An Analysis of Early Studies Released by the Lung Imaging Database Consortium (LIDC)[1]

James C. Ross, James V. Miller, Wesley D. Turner, Timothy P. Kelliher

Rationale and Objectives: To analyze radiologist lung nodule segmentations in the Lung Imaging Database Consortium (LIDC) database and to apply statistical tools to generate estimates of ground truth. This investigation expands on earlier work by considering a larger number of cases from the LIDC database, and results were generated on a per-nodule basis, as opposed to a per-case basis as was done previously.

Materials and Methods: We analyzed nodule data drawn from the 41 most recent computed tomography exams released by the LIDC. We combined radiologist segmentations for a given nodule using different consensus schemes: union, intersection, and simultaneous truth and performance level estimation (STAPLE). We also generated three-dimensional models of the manual segmentations using discrete marching cubes to visualize features of the data.

Results: Using the union as the consensus scheme produced the greatest number of nodule-positive voxels while using the intersection produced the fewest. Considering only nodules for which all readers agreed on nodule presence, STAPLE computed sensitivity averages for readers one, two, three, and four were 0.91, 0.83, 0.90, and 0.77, respectively. Specificity averages were 0.97, 0.98, 0.97, and 0.97. Considering cases for which there was disagreement about nodule presence, sensitivity results become 0.67, 0.74, 0.60, and 0.37. Specificity results in this case are 0.95, 0.95, 0.95, and 0.98. STAPLE-generated *pmaps* exhibited probability values tightly grouped below the 0.25 and above the 0.75 probability levels. Three-dimensional models of manually segmented nodules revealed step-artefacts in the segmentation data.

Conclusions: Radiologists often disagree about nodule presence. Ideally, knowing each reader's sensitivity and specificity a priori is preferred for optimal STAPLE results. Knowing these values and developing manual segmentation tools and imaging protocols that mitigate unwanted segmentation features (such as step artefacts) can result in more accurate estimates of ground truth. Furthermore, a computer-aided detection algorithm's performance is a function of the ground truth estimate by which it is scored.

Key Words. Lung; nodules; ground truth; Lung Imaging Database Consortium (LIDC); computer-aided detection (CAD).

© AUR, 2007

Despite warnings stretching back 40 years—the initial US Surgeon General's Report on Smoking was issued in 1964 and warning labels have been required on cigarettes sold in the United States since 1969—lung cancer remains a serious health threat in the world with an estimated 1 million deaths in 2000 (1). For the United States, lung cancer is the single largest fatal malignancy and is second only to heart disease in yearly fatalities; it makes up 32% of the cancer deaths among men and 25% of the cancer deaths among women. There are few effective treatments and 5-year mortality is approximately 14% (2,3) because largely to advanced stage of the disease

Acad Radiol 2007; 14:1382–1388

[1] From General Electric Global Research, Research Circle, KW-C405, Niskayuna, NY 12309. Received April 28, 2007; accepted August 3, 2007. This publication was supported by the DOD and the Medical University of South Carolina under DOD Grant No. W81XWH-05-1-0378. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Department of Defense or the Medical University of South Carolina. **Address correspondence to:** J.C.R. e-mail: ross@research.ge.com

© AUR, 2007
doi:10.1016/j.acra.2007.08.004

at diagnosis. Despite this, lung cancer screening is not recommended, even for at-risk populations, based largely on statistically powerful studies carried out in the 1970s that showed little benefit to lung cancer screening. However, imaging advances over recent years—most notably the advent of thin slice multidetector computed tomography (MDCT)—have prompted revisiting the notion of screening.

Recent studies of stage I cancers demonstrate that the diagnosis of lung cancer before metastases can give 5-year survival rates as high as 70% (1,4) and other studies (5) equate small, subtle cancers with early-stage disease. But before screening becomes widely accepted, issues involving overdiagnosis, radiation dose, and appropriate population selection must be addressed. However, a dose-limiting screening protocol that enables early nodule detection while limiting overdiagnosis appears within reach. One of the promising vehicles for early detection is the low-dose MDCT scan. MDCT has been shown to be an effective method for detecting small (<1 cm) nodules in the lungs that may be cancerous or precancerous lesions. However, the rigors of reading a large number of MDCT chest scans with upwards of 500 images per scan, combined with the subtlety of some of nodules of interest argue for computer-aided detection (CAD) as a necessity for a rational screening program.

