Characteristics of suspicious features in CT lung-cancer screening images

Philip F. Judy*[a]; Yoshiko Kanasaki[b], Francine L. Jacobson[a], Chiara Del Frate[c] ,
[a]Department of Radiology, Brigham and Women's Hospital and Harvard Medical School, 75 Francis Street, Boston, MA, USA, 02115,
[b]Tottori Univ. Japan;
[c]Univ. degli Studi di Udine, Udine, Italy

ABSTRACT

The high-frequency of suspicious non-malignant image features limit the use of CT lung-cancer screening of asymptomatic individuals. A reference database of 6 radiologists' localizations of suspicious image features was created. The frequency, sizes, shapes, margins and degree of calcification of the images features were determined. The radiologist exam report identified 50% findings reported on 5 or 6 occasions, while the CAD system identified 40%. The radiologist exam report missed 80% to 50% of suspicious, retrospectively identified in CT lung cancer images. Many were less than 4 mm. Radiologists can use a lenient criterion in experiments of previously read screening cases.

Keywords: Lung cancer screening, computer aided detection, computed tomography

1. INTRODUCTION

The chest CT examination is more sensitive than a conventional chest radiograph for the detection of image features that may represent early lung cancers. Consequently, clinical trials have been undertaken and continue to determine whether CT is useful for lung cancer screening.[1,2,3,4.5,6,7,8,9] These trials have established that the majority of the image features radiologists identified during lung-cancer screening were benign nodules. Lung cancer screening programs recognizing this limitation of treating the radiologist findings from a single CT study as positive findings use protocols that involve serial CT studies. A small finding is not regarded as positive until serial examinations detect a significant size increase of the finding. Computer procedures to precisely determine the size of an image feature have been developed.[10,11,12,13] A question remains: which findings should be followed? The obvious criterion is the size of the finding, e.g., non-calcified findings greater that 4mm to 5mm are considered suspicious enough to be followed.[14] Another criterion is the radiologist level of suspicion, estimated in this study by the frequency that the finding is considered suspicious by several radiologists. The limitation of the frequency criterion is that even the reports of several radiologists are not reliable and it is not practical.[15.] However, report frequency might prove useful in an evaluation of a computer aided detection procedure (CAD), that is, CAD should report the findings that most radiologists indicate should be followed that a single radiologist.

Another perspective of unreliability of screening using the CT examination suggested by developers of CAD is the high likelihood that actually malignant findings and image features of concern to radiologists will not be reported leading to false-negative exam. The false-negative exam potentially reassures the individual and delays diagnosis of lung cancer. False-negative screening exams have occurred,[16,17,18] and failures to detect image features of high concern [19,20] have been reported. Presumably, the large amount of image data of one CT examination of the whole chest, in which there may be many small lung nodules or other small abnormalities, exhausts the radiologists, which leads to the failure to report all critical findings. CAD has been developed and studied by several investigators.[21,22,23,24]

In this study, a reference database of 6 radiologists' localizations of suspicious image features has been created. The reference database was a list of unique locations and the frequencies that each of the locations was reported. The potential utility of such a reference database was demonstrated by comparing the performances of a lung cancer-screening program and a computer-aided detection program

Medical Imaging 2008: Computer-Aided Diagnosis, edited by Maryellen L. Giger, Nico Karssemeijer
Proc. of SPIE Vol. 6915, 69151J, (2008) · 1605-7422/08/$18 · doi: 10.1117/12.770260

2. METHODS

2.1 Case selection

The cases were selected from 40 consecutive screening cases scanned between 10/5/2000 and 1/25/2002. These cases were from a lung-cancer screening study of asymptomatic smokers using low-dose, multi-detector row CT. The cases had to be available from the PACS archive, had to be successfully transferred to the workstations used in the study and had to satisfy the section-thickness technical specification described in Scanning Technique. Twenty-eight cases were available for study. The Brigham and Women's Hospital Institutional Review Board approved the retrospective analysis of these images and clinical records for research purposes.

Sixteen of the cases were women and 12 were men. The age range of these individuals at the time of the examinations was 40-81 years (mean age, 57 years). All were asymptomatic and had a current or past history of cigarette smoking. Smoking history ranged between 6 and 110 pack-years with an average of 39 pack-years. One had a past history of a breast cancer, and one had a past history of a melanoma and uterine cancer, and the others had no history of previous cancer. Additional diagnostic imaging procedures were recommended in 13/28 cases as a consequence of screening study.

