OXFORD

JNCI Cancer Spectrum (2021) 6(1): pkab099

https://doi.org/10.1093/jncics/pkab099
First published online December 31, 2021
Commentary

Advancing Research on Medical Image Perception by Strengthening Multidisciplinary Collaboration

Melissa Treviño, PhD,[1,2] George Birdsong, MD,[3] Ann Carrigan, PhD,[4] Peter Choyke, MD,[5] Trafton Drew, PhD,[6] Miguel Eckstein, PhD,[7] Anna Fernandez, PhD,[8,9] Brandon D. Gallas, PhD,[10] Maryellen Giger, PhD,[11] Stephen M. Hewitt, MD, PhD,[12] Todd S. Horowitz, PhD,[1,*] Yuhong V. Jiang, PhD,[13] Bonnie Kudrick, MS,[14] Susana Martinez-Conde, PhD,[15] Stephen Mitroff, PhD,[16] Linda Nebeling, PhD, MPH,[1] Joseph Saltz, MD, PhD,[17] Frank Samuelson, PhD,[10] Steven E. Seltzer, MD,[18,19] Behrouz Shabestari, PhD,[20] Lalitha Shankar, MD, PhD,[21] Eliot Siegel, MD,[22] Mike Tilkin,[23] Jennifer S. Trueblood, PhD,[24] Alison L. Van Dyke, MD, PhD,[8] Aradhana M. Venkatesan, MD,[25] David Whitney, PhD,[26] Jeremy M. Wolfe, PhD[19,27,28]

[1]Behavioral Research Program, National Cancer Institute, Rockville, MD, USA, [2]Clinical Research in Complementary and Integrative Health Branch, National Center for Complementary and Integrative Health, Rockville, MD, USA, [3]Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA, USA, [4]Australian Institute of Health Innovation, Macquarie University, Sydney, NSW, Australia, [5]Molecular Imaging Program, National Cancer Institute, Bethesda, MD, USA, [6]Department of Psychology, University of Utah, Salt Lake City, UT, USA, [7]Department of Psychological & Brain Science, University of California, Santa Barbara, CA, USA, [8]Surveillance Research Program, National Cancer Institute, Rockville, MD, USA, [9]Booz Allen Hamilton, McLean, VA, USA, [10]Division of Imaging Diagnostics, and Software Reliability, US Food and Drug Administration, Silver Spring, MD, USA, [11]Department of Radiology, University of Chicago, Chicago, IL, USA, [12]Laboratory of Pathology, National Cancer Institute, Bethesda, MD, USA, [13]Department of Psychology, University of Minnesota, Minneapolis, MN, USA, [14]Transportation Security Administration, Springfield, VA, USA, [15]Department of Ophthalmology, SUNY Downstate Health Sciences University, Brooklyn, NY, USA, [16]Department of Psychology, The George Washington University, Washington, DC, USA, [17]Department of Biomedical Informatics, Stony Brook University, Stony Brook, NY, USA, [18]Department of Radiology, Brigham and Women's Hospital, Boston, MA, USA, [19]Department of Radiology, Harvard Medical School, Boston, MA, USA, [20]Division of Health Informatics Technologies, National Institute of Biomedical Imaging and Bioengineering, Rockville, MD, USA, [21]Cancer Imaging Program, National Cancer Institute, Rockville, MD, USA, [22]Department of Diagnostic Radiology and Nuclear Medicine, University of Maryland School of Medicine, Baltimore, MD, USA, [23]American College of Radiology, Reston, VA, USA, [24]Department of Psychology, Vanderbilt University, Nashville, TN, USA, [25]Department of Abdominal Imaging, University of Texas MD Anderson Cancer Center, Houston, TX, USA, [26]Department of Psychology, University of California, Berkeley, CA, USA, [27]Department of Surgery, Brigham and Women's Hospital, Boston, MA, USA and [28]Department of Ophthalmology, Harvard Medical School, Boston, MA, USA

*Correspondence to: Todd S. Horowitz, PhD, Behavioral Research Program, National Cancer Institute, 9609 Medical Center Dr, Rockville, MD 20850, USA (e-mail: todd.horowitz@nih.gov).

