JMI Journal of Medical Imaging

GUEST EDITORIAL

Special Issue Editorial: Medical Image Perception and Observer Performance

Karla Evans,[a] Asli Kumcu,[b] and Elizabeth A. Krupinski[c,*]
[a]University of York, Department of Psychology, York, United Kingdom
[b]Ghent University, TELIN Department, Ghent, Belgium
[c]Emory University, Department of Radiology & Imaging Sciences, Atlanta, United States

Medical image perception is a multidisciplinary field with the goal of better understanding the role of perceptual and cognitive factors in medical image interpretation. The Medical Image Perception Society (MIPS) provides a venue for radiologists, physicists, engineers, statisticians, psychologists, and others to meet and discuss the latest research on medical image perception at its biannual conference. The MIPS XIX Conference was held July 17 through July 20, 2022, in York, England, and was organized and hosted by multinational team of Karla Evans, PhD, from the Complex Cognitive Processing Lab at the University of York; Asli Kumcu, from the Image Processing and Interpretation group at Ghent University in Belgium; and Elizabeth Krupinski, PhD, from the Department of Radiology & Imaging Sciences at Emory University in the USA. This special issue contains 17 papers from the conference covering a broad spectrum of topics. Our next MIPS meeting will be held in 2024 at Vanderbilt University in Nashville, Tennessee, hosted by Frank Tong, PhD.

The MIPS meeting has a long history of sponsoring trainees' attendance and participation (128 to date). The MIPS XIX meeting provided financial support for trainee attendance using NIBIB/NCI 1R13EB031602-01and through MIPS member dues. There were six MIPS trainee scholars from six institutions and four countries. All scholars had abstracts that were accepted and presented at the conference. MIPS scholars included: Stephen Adamo (United States), Ann Carrigan (Australia), Andrew Chen (United Stats), Asli Kumcu (Belgium), Yueran Ma (Wales), and Haoqi Wang (United States).

The keynote speaker was Kenneth Young, PhD, Head of Research at the National Coordinating Centre for the Physics of Mammography (NCCPM), which is part of the National Health Service in Britain (equivalent to the US NIH). His talk covered a number of important topics including physics of breast cancer imaging with X-rays, measurement of radiation doses in mammography, risk and benefit analysis for mammographic screening, optimization of mammographic image quality and radiation dose, measurements of breast composition, standards and performance of digital mammography systems, objective assessment and clinical relevance of image quality, simulation of the mammographic imaging process, and digital breast tomosynthesis.

In "Eye-tracking differences between free text and template radiology reports: a pilot study," Krupinski et al. found that template-based radiology reports have significant potential to alter the way radiologists view images and report on them, spending more time viewing the report monitor rather than diagnostic images compared to free text dictation.

1 Visual Search & Perceptual Processes

In "Gaze patterns reflect and predict expertise in dynamic echocardiographic imaging," Laubrock et al. studied differences between experts and medical students in this underexplored application

*Address all correspondence to Elizabeth A. Krupinski, ekrupin@emory.edu

area. As shown on the cover of this special issue, students fixated on highly salient but diagnostically less relevant areas such as the mitral valve (the larger central image), whereas experts fixated on diagnostic regions earlier and more often (smaller image) while performing significantly better at discriminating patients from healthy controls. The results may be useful in optimizing echocardiography training and user interfaces, and for objectively evaluating echocardiographic expertise in trainees.

Adamo et al. used the Open Virtual Clinical Trials (VCT) framework to simulate breast anatomy and lesions in "Assessing satisfaction of search in virtual mammograms for experienced and novice searchers." Experienced and novice searchers were instructed to search for up to two lesions (masses and calcifications) per image. The results show that experienced and novice searchers made a significant amount of SOS errors, similarity had little impact on experienced searchers, but novices tended to miss a dissimilar second lesion compared to when it was similar to a detected first lesion, and both types of searchers were faster at finding similar compared to dissimilar second lesions.

The results of "Visual hindsight bias for abnormal mammograms in radiologists," by Schill et al. suggest that the phenomenon of 'missed' lesions on a previous mammogram, subsequently found 'in hindsight' by another radiologist only after diagnosis, is an effect of 'visual hindsight' and not a decision level error. The study employs an object distortion method to provide either an undegraded or a noisy image as a prior, followed by a block of the same image at varying noise levels. When radiologists were first presented with an undegraded version, they subsequently performed significantly better at detecting an abnormality on the noisiest image, compared to blocks starting with a noisy image. Thus, a lesion remained imperceptible until prior information was available. The study supports the phenomenon of visual hindsight – a lesion 'seen' only in hindsight was likely never visually perceptible – with important legal implications.