2.2 Examination protocol

All studies were performed on a multi-detector row, spiral CT scanner (Somatom Volume Zoom, Siemens Medical Solutions, Erlangen, Germany) using the following protocol: 4×1mm detector collimation; 0.5 sec/rotation; table feed, 8mm feed/rotation; 140kVp; 40mAs per rotation; and B50f reconstruction filter. The CT section thickness was either 1.25mm or 2.0mm (respectively in 11and 17 cases) and sections were obtained using 1.0mm section increments. The average number of sections per study in our data was 281, with a range of 226 to 312. This technique is essentially identical to the NLST protocol that we used with National Lung Screening Trial (NLST) participants at our institution. The NLST protocol specifies 120kV rather than 140 kV.

2.3 Radiologist image analysis protocol

On 6 separate occasions radiologists evaluated the study cases. They indicated the locations of suspicious image features in the CT sections. Radiologists were likely looking for nodular image features, however generalizing the task to identifying all potentially malignant image features provided the opportunity to report non-nodular image features. For this identification of the locations of suspicious features, the radiologists were encouraged to use a lax criterion rather a strict criterion. They were asked to report nodules that were calcified. The presentation was randomized for these 6 evaluations. In order to be concise in the rest of the presentation, a "radiologist's suspicious image feature" will be called a "radiologist finding" or when the context is clear, just "finding."

Three different workstations were used during the evaluations, however they were similar. The workstations had 48 cm monitors. Images were viewed using stack sequential display. The radiologists could scroll rapidly between images under mouse control. Radiologists used the sequential display techniques to differentiate between blood vessels and nodules.[25] The radiologists were allowed to change window and level displays value as they do in clinical evaluation of screening case. Most manipulations involved toggling between preset display values.

For 2 evaluations the radiologists used the Department's clinical PACS workstation (AGFA Impax, Ridgefield Park, NJ) to identify suspicious image features. The 512×512 CT images were displayed at a size of 37.0×37.0cm on the 1,280×1,024-line monitor. A display lung window of 2000 HU with a level of –600 HU, soft tissue window of 360 HU with a level of 36 HU, and bone window of 2500 HU with a level of 250 HU were used. The radiologists tabulated the patient table location. The one of the evaluations was a composite of 2 radiologist reports, that is, one radiologist read some of the cases and a second read the rest of the cases. The radiologists' data were entered into an Access database.

On one occasion, a fourth radiologist used a computer workstation dedicated to lung CT analysis (LungCARE, Syngo;VX26BSL02P18, Siemens Medical Solutions, Erlangen, Germany) to identify all the findings. The 512×512 CT image was displayed at a size of 27.0×27.0cm on the 1,280×1,024-line monitor (SCM21107-M, Siemens Medical Solutions, Erlangen, Germany). The initial display mode used to identify the radiologist findings was a maximum intensity projection (MIP) with a slab of 10 mm.[26,27] A further evaluation of the findings used the thin sections (1.25 or 2 mm). A lung display window of 2000 HU with a level of –600 HU was used. Using a mouse pointer of the computer workstation, she selected the center of the finding. Pressing the mouse button, the location information as x, y, z co-ordinates was stored by the computer. After she completed the reading of study cases, she used the same computer tool to identify the locations from the evaluations using the clinical PACS workstation. The x, y, z co-ordinates of each finding along with the frequency the feature was reported and transferred to the Microsoft Access database. At this point in the creation of reference database, radiologist findings were summarized in the reference database.

On 3 additional occasions, radiologists used the image evaluation software of CAD system (ImageCheckerCT, R2 Technology, Inc. Sunnyvale, CA). One radiologist was a reader from the condition using the clinical PACS workstation. Another radiologist was the reader who evaluated the case using used computer workstation dedicated lung CT analysis. The third radiologist had not participated in any previous evaluations. The each 512×512 CT image was displayed at a size of 32.0×32.0cm on the 1,600×1,200-line color monitor (Color Graphic Display:P1130, Dell Computer Corporation, Round Rock, TX). Using a mouse pointer of the computer workstation, the center of the finding was selected and location information as x, y, z co-ordinates was stored by the computer of the CAD system. The location data from the 3 radiologists were directly transferred to the reference database.