Abstract

Medical image interpretation is central to detecting, diagnosing, and staging cancer and many other disorders. At a time when medical imaging is being transformed by digital technologies and artificial intelligence, understanding the basic perceptual and cognitive processes underlying medical image interpretation is vital for increasing diagnosticians' accuracy and performance, improving patient outcomes, and reducing diagnostician burnout. Medical image perception remains substantially understudied. In September 2019, the National Cancer Institute convened a multidisciplinary panel of radiologists and pathologists together with researchers working in medical image perception and adjacent fields of cognition and perception for the "Cognition and Medical Image Perception Think Tank." The Think Tank's key objectives were to identify critical unsolved problems related to visual perception in pathology and radiology from the perspective of diagnosticians, discuss how these clinically relevant questions could be addressed through cognitive and perception research, identify barriers and solutions for transdisciplinary collaborations, define ways to elevate the profile of cognition and perception research within the medical image community, determine the greatest needs to advance medical image perception, and outline future goals and strategies to evaluate progress. The Think Tank emphasized diagnosticians' perspectives as the crucial starting point for medical image perception research, with diagnosticians describing their interpretation process and identifying perceptual and cognitive problems that arise. This article reports the deliberations of the Think Tank participants to address these objectives and highlight opportunities to expand research on medical image perception.

Received: June 9, 2021; **Revised:** October 20, 2021; **Accepted:** November 11, 2021

Disease detection and diagnosis rely on diagnosticians (ie, anyone tasked with interpreting a medical image, including, but not limited to, radiologists and pathologists) searching for abnormalities in medical images. Across all image acquisition (from light microscopy to magnetic resonance) and presentation methods (eg, via viewing glass slides through a microscope or radiographs at computer workstations), diagnosticians are tasked with combining their vast medical knowledge with contextual information about the patient to interpret a high volume of complex images quickly and accurately. This is a challenging endeavor. Diagnosticians face many well-known difficulties, from time constraints (1) and information overload (2) to repeated interruptions (3) and burnout (4). These difficulties affect health-care quality and patient safety and can inflict financial burdens on practices (5-7).

Even in ideal circumstances, human observers inevitably make mistakes. Errors are a well-documented reality of radiology (8-10) and pathology (11) practice. False positives can lead to unnecessary treatments, and false negatives may mean that disease goes untreated. Although some errors may be due to imaging limitations, diagnosticians sometimes fail to perceive visible abnormalities. These perceptual and cognitive errors (11-13) reflect inherent limitations in human memory, search, interpretation, and decision-making capabilities (14). Medical image perception research can improve clinical care, most notably display characterization for digital imaging guidelines (15), optimizing environments for diagnosticians (16,17), and evaluating the impact of computer-aided detection on observer performance for US Food and Drug Administration approval (18). Understanding the perceptual and cognitive mechanisms underlying diagnostician performance can therefore help to improve diagnostic accuracy (11-13). However, research on perceptual factors has received far less attention and resources than research on new technologies.

To advance research on medical image perception and cognition, the National Cancer Institute (NCI) convened the "Cognition and Medical Image Perception Think Tank" with the goal to advance research on medical image perception and cognition. The Think Tank brought together radiologists, pathologists, researchers working in medical image perception and adjacent fields of cognition and perception, and representatives from interested federal government agencies (including NCI, National Institute of Biomedical Imaging and Bioengineering, US Food and Drug Administration, and US Department of Homeland Security). Here, we report on the Think Tank's deliberations and highlight opportunities to enhance research in medical image perception and cognition.

Note that we use the term "medical image" to cover a wide range of imaging techniques, from images of individual cells to whole-body medical images, from glass slides to 3-dimensional computer renderings. In addition to detection and diagnosis, medical imaging is also used to guide surgical interventions. The term would also cover, for example, a dermatologist visually examining a patient's skin.

