The Lung Image Database Consortium (LIDC): A Quality Assurance Model for the Collection of Expert-Defined "Truth" in Lung-Nodule-Based Image Analysis Studies

S. G. Armato III, R. Y. Roberts, G. McLennan, M. F. McNitt-Gray, D. Yankelevitz, E. A. Kazerooni, E. J. R. van Beek, H. MacMahon, D. R. Aberle,
C. R. Meyer, A. P. Reeves, C. I. Henschke, E. A. Hoffman, B. Y. Croft, L. P. Clarke

University of Chicago, University of Iowa, UCLA, Cornell University, University of Michigan, National Cancer Institute

ABSTRACT

CAD development requires the initial establishment of "truth." p4otential inconsistencies in "truth" data must be identified and corrected before investigators can rely on this data. We developed a quality assurance (QA) model to supplement the "truth" collection process for lung nodules on CT scans. A two-phase process was established for the interpretation of CT scans. The final set of marks underwent QA, which consisted of identification of potential errors that occurred during the reading process and error correction. Six categories of potential error were defined, and any nodule with a mark that satisfied the criterion for one of these categories was referred to the radiologist who assigned the mark in question. The radiologist either corrected the mark or confirmed that the mark was intentional. 829 nodules were identified in 100 CT scans. The QA process yielded 86 potential errors. The establishment of "truth" must include QA to guarantee the integrity of the "truth" that will provide the basis for training and testing CAD systems.

Figure 1. Radiologist annotations: "nodule $\geq$ 3mm" (left), represented by an outline constructed by the radiologist, "nodule < 3mm" (center), represented by a hexagon positioned at the nodule centroid, and "non-nodule $\geq$ 3mm" (right), represented by an "x" at the lesion centroid.

METHODS

The Lung Image Database Consortium (LIDC) is developing a publicly available resource for the medical imaging research community (1-3). A two-phase process was established for the asynchronous interpretation of CT scans by a panel of four radiologists at different institutions. During the initial "blinded read" phase, radiologists independently marked lesions they identified through a computer interface as "nodule $\geq$ 3mm," "nodule < 3mm," or "non-nodule $\geq$ 3mm" (Figure 1). During the subsequent "unblinded read" phase, the blinded read results of all radiologists were revealed to each radiologist, who then independently reviewed their marks along with the anonymous marks of their colleagues; a radiologist's own marks could be left unchanged, deleted, switched in terms of lesion category, or additional marks could be added. This approach was developed to identify, as completely as possible, all nodules in a scan without requiring forced consensus (Figure 2). Recognizing the extent to which the scientific community will rely on this information, the post-unblinded-read data was subjected to a specially designed quality assurance (QA) process.

Figure 2. The two-phase reading process for the asynchronous interpretation of CT scans at different sites. A scan is sent to four sites, and an experienced thoracic radiologist at each site identifies appropriate lesions through the blinded read phase, the annotations of which are recorded in XML files. A single XML file that merges blinded read annotations is sent to the same radiologists to initiate the unblinded read phase. The single XML file that merges unblinded read annotations from the unblinded reads provides the basis for the QA process.

The QA process consists of (1) the manual identification of potential errors that occurred during the reading process and (2) the resolution of these errors by the radiologists who assigned the marks in question. All marks placed by the four radiologists during the unblinded reads later were visually inspected, and individual marks were grouped into discrete nodules. "Non-nodule $\geq$ 3mm" marks were not considered unless they were spatially contiguous with nodule marks. The assignment of marks to specific lesions provided the basis for the QA process.

Six QA categories were defined:

1) errant marks on non-pulmonary regions of the image or stray marks within the lungs (Figure 3)
2) marks from multiple categories assigned to the same lesion by the same radiologist (Figure 4)
3) more than a single "nodule < 3 mm" mark assigned to the same lesion by a single radiologist (Figure 5)
4) "nodule $\geq$ 3 mm" contours for a single lesion that are recorded as multiple lesions across the sections
5) "nodule $\geq$ 3 mm" contours for a single lesion that are not contiguous across sections
6) lesion marked as "nodule $\geq$ 3 mm" by 3 radiologists that was assigned no mark at all by the fourth (Figure 6)

Figure 3. QA category 1 error: this mark was removed during the QA process.