2.4 Determination of independent feature locations

The radiologist location reports from the computer workstation that was dedicated lung CT analysis and the 3 radiologists' location reports using the CAD system display contained duplicate findings. Computer analysis was performed to eliminate and count duplicate reports. The purpose of the analysis was to obtain the inventory of independent locations along with frequency the location was reported. To decide which of the image features from one source were associated with image features in a second source, the distance between all possible pair of locations was calculated and the cumulative distribution of the pairs whose distances were less than a specified value was determined. Figure 1 shows the cumulative distributions for each of the 3 radiologists compared the independent locations identified on 4 previous occasions. One radiologist visually determined and collated the independent locations. The distributions increased rapidly from 0 to 3 mm then became flat. Above 6 mm the distribution began to increase again. The increase above 6 mm was assumed to be associated with random matches of different features. Consequently, feature pairs with a distance of less than 5mm identified features were considered previously identified by the other 4 readers. The new radiologist findings were added serially to database in order to accumulate independent locations. One radiologist once reported image feature as three distinct features that were separated by less 5 mm. The three distinct reports were treated as a single finding.

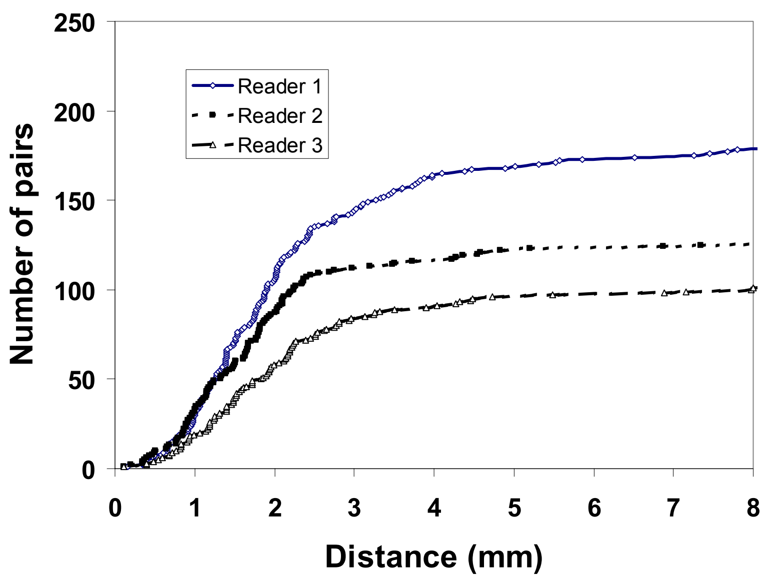

Fig 1. Cumulative count of pairs of findings less than a specified distance value. One location was obtained from one of three readers using stack display, while the other was determined by a reader using reports from 4 readers.

2.5 Other visual evaluations

Additional data was collected at the time of clinical exam. After the report was dictated, the radiologist tabulated the patient table location, lung segment. The fourth radiologist used a computer workstation dedicated to lung CT analysis to determine the spatial locations of the clinical findings. The clinical evaluation was the consensus of two radiologists and a stricter reporting criterion was used.

After the reference database of complete, the first author visually evaluated features the 614 independent location twice. A Macintosh G5 with Cinema HD Display and OsiriX display software was used. Feature was zoomed to about 1 cm on the display. On the first replication, the largest and orthogonal diameter of feature was determined and shape, margin, and calcification classification were developed. On the second replication, the largest and orthogonal diameter of feature was determined and shape, margin, and calcification classification were made.

2.6 CAD evaluations

A CAD program (ImageCheckerCT, R2 Technology, Inc. Sunnyvale, CA) was applied to the CT image data. This prototype program's findings are similar to the R2's first released version. The CAD software was designed to report all findings that are at least 4 mm in diameters. In addition, the software also reports some findings smaller than 4mm, however to avoid including too many CAD reports that might be of less clinical significance, the CAD software selected a subset of these small features with very strong nodular characteristics. The x, y, z co-ordinates were recorded and transferred to the Microsoft Access database.