Identifying Research Gaps

The first challenge for Think Tank participants was to identify research gaps and critical, unsolved problems in pathology and radiology from the diagnosticians' perspective. Information overload, integrating large amounts of information from various sources, and artificial intelligence (AI) were identified.

Information Overload in the Digital Era

In radiology, the transition from analog to digital produced qualitative changes in image interpretation. A radiologist scrutinizing a single chest radiograph for a lung nodule is typically trying to find a target with a very low signal-to-noise ratio from a single image. The same radiologist reviewing chest computed tomography is more often searching for a target with a fairly high signal-to-noise ratio in a scan comprising hundreds of image slices. In addition, radiologists' decisions now incorporate much more information from nonimaging sources, such as electronic health records and genetic panels. The development of radiomics (19) will allow extraction of nonvisual information from image data, which will have to be integrated into the diagnosis.

Pathologists face similar challenges moving from glass slides to digital images, which changes how the pathologist interacts with the slide. Modern pathology diagnosis goes beyond assessing the presence or absence of diseased tissues and now includes prognostic and predictive elements as well as the integration of ancillary stains and additional modalities such as molecular profiles (20). Pathologists aggregate all these predictive covariates and biomarkers with patient demographics and history to make integrated diagnoses used by clinicians in determining patient management (21). The number of inputs the diagnostician must consider keeps growing, intensifying time pressures.

Integrating Information From Various Sources

Workflow

Diagnosticians move from case to case, organ to organ, and image modality to image modality while dealing with texts, email, pager messages, colleagues, trainees, telephone calls, and other interruptions. Moreover, the task of the diagnostician is to integrate information from diverse sources, whereas current decision-support tools (eg, clinical algorithms, AI-supported findings) are typically narrowly focused.

Fatigue and Workload

As diagnosticians' workload has increased (22), physical and mental fatigue among diagnosticians are at unprecedented levels, contributing to both perceptual and interpretive errors. Multiple studies have reported a decrease in detection accuracy because of fatigue (23). For instance, if a tired radiologist scrolls too quickly through multislice images, the act of displaying all slices on screen can produce the false impression that the case has been adequately scrutinized. Errors can also result from lack of focus, distractions, satisfaction of search (when detection of 1 abnormality interferes with detecting others), and/or the failure to record what is perceived. Errors caused by physical and mental fatigue become more apparent at the end of the day (24).

Artificial Intelligence

AI tools are routinely proposed as the solution to many challenges facing diagnosticians. However, there is a mismatch between what diagnosticians need from AI and how AI is typically developed. In news reports, AI solutions are often pitched as "outperforming" the diagnostician, implying that AI systems are on the cusp of replacing the diagnostician (25) or at least serving as an equivalent reader (26). This is a flawed analysis. First, as with prior computer-aided detection systems (27), an AI algorithm that performs well in a laboratory setting may not perform as well once deployed in real-world clinical practice. Diagnosticians may come to over-rely on computer recommendations (28,29) or, conversely, learn to ignore them (30). Second, training biases (31) and "off label" use of AI for tasks it was not designed for (eg, using AI as a primary reader when it is intended as a second reader) may cause interpretation errors. Finally, human diagnosticians are more than collections of image-interpretation routines. They perform noninterpretative functions, such as interacting with patient-facing clinicians, dealing with patient responses, and managing quality control and other administrative functions, none of which AI can yet perform.

A better approach is to consider AI as a trained assistant, making the diagnostician's work easier by automating aspects of the data reduction process and presenting information that can aid the decision process. This "trained assistant" role has already gained clinical utility in some areas, such as interpreting Pap smears (32), identifying focal lesions on screening mammograms (as in computer-aided detection) (33), and looking for errors in prescriptions (34). Such tools are intended to provide optimal presentation of findings, ensure complete and precise evaluation of the information, improve accuracy, and reduce cost. This model leaves the task of integration to the diagnostician.

