

Human, Animal & Computer-Based Medical Image Interpretation: What Can We Learn?

Elizabeth A. Krupinski ^{1*}, Michel de Lange,² Victor M. Navarro ³, Edward A. Wasserman ³, Richard M. Levenson ⁴

¹ *Department of Radiology & Imaging Sciences, Emory University; ekrupin@emory.edu*

² *Department of Statistics University of Amsterdam michel_de_lange@yahoo.co.uk*

³ *Department of Psychological & Brain Sciences, University of Iowa; victor-navarro@uiowa.edu; ed-wasserman@uiowa.edu*

⁴ *Department of Pathology & Laboratory Science, UC Davis; levenson@ucdavis.edu*

Rationale

It is well known that human observers, even the most expert mammographers, can miss lesions in breast images as well as make false positives. There has been much research into the nature and causes of these errors, but there is still much to learn. We have a set of mammographic images that in separate studies were evaluated by mammographers, pigeons (whose visual systems are similar in many ways to those of humans), and a machine learning algorithm. Perhaps by studying these different “observers” we can learn more about the causes of interpretation errors and thereby develop new tools and techniques to obviate them.

Methods

Two sets of mammograms, one with masses and one with microcalcification clusters in half the images, were viewed by the three sets of observers in 3 independent studies: 1) mammographers (faculty and senior residents), 2) pigeons, and 3) a machine learning algorithm. Detection performance was measured for the microcalcification images and discrimination performance (benign vs malignant) for the mass images.

Results

For all three sets of “observers”, performance was better with the calcification than for the mass images and in all cases, even with the pigeons, performance was significantly better than chance. In the pigeon and machine learning studies, performance effectively transferred from training to test sets. In the pigeon study there were inter-observer differences similar to those seen in human observer studies. Some of the pigeons readily learned the task and transferred their acquired skills to new test images; while others had difficulty learning the task and generalizing to new images.

Conclusions

Analysis of image interpretation data by “observers” other than experienced trained observers may help us better understand the mechanisms of medical image perception and may prove useful in quality assessment by serving as surrogates in several types of studies. The inter-observer differences observed in the pigeon study parallel what is often observed in radiology residents – many tend to excel in some image interpretation tasks or modalities, but never excel in others.