The radiologists' reports and the CAD programs provided two independent sources of the location of the suspicious findings. To decide which of these locations in the reference database were associated with CAD finding, the distance between all possible pair of locations was calculated and the cumulative distribution of the pairs whose distances were less than a specified value was determined. The distributions increased rapidly from 0 to 3 mm then became flat. Above 10mm the distribution began to increase again. This increase was assumed to be associated with random matches of different features. Consequently, feature pairs with a distance of less than 5mm identified the pairs that were considered the same feature. Consequently, the procedure to determine whether a CAD finding was in reference database was identical to the procedure that was used to determine whether a radiologist finding was in the reference database.

3. RESULTS

Seventy two percent (72%) of the findings measured were less than 4 mm. There is a slight decrease in finding size as a function of report frequency. The decrease was -0.19 mm/report-frequency ($t=-389$, $p=0.011$). However unpaired t-tests determined that there were not significant differences between image feature-size pairs, although an F-test, Levene's test for equality of variances, demonstrated that the findings reported some significant differences in variance of the feature-size pairs. For example, findings reported only once had significantly greater size variation than the findings reported 7 times ($p=0.017$). Inspection of the large findings indicated that these finding were probably scars or inflammations, which tend to be substantially larger than nodular features. Interestingly, 59% of the findings reported by all 6 radiologists were less than 4 mm. No systematic differences between the sizes of difference shapes, margins, and calcification classifications could be demonstrated, because the small features could not be reliable classified.

The reference database for the 28 cases consisted of 614 independent findings (23.6 findings per case). The minimum number of findings per case was 7 and maximum number of finding per case was 49. The radiologists identified 79 to 348 findings per occasion during the 7 readings (6 experimental reading and 1 clinical reading). Seventy-nine of these findings were identified during the clinical evaluation. The average number of findings identified per occasion was 194. On average 29% of the findings in reference database were detected on each reading occasion. Every case had findings that were identified by at least 2 radiologists. Figure 2 presents fraction of findings identified the CAD program and the clinical exam as function of number of reader reports required to interpret the finding as positive (reader threshold). For example, if reader threshold was 4, then the findings for which 4, 5, or 6 readers reported the finding were considered positive. The number of findings per case decreased exponentially as the reader threshold increased.

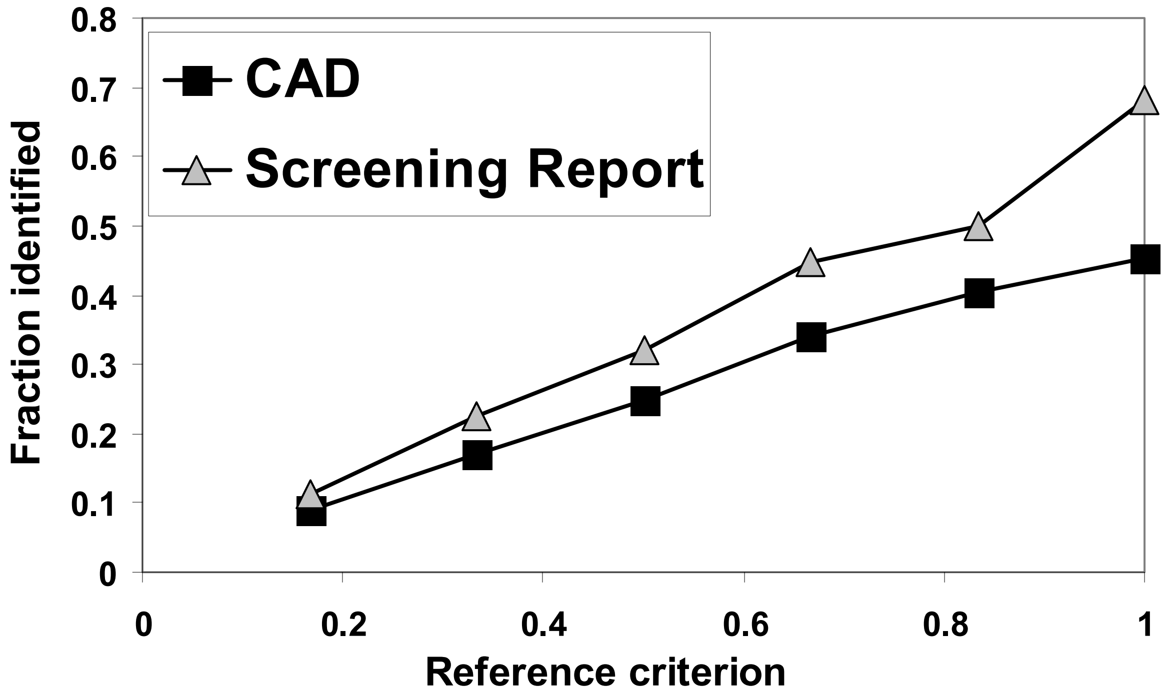

Fig 2. The fraction of positive findings in the reference database identified by CAD compared to positive findings in the reference database identified during the clinical evaluation.

