Developing personalized decision support tools in radiology

Georgia Tourassi

Integrating eye-tracking technology can improve the clinical utility of computer-assisted detection technology by matching decision support to the individual radiologist.

Computer-assisted detection (CADe) technology was introduced to help radiologists by guiding their attention to suspicious features that deserve careful consideration when imaging a patient. For example, when screening women for breast cancer, CADe technologies can be used to highlight suspicious lesions on a mammogram image. CADe promised to reduce the risk of false-negative errors, bridge the performance gap between radiologists, and alleviate visual strain and cognitive fatigue by offering an independent ‘pair of eyes.’ Such support is highly desirable for breast cancer—and other image-based—screening programs because of the inherent image complexity and large amount of image data that must be interpreted for each patient. CADe technology received US Food and Drug Administration (FDA) approval 15 years ago, but despite continuing technical improvements, cost-effectiveness within an acceptable range, and widespread clinical use, its clinical utility remains one of the most controversial issues in radiology. For example, radiologists’ diagnostic accuracy often improves when using CADe, but in some cases radiologists reportedly experience no measurable impact—or even a detrimental effect—when using such systems. Such disparate findings have called into question the complexities of the human-CADe interaction and whether CADe is good enough for clinical use.

The CADe research community has accordingly pursued improvements to the technology with respect to its standalone accuracy (i.e., without the radiologist in the loop). Although such research and development efforts are necessary to advance the techniques, an equally important but largely underappreciated issue is that human users—the radiologists—respond differently when asked to make high-risk decisions under the influence of imperfect advice. That is, individual radiologists make different diagnostic decisions (and errors), and also make different judgments on whether they will accept or reject a CADe opinion. Thus, when introducing a highly accurate but still imperfect clinical decision-support system for providing a diagnosis, the ‘independent’ second-reader role the FDA has approved for CADe might not be the most effective way to realize the full potential of the technology.

We believe that tailoring or customizing decision support to the individual radiologist and case context is a more effective way to maximize the clinical effectiveness of CADe in terms of helping radiologists improve their diagnostic accuracy compared with conventional CADe. Existing CADe systems are already highly accurate in assessing image content in an independent manner, but are not sophisticated enough to select useful information without overwhelming the individual user.
Table 1. Standalone performance of user-adaptive CADe. For comparison, the standalone performance of the same CADe system operating in conventional, one-size-fits-all mode is 85.7% sensitivity with 3.15 false positives per image.

<table>
<thead>
<tr>
<th>Radiologist</th>
<th>User-adaptive CADe</th>
<th>Sensitivity (Image)</th>
<th>False positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100%</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>85.7%</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>92.9%</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>92.9%</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>100%</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>92.9%</td>
<td>0.40</td>
<td></td>
</tr>
</tbody>
</table>

We began with a pilot study\textsuperscript{10} to test whether a ‘context-sensitive’ CADe tailored to the individual radiologist’s visual search and cognitive pattern is more effective than the conventional ‘one-size-fits-all’ CADe, which gives the same visual cues without taking into account the radiologist’s individual needs for the specific case.

We asked six radiologists to interpret 20 single-view mammograms while their gaze was monitored via an eye-tracking device.\textsuperscript{10} For each image location reported by the radiologist as suspicious, three gaze metrics were collected: total dwelling time, time from beginning of reading to when the radiologist fixates on the reported location for the first time, and number of returns, which is the number of times the radiologist refixates on a particular image location. Image locations that attracted prolonged dwelling but were not reported by the radiologists were also recorded. Then we used our in-house CADe system to analyze the same images. Instead of deploying the system with the same operating threshold for all readers, we designed the system to operate with local image thresholds guided by the radiologist’s dwelling and reporting characteristics for the specific image location. Specifically, we applied three intuitive rules. Rule 1 is that if the radiologist reports a location, the CADe should operate with a lax decision threshold. Rule 2 states that if the radiologist dwells on a location but does not report it, the CADe system should operate with a moderate decision threshold. Rule 3 stipulates that for all image regions that attract short or no dwell, the system should operate with a strict threshold to avoid too many false-positive errors. We found that performance differed across users (since the operating mode is user-dependent) but that the user-adapted CADe had higher sensitivity and higher specificity than the same CADe system operating in a conventional mode with the same fixed threshold for all users (see Table 1).

We also studied the impact CADe would have on radiologists’ performance if they used it as a second reader. Figure 1 shows that half of the radiologists would benefit substantially with user-adaptive CADe, but the other half would get the same benefit as they would from the conventional CADe. Thus, leveraging the individual user’s perceptual and cognitive behavior can increase the potential benefit of a typical CADe system. With conventional CADe the radiologists’ average sensitivity would increase significantly from 85.7 to 89.3%, whereas with user-adaptive CADe, the improvement would be even larger at 94.1%.

In a follow-up study using the same data from this group of radiologists, we established strong links between medical image content and the perceptual and cognitive behaviors the radiologists displayed when viewing a mammographic image.\textsuperscript{11} For example, we observed that merging the radiologists’ case-specific gaze metrics and cognitive behaviors with computer-extracted features that capture the textural content of the image could predict the radiologists’ risk of diagnostic error. A secondary finding was that group-based understanding of the cognitive and perceptual behavior of the radiologists (less-experienced versus experienced practitioners) cannot adequately capture individual behavior. These results further encourage a paradigm shift in the way we develop and integrate computerized decision support systems in radiology. A human-centered CADe design is not only feasible but could be superior to the one-size-fits-all approach currently available in clinical practice.

Our next steps are to validate our findings with a large-scale study and continue to investigate innovative ways to establish a synergistic collaboration between humans and CADe. We have been exploring ways to capture richer spatiotemporal gaze and image context beyond what we included in the pilot study. Apart from the algorithmic challenges, there are some practical usability concerns with everyday use of eye-tracking technology in the radiology reading room. However, eye-tracking devices are becoming rapidly more user-friendly, robust, and integrated with soft-copy displays. We believe that with continuing advances in eye-tracking technology and related data analytics, the future of computer aids in radiology is very exciting. We envision a smart CADe system that anticipates and dynamically adapts to meet the evolving needs of the individual radiologist and case under review.

Continued on next page
This work was partially supported by NIH/NCI (National Institutes of Health/National Cancer Institute) R56 CA101911 and the Department of Energy (Laboratory Directed Research and Development Fund). This article has been authored by UT-Battelle, LLC, under contract DE-AC05 00OR22725 with the US Department of Energy. The United States Government retains—and the publisher, by accepting the article for publication, acknowledges—a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce this manuscript, or allow others to do so, for United States Government purposes.

Author Information

Georgia Tourassi
Biomedical Sciences and Engineering Center
Oak Ridge National Laboratory
Oak Ridge, TN

Georgia Tourassi is director of the Biomedical Science and Engineering Center and the Health Data Sciences Institute at the Oak Ridge National Laboratory. She is also adjunct professor of radiology at Duke University, Durham, NC. Her current research is in the areas of biomedical imaging informatics and big health data analytics.

References