Use of Image Processing Presets in Chest Radiography

Elizabeth Krupinski, Theron Ovitt, and Michael Evanoff

IMAGE PROCESSING in radiology is used to improve the perception of diagnostic information in the image so that the radiologist can make a more confident and accurate decision.¹ With the advent of computers in radiology for displaying images, there is a huge variety of image processing techniques being developed. The most abundant types of image processing techniques are point processes (eg, window-level operations) and area processes (eg, spatial filtering operations).² The goal of the present study was to determine the influence of various image processing functions on the accuracy of diagnosing chest images displayed on a computer monitor. A secondary goal was to determine how often the various processing functions were used. The value of such information is twofold. On the one hand it helps those designing the image processing functions determine if the functions are actually performing their intended function to improve diagnostic accuracy. On the other hand, the patterns of processing use may help in the design of the computer interface, making systems more "user-friendly".

METHODS AND MATERIALS

A total of 168 computed radiography chest images were selected from cases at the Toshiba General Hospital and the University of Arizona Department of Radiology. A total of 104 of the cases contained subtle instances of nodule/mass ($n = 38$), pneumonia/interstitial disease ($n = 51$), or bronchitis/bronchiectasis ($n = 15$) that were missed by the original reader. A total of 64 of the cases were lesion-free with respect to chest pathologies.

Six image processing presets were developed for viewing chest images on a $2K \times 2.5K$ monitor. The development was an iterative process using an experienced chest radiologist and an experienced pulmonologist to evaluate the presets. The six presets were: linear (similar to film), lung window, mediastinum window, black and white reverse, low-pass and high-pass filters. The monitor on which the images were displayed was perceptually linearized before the experiment.³

Six radiologists experienced in reading chest exams read the images in two 1-hour sessions. For each image, the radiologists determined the diagnostic category (normal, nodule, pneumonia, bronchiectasis) for the case and reported their confidence in that decision using a 6-level scale. Readers gave two reports: the first without use of the presets (images appeared using the linear, film-like function); then they could use any of the presets and revise the original decision. The computer recorded each time any image processing function was used and how long the radiologists took to diagnose each image. At the end of the study, the radiologists were asked to fill out a questionnaire rating the presets and overall image quality.

RESULTS

The decision confidence ratings were used to generate Receiver Operating Characteristic (ROC) area under the curve (Az) values for the two conditions (reading without presets, reading with presets). The overall Az for the first reading without use of presets was 0.8506 (sd = 0.0124), and the Az for the second readings with presets was 0.8497 (sd = 0.0129). There was no statistically significant difference in diagnostic accuracy between the two reading conditions. The correlation between confidence rating for decisions with and without use of presets was 0.901.

The false-negative (FN) errors could be classified in two ways: omission errors (lesion present but not reported) or misclassification errors (lesion present and reported, but as the wrong type of lesion). False-positive (FP) decisions occurred when no lesion was present, but one was reported. There was a total of 284 errors made of 1,008 total decisions by the six radiologists, and 75% of the errors were FN-FN decisions (FN for both decisions). Of these FN-FN errors, 52% were misclassifications for both decisions; 40% were omissions for both decisions; 4% were missed on the first decision and misclassified on the second; and 4% were misclassified on the first decision and missed on the second. A total of 7% of the errors were FN-TP (true-positive). Of these, the original FN was an omission 74% of the time and a misclassification 26% of the time and 7% of the decisions were TP-FN. Of these, 53% were reported on the first decision, then not reported on the second, and 47% were reported and classified correctly on the first decision, then misclassified on the second. There was a total of 32 false-positive errors. Twenty-eight percent were FP-TN and 72% were TN-FP.

On average, the linear preset was used 1.29 times per case during the second decision process; lung window 1.00 times on average; mediastinum 0.418

From the Department of Radiology, University of Arizona, Tucson, AZ.

Address reprint requests to Elizabeth Krupinski, PhD, Department of Radiology, University of Arizona, Tucson, AZ 85724.

*Copyright © 1997 by W.B. Saunders Company
0897-1889/97/1003-1060\$3.00/0*

times on average; reverse 0.519 times on average; low-pass 0.042 times on average; and high-pass 0.568 times on average per case. For all of these presets, there were no differences between lesion types (ie, used the same amount for nodule, pneumonia, bronchiectasis and normal cases).

The processing functions were rated as good or excellent (choices were poor, fair, good, excellent) by 4 of 6 readers for linear; 5 of 6 for lung; 4 of 6 for mediastinum; 5 of 6 for reverse; 1 of 6 for low-pass; and 5 of 6 for high-pass. All readers rated image resolution adequate for diagnosis. When asked how they would prefer the images to appear for the first time viewing, all readers preferred some combination of lung and the high-pass filter.

DISCUSSION

The ROC results indicate that the diagnostic accuracy of the radiologists was not influenced by the use of the image processing tools. However, there were changes in decisions made from the first (without presets) to the second decision (with presets). The FN-TP and TP-FN decisions balanced each other out, although the TP-FN decisions are obviously not desirable—especially the 53% that were originally reported then not reported at all after image processing. The FN-FN category was the most prevalent (75% of all errors). On the one hand, it is fortunate that 52% were misclassification errors for both decisions—at least the abnormality was detected, just not classified correctly. On the other hand, 40% were missed with and without the

use of image processing tools. This suggests that there are some lesions that are not helped by image processing. No consistent pattern was found for the types of lesions missed. These diagnostic accuracy results suggest overall, that image processing might best be used with caution—not all decision changes made with image processing tools are for the better.

The data for frequency of preset use and the subjective ratings suggest that most radiologists prefer the lung and/or high-pass filter settings for viewing chest images. It might be advantageous for workstation design if this setting was used to present chest images originally to the radiologists. The fact that there was no obvious pattern to the use of presets correlated to type of lesions suggests that radiologists may use image processing tools in a more global way to make images look the way they prefer, rather than to improve the visibility of specific lesion types. In a sense, if a lesion can be detected without image processing, the image processing may just be used to help confirm the diagnosis. If a lesion cannot be detected without image processing, it may be unlikely that image processing will help.

REFERENCES

1. Kundel HL, Revesz G, Stauffer HM: Evaluation of a television image processing system. *Invest Radiol* 3:44-50, 1968
2. Lindley CA: *Practical Image Processing*. New York, NY, Wiley, 1991
3. Pizer SM: Intensity mappings to linearize display devices. *Computer Graphics & Image Processing* 17:262-268, 1981