4. Discussion

The reference database contained 4 radiologist findings per case that were identified by at least 2 radiologists. This value was substantially larger than we had anticipated. Consequently, a complete inventory of radiologist findings will be large compared to the number of suspicious findings typically reported in lung cancer screening. In this study using a lax criterion, on average less than 30% of the findings in the reference database were identified on each reading occasion. An automatic and organized inventory of findings will be necessary in serial lung cancer screening program because there are a large number of findings. These results suggest that an important role of CAD in CT lung cancer screening is maintaining and identifying an inventory of findings from serial screening exams.

The CAD system identified 16% positive findings in the reference database, if a positive finding is defined as a finding reported on 2 evaluations of the cases and 47% of the highly suspicious positive findings; those findings reported on all 6 experimental evaluations of the cases. The CAD system identified 1.6 false-positives per case. The CAD system performance was not substantially below to clinical evaluation of the test cases. A reason for the lower sensitivity of the CAD program on this reference database as compared to other evaluations of CAD was that the CAD program was designed to mainly detect and report nodules that are 4mm in diameter or larger and at least 60% of the findings in these cases were smaller than 4mm in diameter. The CAD performance improved slightly when smaller feature were excluded (Figure 3). By design, in order to reduce the false-positive rate, the CAD system probably dismissed the majority of these small findings. Some of these dismissed CAD findings may be present in the reference database and could be included in an inventory maintained by a CAD system and not presented to the radiologists until a subsequent

CT examination increases the likelihood that the finding is malignant, for example, the image feature increased in size. While the consequence of working-up small or ambiguous features is uncertain [28,29], malignancies are small early in their natural history.

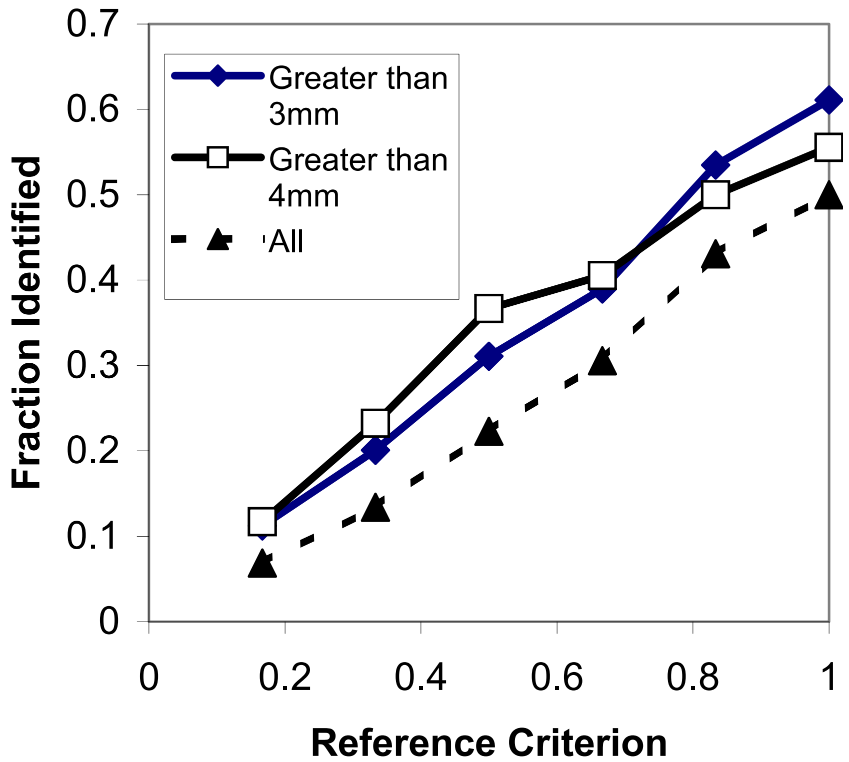

Fig 3. The fraction of positive findings in the reference database identified by CAD compared to positive findings in the reference database modified to eliminate smaller features.