Park et al. found that two dissociable and temporally separated cognitive processes may account for diagnostic performance. In "Contributions of global and local processing on medical image perception," normal and abnormal mammograms were presented to observers for 500 and 2500 ms duration consecutively. While both experienced and inexperienced observers performed at above-chance at both durations, experienced participants outperformed the latter group at 500 ms, indicating better ability to extract global gist. In addition, on the second exposure of longer duration, only the experienced participants' task performance and confidence ratings increased, indicating local information processing.

2 Technology Evaluation

Krupinski et al. evaluated portable MRI images produced using a deep learning-based advanced reconstruction scheme to improve image quality to determine if diagnostic performance was similar to images acquired at 1.5 T. Their paper "Image quality assessment of advanced reconstruction algorithm for point-of-care MRI scanner" found the scheme was successful for hemorrhage, but for acute ischemic stroke the scheme could still be improved. Portable MRI could be useful for neurocritical care especially in remote and/or resource poor locations.

In "Lesion detection in DBT: Human reader experiments indicate no benefit from the integration of information from multiple planes," Sechopoulos et al. investigated whether the fact that the increased number of images in digital breast tomosynthesis increases reading time, might be mitigated by a perceptual benefit to viewing a mass in the 3D tomosynthesis volumes. Using simulations, 11 readers reviewed 1600 images in two-alternative forced-choice experiments. Spherical lesion detection was higher in 2D than 3D, for both DBT- and bCT-like images, but equivalent for capsule-shaped signals. Average reading time was up to 134% higher for 3D viewing. For low-contrast lesions, there seems to be little inherent visual perception benefit to reviewing the entire DBT or bCT stack.

3 Artificial Intelligence Applications

"*E pluribus unum*: prospective acceptability benchmarking from the Contouring Collaborative for Consensus in Radiation Oncology crowdsourced initiative for multiobserver segmentation"

by Lin et al. sought to characterize whether aggregate segmentations generated from multiple nonexperts could meet or exceed recognized expert agreement. They concluded that nonexperts could potentially generate consensus segmentations for most ROIs with performance approximating experts, suggesting nonexpert segmentations as feasible cost-effective AI inputs.

4 Model Observers

Model observers are always a hot topic at MIPS and Valeri et al. investigated this in "UNet and MobileNet CNN-based model observers for CT protocol optimization: comparative performance evaluation by means of phantom CT images." The goal was to develop and characterize a model observer based on convolutional neural networks, trained to mimic human observers in image evaluation in terms of detection and localization of low-contrast objects in CT scans. They found very good agreement between human and model observers, and between the two algorithms, supporting feasibility of using the approach for CT protocol optimization programs.

In "Assessing saliency models of observers' visual attention on acquired facial differences," Wang et al. evaluated the performance of three models in predicting the saliency of head/neck impairments due to congenital or acquired causes such as cancer. Accurate saliency models could enable new psychosocial interventions to help patients anticipate staring behaviors. While the face-specific saliency model outperformed graph-based and artificial neural network methods, none of the models adequately predicted visual attention compared to eye tracking-based saliency ground truth. The study offers suggestions for improving face-specific models.

5 Training & Expertise

In a bid to make learning more fun for radiology trainees, Banerjee et al. used a gamified workstation to teach skills related to identification of pulmonary nodules in trainees in their study "RADHunters: gamification in radiology perceptual education." The RADHunters game was positively evaluated and although performance of trainees using it increased, so did that of the control group.

Auffermann et al. also sought to improve trainees' abilities to identify pulmonary nodules in "Perceptual training and teaching medical students how to window and level chest radiographs." After a training session, students who were explicitly taught to adjust display settings could localize nodules more accurately and felt significantly more confident in their ability to localize nodules, compared to the control group.

"Taking a second look and zooming out: does this help with abnormality detection in chest radiography?" by Zhu et al. was a training study testing the effectiveness of the "second look and minification technique" (SMLT) aimed are developing a holistic understanding of images used for perceptual training in detection of nodules and other actionable findings in chest radiographs did not find the training significantly increased the accuracy of detection. However strongly positive survey responses lead the authors to argue that SMLT training approach should not be discounted even in the light of this null finding and suggest possible reasons for not finding immediate improved performance after training.