Figure 4. QA category 2 error: A lesion that was marked as both a "non-nodule $\geq$ 3 mm" and a "nodule $\geq$ 3 mm" by the same radiologist, who removed the latter mark during the QA process.

Figure 5. QA category 3 error: A lesion that received two "nodule < 3 mm" marks from the same radiologist on different sections. The second mark was removed during the QA process.

Figure 6. QA category 6 error: A lesion marked as a "nodule $\geq$ 3 mm" by three radiologists; the fourth radiologist assigned no mark at all. The QA process confirmed that the "no mark" was intentional.

Any lesion with a mark that satisfied the criterion for one of these categories was referred to the radiologist who assigned the mark. That radiologist either (1) corrected the mark to resolve the inconsistency or (2) confirmed that the mark was intentional. Any modifications to the annotated marks were stored in revised XML files that represent the final reads of the scans.

RESULTS

After the unblinded reads, a total of 829 nodules were identified by at least one radiologist in the 100 CT scans. After review of the radiologists' marks, the QA process

yielded 42 scans (42.0%) with 86 QA issues, of which 82 (95.3%) resulted in modifications by the radiologists. The four potential errors that were not modified pertained to four separate lesions to which three radiologists assigned "nodule $\geq$ 3 mm" marks and the fourth radiologist assigned no mark at all (category 6 error); on further review during the QA process, the lone radiologist who did not mark each of these four lesions confirmed that their decision to assign no mark was intentional (see Figure 6). The final set of XML files included a total of 812 nodules, as identified by at least one radiologist. 17 lesions were no longer considered "nodules" after the QA process (Figure 7).

The method through which expert observers define "truth" is an important aspect of CAD studies. Potential inconsistencies in the acquired "truth" data must be identified and corrected before investigators can rely on this data for CAD performance assessments. The LIDC QA model identified errors in the "truth" marks from 42 of 100 CT scans; if left uncorrected, these errors could adversely affect the assessment of CAD methods applied to these scans. A detailed process of establishing "truth" for lung nodule detection studies must incorporate a QA model to guarantee the integrity of the "truth" that will provide the basis for training and testing CAD systems.

Figure 7. A lesion that received a "nodule < 3 mm" mark (the only nodule mark assigned to this lesion) and, in another section, a "non-nodule $\geq$ 3 mm" mark from one radiologist. The "nodule < 3 mm" mark was removed during the QA process, thus eliminating "nodule" status for this lesion in the final assessment.

REFERENCES

1. L. P. Clarke, B. Y. Croft, E. Staab, et al., "National Cancer Institute initiative: Lung image database resource for imaging research," Acad Radiol, **8**, 447–450, 2001.
2. S. G. Armato, III, G. McLennan, M. F. McNitt-Gray, et al., "Lung Image Database Consortium: Developing a resource for the medical imaging research community," Radiology, **232**, 739–748, 2004.
3. L. E. Dodd, R. F. Wagner, S. G. Armato, III, et al., "Assessment methodologies and statistical issues for computer-aided diagnosis of lung nodules in computed tomography: Contemporary research topics relevant to the Lung Image Database Consortium," Acad Radiol, **11**, 462–475, 2004.
4. S. R. Rogers, M. S. Brown, J. G. Goldin, et al., "Automated lung nodule detection in CT: Nodule inclusion criteria for determining ground truth," Radiology, **225(P)**, 407, 2002.
5. D. Wormanns, K. Ludwig, F. Beyer, et al., "Detection of pulmonary nodules at multirow-detector CT: Effectiveness of double reading to improve sensitivity at standard-dose and low-dose chest CT," Eur Radiol, **15**, 14–22, 2005.
6. J. K. Leader, T. E. Warfel, C. R. Fuhrman, et al., "Pulmonary nodule detection with low-dose CT of the lung: Agreement among radiologists," AJR, **185**, 973–978, 2005.
7. C. L. Novak, J. Qian, L. Fan, et al., "Inter-observer variations on interpretation of multi-slice CT lung cancer screening studies, and the implications for computer-aided diagnosis," SPIE Proceedings, **4684**, 68–79, 2002.

Supported in part by U01 grants CA091085, CA091090, CA091099, CA091100, CA091103