The effectiveness of nodule growth measuring procedures depends on several factors, among these factors are: reliable size measurements, accurate matching of same finding on sequential CT images, and a complete identification of all suspicious findings. The last factor, the need for a complete CAD inventory of all the suspicious findings, is the topic of this investigation. We believe that an important role of CAD, when it is applied to CT lung cancer screening, is to create an inventory of findings that radiologists will classify as image features of concern. The effectiveness of CAD fulfilling this role is measured by comparison of the CAD inventory of findings to an inventory radiologist findings. Completeness of the inventory means, that all of the findings that the radiologist could be expected to identify are included, even those of identified at low rates. Completeness of the radiologist inventory was achieved by numerous evaluations, six, of the CT images by radiologists using a low-suspicion detection criterion. The radiologists' level of suspicion regarding a finding will be characterized by the frequency that radiologists identify the finding rather than the rating of the level of suspicion.

REFERENCES

[1] M. Kaneko, K. Eguchi, H. Ohmatsu, R. Kakinuma, T. Naruke, K. Suemasu, N. Moriyama, “Peripheral lung cancer: screening and detection with low-dose spiral CT versus Radiography,” Radiology **201**, 798-802 (1996).

[2] S. Sone, S. Takashima, F. Li, Z. Yang, T. Honda, Y. Maruyama, M. Hasegawa, T. Yamanda, K. Kubo, K. Hanamura, K. Asakura, “Mass screening for lung cancer with mobile spiral computed tomography scanner,” Lancet **351**, 1242-1245 (1998).

[3.] C. I. Henschke, D. I. McCauley, D. F. Yankelevitz, D. P. Naidich, G. McGuinness, O. S. Miettinen, D.M. Libby, M. W. Pasmantier, J. Koizumi, N.K. Altorki, J. P. Smith, “Early lung cancer action project: overall design and findings from baseline screening,” Lancet **354**, 99-1051 (1999).

[4] S. J. Swensen, J. R. Jett, J. A. Sloan, D. E. Midthun, T. E. Hartman, A. M. Sykes, G. L. Aughenbaugh, F. E. Zink, S. L. Hillman, G. R. Noetzel, R. S. Marks, A. C. Clayton, P. C. Pairolero, “Screening for lung cancer with low-dose spiral computed tomography,” Am J Respir Crit Care Med **165**, 508-51 (2002).

[5] S. Diederich, D. Wormanns, M. Semik, M. Thomas, H. Lenzen, N. Roos, W. Heindel, “Screening for early lung cancer with low-dose spiral CT: prevalence in 817 asymptomatic smokers,” Radiology **222**, 773-781 (2002).

[6] S. J. Swensen, J.R Jett, T.E. Hartman, D. E. Midthun, J. A. Sloan, A. M. Sykes, G. L. Aughenbaugh, M. A. Clemens, “Lung cancer screening with CT: Mayo Clinic experience,” Radiology **226**, 756-761 (2003).

[7] C.I. Henschke , D.F. Yankelevitz, D.M Libby, M.W. Pasmantie, J.P. Smith O.S. Miettinen , “Survival of patients with stage I lung cancer detected on CT screening,” N Engl J Med **355**, 1763-1771 (2006).

[8] PB, Jett JR, Pastorino U, Tockman MS, Swensen SJ, Begg CB, “Computed tomography screening and lung cancer outcomes,” JAMA;**297(9)**, 953-61 (2007)

[9] ACRIN Protocol 6654 Contemporary Screening for the Detection of Lung Cancer http://www.acrin.org/6654_protocol.html (Accessed 1/22/2008).

[10] B. Zhao, D. Yankelevitz, A. Reeves, C. Henschke, “Two-dimensional multi-criterion segmentation of pulmonary nodules on helical CT images,” Med Phys **26**, 889-895 (1999).

[11] D. F. Yankelevitz, A. P. Reeves, W. J. Kostis, B. Zhao, C. I. Henschke, “Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation,” Radiology **217**, 251-256 (2000).

[12] J. P. Ko, M. Betke, “Chest CT: automated nodule detection and assessment of change over time- preliminary experience,” Radiology **218**; 267-273 (2001).

[13] A.P. Reeves, “Measurement methods for small pulmonary nodules,” Radiology.;**246(1),** 333-4 (2008).