An AI "Diagnostic Cockpit of the Future"

A metaphor for how perception and cognition research can guide the integration of AI tools is the "diagnostic cockpit of the future" (35). During World War II, thousands of B-17 bombers crashed on landing without any sign of mechanical failure. Psychologists Paul Fitts and Alphonse Chapanis noticed that the controls for the landing gear and the flaps were identical, so that pilots were often mistaking one for the other. By changing the shape of the controls, they were able to avert many crashes (36). Similar human factor input could create a diagnostic cockpit that reduces medical error. This cockpit would include a set of tools to "aggregate, organize, and simplify medical imaging results and patient-centric clinical data" (35, p. 579). The aggregation of multiparametric data is typically not a strength of human observers but is well suited to computer applications.

Questions and Opportunities for Medical Image Perception Research

What are the effects of information overload on medical decision making and decision quality? Are there technical solutions, decision-making aids, or strategies that could be used to more efficiently integrate complex information? Can such decision-making strategies capture the ability of the patient's clinical history to provide a useful set of differential diagnoses?

More research is needed on documenting and mitigating the costs of task switching. For example, can system-level dashboards that integrate tools minimize the burden of switching? What interventions might be available to detect and alleviate the effects of fatigue on interpretive accuracy?

Pathology has been understudied, relative to radiology, presenting many research opportunities. Although in some cases pathologists make binary assessments (ie, cancer vs not cancer), in most cases their assessments require more nuanced grading or subtyping. Because treatment decisions frequently hinge on tumor classification, interobserver variability among pathologists is an important problem. How do these classification errors change as pathology transitions to using digital images?

We know even less about perceptual and cognitive factors in fields such as dermatology (37,38) and ophthalmology. Meanwhile, the COVID-19 pandemic has highlighted the importance of telemedicine (39), which may pose novel cognitive challenges.

Addressing Clinical Questions Through Perceptual Research

Is medical image perception research on nonexpert observers clinically relevant? The relevance argument assumes that diagnosticians and nondiagnosticians have the same visual, perceptual, and cognitive faculties. Thus, the perceptual and cognitive bottlenecks found in nonexpert observers will also apply to diagnosticians. Expertise changes the use of these faculties but does not create a new visual system.

Clinical relevance also requires that artificial tasks and stimuli used in these studies capture at least some key elements of their clinical counterparts. Perceptual studies often trade the specificity and messiness of clinical settings for experimental control. Complex stimuli are simplified, narrative reports replaced with binary responses, clinical history is absent, and the life-and-death stakes of the clinic are eliminated. Many diagnosticians are skeptical of such abstracted medical image interpretation. However, such methodological choices are required to make progress on some topics because parametric studies with experts are often impractical. Expert participants need to be reserved for relatively short, carefully crafted experiments.

These constraints motivate a "reverse translation" paradigm (see Figure 1). The process begins with a dialogue between diagnostician and researcher (and diagnosticians are often researchers!), where the diagnostician identifies a problem in the clinical setting (just as the diagnosticians at the Think Tank did in the section on identifying research gaps). The researcher then abstracts out a basic question about the underlying cognitive and perceptual capabilities required and develops laboratory versions of the problem to study the basic science ("use-inspired basic research") (40,41). These studies may use observer models or nonexpert observers because diagnosticians' time is severely limited (42). After multiple iterations, working out the basic science, the researcher can return to the clinical setting to test the resulting hypothesis with a plausible, well-focused experiment with clinician observers.

We stress that this reverse translation strategy can succeed only if researchers study questions that diagnosticians want answered. We propose that closer collaboration between diagnosticians and basic researchers is critical to ensuring that perception research investigates problems that are clinically significant (see Challenges and Solutions for Transdisciplinary Research below).