6 Image Characteristics

The Chen and Loew study, "Is it blur or is it texture? An analysis of computational blur and texture measures commonly used in medical image analysis," proposes that the output of a blur measure used for image quality assessment can be influenced by texture information in an image, even though blur measures are usually not evaluated with respect to the presence of texture in the image. The finding that a subset of texture measures outperform established blur measures in blur classification of mammograms suggests that blur measures might not be always a most suitable tool for classifying blur levels in medical image analysis.

A computational approach to understand what constitutes an implicitly perceived symmetry signal between the left and right breast mammograms indicates that the global symmetry signal expert observers are sensitive to is a textural signal embedded in the parenchyma of the mammograms was reported on by Kyle-Davidson et al. in "Signals of global symmetry are important for abnormality detection in mammograms." The findings suggest that the presence of an

abnormality alters the similarity of the textural signal between the two breasts contributing to the medical gist signal that affords experts to detect cancer in fraction of a second even years prior to onset of cancer.

"Visual adaptation to medical images: a comparison of digital mammography and tomosynthesis," by Parthasarathy et al., found that observers could adapt to specific image features – breast density or image noise – to varying degrees on each modality. In particular, the graininess present in tomosynthesis images affected the observers' noise adaptation to a greater degree than mammography. However, observers exhibited little to no adaptation between the two modalities, for as-yet unstudied reasons. This study corroborates and extends previous findings that visual adaptation can increase an observer's sensitivity to features differing from the adapting stimulus, with the potential to improve diagnostic performance.

7 Statistical Methods

Statistical methods have always been a mainstay at MIPS, and in "Roe and Metz identical-test simulation model for validating multi-reader methods of analysis for comparing different radiologic imaging modalities," Hillis discusses simulating multi-reader multi-case data to emulate confidence-of-disease ratings from imaging studies to show how to formulate a Roe and Metz identical-test model and to show its usefulness for validating the error covariance constraints employed by the Obuchowski–Rockette (OR) method. His results show the importance of the OR model constraints for avoiding negative variance estimates using data simulated from the Roe and Metz identical-test model, and that negative variance estimates can occur at nontrivial rates when the two tests are not identical but close to identical.

8 Summary

From the first Far West Image Perception Conference 40 years ago, to the first Medical Imaging Conference 27 years ago, research in medical image perception has continued evolving, from studying old topics in new ways (e.g., hindsight bias, training using simulators, eye-tracking for reporting optimization, image features, statistical analyses) to new application areas (e.g., automated saliency assessment for facial differences, satisfaction of search in virtual mammograms) to newer methodologies and modalities (e.g., neural networks for protocol optimization, crowdsourcing, tomosynthesis, adaptation mechanisms). This special issue reflects the wide variety of topics covered by medical image perception. We hope these papers are of interest and invite readers to join us at our next MIPS XX meeting in 2024 in Nashville!

Acknowledgments

The Medical Image Perception XIX Conference was supported in part by a grant from the National Institutes of Health, Institute of Biomedical Imaging and Bioengineering, 1R13EB031602-01.

Karla Evans is an associate professor and head of the Complex Cognitive Processing Lab at the Psychology Department, University of York, UK. She received her PhD from Princeton University, and performed postdoctoral research at MIT, Harvard Medical School, and the Brigham and Women's Hospital in the US. Using a variety of techniques her work focuses on understanding how perceptual information within and across senses and memory gets integrated into a unified complex percept of the world and is applied to real world tasks. Of a specific interest in the last 12 years is the understanding of visual expertise in medical image perception.

Asli Kumcu is a researcher at TELIN-IPI, an imec research group at Ghent University, Belgium. She received a BS degree in electrical engineering from Purdue University in 2002, MS degree in medical imaging from University of Leuven in 2008, and is currently pursuing her PhD. She was with the healthcare divisions of Philips and Barco for 9 years. Her research interests include optimization and validation of medical imaging devices using human and virtual observers. Her focus lies in assessing image/video quality of diagnostic devices and usability of interactive devices. She has served on the Medical Image Perception Society (MIPS) leadership committee since 2015.

Elizabeth A. Krupinski is a professor and vice-chair for Research at Emory University in the Department of Radiology and Imaging Sciences and a subject matter expert for the Emory Telehealth Program. She received her BA degree from Cornell, MA degree from Montclair State, and PhD from Temple, all in experimental psychology. Her interests are in medical image perception, observer performance, decision making, and human factors. She is a former president of the American Telemedicine Association, former chair of Society for Imaging Informatics in Medicine, former chair of the SPIE Medical Imaging Conference, president of the Society for Education and the Advancement of Research in Connected Health, and president of the Medical Image Perception Society. She is editor of the *Journal of Digital Imaging* and *Telemedicine Reports.*