[14] C. I. Henschke, D. F. Yankelevitz, D. P. Naidich, D. I. McCauley, G. McGuinness, D. M. Libby, J. P. Smith, M. W. Pasmantier, O. S. Miettinen, “CT screening for lung cancer: suspiciousness of nodules according to size on baseline scans,” Radiology **231**, 164-168 (2004)

[15] J.K. Leader, T.E. Warfel, C.R. Fuhrman, S.K. Golla, J.L. Weissfeld , R.S. Avila, W.D. Turner , B. Zheng , “Pulmonary nodule detection with low-dose CT of the lung: agreement among radiologists,” AJR ;**185(4),** 973-8 (2005).

[16] T. Nawa, T. Nakagawa, S. Kusano, Y. Kawasaki, Y. Sugawara, H. Nakata, “Lung cancer screening using low-dose spiral CT: results of baseline and 1-year follow-up studies,” Chest **22,** 15-20 (2002).

[17] R. Kakinuma, H. Ohmatsu, M. Kaneko, K. Eguchi, T. Naruke, K. Nagai, Y. Nishiwaki, A. Suzuki, N. Moriyama, “Detection failures in spiral CT screening for lung cancer: analysis of CT findings,” Radiology **212**, 61-66 (1999).

[18] F. Li, S. Sone, H. Abe, H. MacMahon, S. G. 3rd Armato, K. Doi, “Lung cancers missed at low-dose helical CT screening in a general population: comparison of clinical, histopathologic, and imaging findings,” Radiology, **225**, 673-683 (2002).

[19] C. I. Henschke, D. P. Naidich, D. F. Yankelevitz, G. McGuinness, D. L. McCauley, J. P. Smith, D. Libby, M. Pasmantier, M. Vazquez, J. Koizumi, D. Flieder, N. Altorki, O. S. Miettinen, “Early lung cancer action project: initial findings on repeat screenings,” Cancer **92**, 153-159 (2001).

[20] S. J. Swensen, J. R. Jett, T. E. Hartman, D. E. Midthun, J. A. Sloan, A. M. Sykes, G. L. Aughenbaugh, M. A. Clemens, “Lung cancer screening with CT: Mayo Clinic experience,” Radiology **226**, 756-761 (2003).

[21] S. G. 3rd Armato, F. Li, M. L. Giger, H. MacMahon, S. Sone, K. Doi, “Lung cancer: performance of automated lung nodule detection applied to cancers missed in a CT screening program,” Radiology **225**, 685-692 (2002).

22 M. N. Gurcan, B. Sahiner, N. Petrick, H. P. Chan, E. A. Kazerooni, P. N. Cascade, L. Hadjiiski, “Lung nodule detection on thoracic computed tomography images: preliminary evaluation of a computer-aided diagnosis system,” Med Phys. **29**, 2552-2558 (2002).

[23] B. Zhao, G. Gamsu, M. S. Ginsberg, L. Jiang, L. H. Schwartz, "Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm," J Appl Clin Med Phys **4**, 248-260 (2003).
[24] M. S. Brown, J. G. Goldin, R. D. Suh, M. F. McNitt-Gray, J. W. Sayre, D. R. Aberle, "Lung micronodules: automated method for detection at thin-section CT-initial experience," Radiology **226**, 256-262 (2003).
[25] S.E. Seltzer, P. F. Judy, D. F. Adams, F. L. Jacobson, P. Stark, R. Kikinis, R. G. Swensson, S. Hooton, B. Head, U. Feldman, "Cine Viewing of Spiral Chest CT Images," Radiology **197**, 73-78 (1995).
[26] S. Napel, G. D. Rubin, R. B. Jr. Jeffrey, "STS-MIP: a new reconstruction technique for CT of the chest" J Comput Assist Tomogr **17**, 832-838 (1993).
[27] J. F. Gruden, S. Ouanounou, S. Tigges, S. D. Norris, T. S. Klausner, "Incremental benefit of maximum-intensity-projection images on observer detection of small pulmonary nodules revealed by multidetector CT," AJR , 149-157 (2002).
[28] R. F. Munden, K. R. Hess, ""Ditzels" on chest CT: survey of members of the Society of Thoracic Radiology," AJR **176**, 1363-1369 (2001).
[29] L. F. Rogers, """Ditzels": Little Things Mean a Lot," AJR **176**, 1355 (2001)