regarding the use of thin slice acquisition, mainly due to dose-related issues. With the advent of multi-detector scanners, acquisition of thin slice data is common practice. The high noise levels in the thin slice data can be compensated for by averaging the effect of noise by acquiring more data [8]. Patients can be scanned at thin slice, and thick slice data can then be reconstructed for review. CAD software can run in the background and evaluate the thin slice data to optimize its performance, while the radiologist primarily reads thick slice data to avoid being deluged by a large number of images.

Most commonly the use of CAD software is advocated as a second reader rather than as a primary reader. There are several publications verifying the utility of CAD as a second reader [8,14]. During the second read the CAD software may find a set of nodules complementary to those detected by the radiologist in the first read. Thus, it can boost the overall sensitivity of the radiologist in detecting nodules on chest CT scans. Moreover, CAD can also be used on cases that were scanned for a different clinical indication (e.g. PE, diffuse lung disease) that may have nodules as incidental findings.

A major limitation of this study is the comparatively small number of cases and nodules included in the study. We had a total of 28 cases in the study with 33 nodules. There were 11 cases with no nodules that fit our inclusion criteria and therefore were included in the study. The maximum sensitivity of the CAD software reported in the study is 76%. This number might be relatively lower than the sensitivities reported by some of the other studies. The low sensitivity can also be attributed to the small sample population and number of nodules included in the study. Nevertheless our results are in line with those published previously suggesting improved accuracy of the CAD software with decrease in reconstruction interval.

**Conclusion**
The current study demonstrates that the performance of the CAD software decreases with an increase in the slice thickness of the datasets. Although there was only a mild increase in the slice thickness, the decrease in sensitivity was substantial. No significant differences were observed between 0.9 mm and 1.8 mm datasets. There were significant differences between the 0.9 mm to 2.7 mm and 1.8 mm to 2.7 mm datasets. The false positive markings increased from 8.5 to 10 with the increase in slice thickness. We conclude that CAD performance is vulnerable to even small increases in slice thickness. Optimal performance of CAD software is achieved using the thinnest of the slice data among the narrow slice collimations currently in use.
References

“http://www.lungusa.org/”