In this environment, the National Cancer Institute formed the Lung Imaging Database Consortium (LIDC) to study the barriers to effective CAD development and to develop a database as a national resource that can be used to expedite development (6). One of the major barriers to the development and evaluation of effective CAD devices is the absence of rigorous ground truth (GT). Unlike other cancer screening applications, the determination of malignancy in a discovered lung lesion is not undertaken for small, subtle lesions because of the likelihood of morbidity during the procedure. Although histologic analysis of biopsy samples provides the most reliable assessment of malignancy, radiologic diagnosis based on medical images serves as a surrogate for an actual malignancy determination. Performance of CAD algorithms tends to be very sensitive to the choice of GT. Consider the chart in Fig 1, which represents a set of lung scans read by three radiologists. The data were generated based on lung cancer screening exams in the GE Global Research cancer database. The y-axis shows how many nodules were detected by one, two, or all three radiologists reading the exams. Depending on whether single-radiologist GT, majority GT, or unanimous GT is chosen, the performance

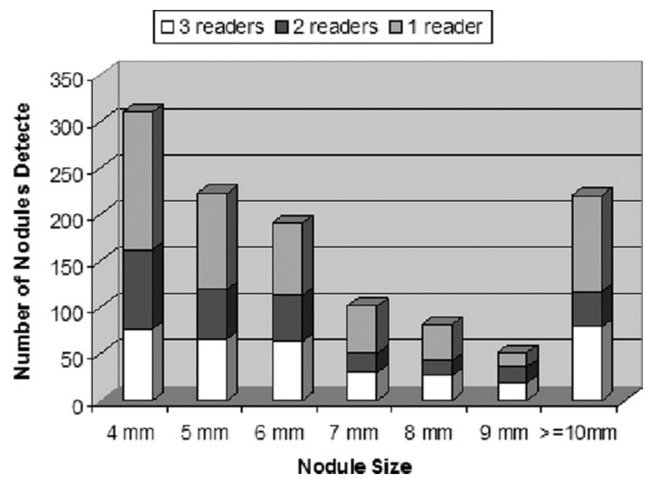

Figure 1. The variability of nodule ground truth based on radiologic consensus.

of a typical CAD system will change drastically. As an example, a CAD algorithm with 100% sensitivity on unanimous GT could have much lower sensitivity on single reader GT, and an algorithm with 100% specificity for single-reader GT could gain hundreds of false positives using unanimous GT.

To get the best possible GT under these constraints, and to capture nodule characteristics such as spiculations, lobulations, and density, the LIDC asked multiple radiologists to perform a series of blinded and unblinded reads on a number of CT lung images. Each radiologist first read the lung cases without knowledge of how his or her colleagues marked the exams and placed contours around the periphery of nodules found. The radiologists also described the nodules by scoring features such as texture, lobulation, spiculation, subtlety, margin, and sphericity. After all initial reads where completed, a further round of unblinded reads were performed in which the radiologists could modify their own contours based on the contours of their colleagues. In the end, all four sets of final contours were saved and provided as annotations to the exams. This allows CAD and cancer researchers to determine an appropriate method to combine the GT for their research to maximize detection and diagnostic capabilities of their algorithms.

In this article, we investigate nodule data drawn from the first 41 LIDC datasets released. We use statistical tools such as Simultaneous Truth and Performance Level Estimation (STAPLE) (7) available in the Insight Segmentation and Registration ToolKit (8) to both extract estimations of GT contours from the multiple radiologist marks and to characterize reader performance. We also

Download English Version:

https://daneshyari.com/en/article/4219470

Download Persian Version:

https://daneshyari.com/article/4219470

Daneshyari.com