Case Study: Prevalence

Research on the prevalence effect provides an illustrative example: 1) Low prevalence has been identified as a problem for diagnosticians (43). 2) Perception researchers create *visual search* experiments as an abstract approximation of medical image interpretation. For example, observers might search for semitransparent images of tools (*targets*) among nontool objects embedded in noisy backgrounds (44). 3) By using artificial search tasks and nonexpert observers, researchers were able to vary prevalence from 0.078% (45) to 98% (46), measure the development of prevalence expectations (47), and more. 4) Multiple studies converged on the hypothesis that observers become less

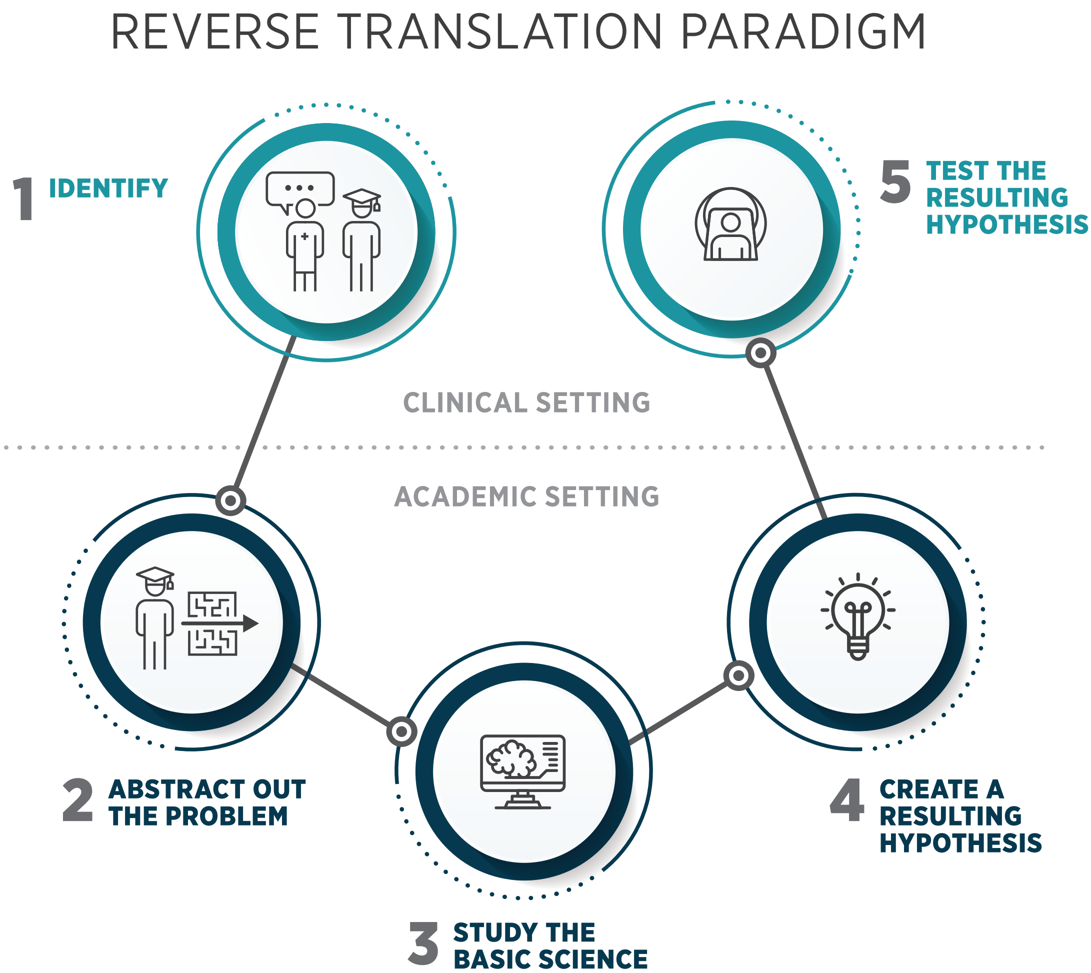

Figure 1. The reverse translation paradigm. Diagnosticians and researchers work together to identify relevant problems in the clinical setting (1). Perception researchers abstract the problem to create a laboratory version of the problem (2) and conduct experiments to study the basic science involved (3). From these studies, researchers derive a hypothesis (4), which can be tested in the clinical setting (5).

likely to call out targets (becoming more "conservative" in signal detection terms) as prevalence decreases. Consequently, observers are more likely to miss targets in lower prevalence than in higher target prevalence scenarios (48). 5) Returning to the clinical setting to test the hypothesis, this *criterion shift* has been shown to occur in both mammography (49) and cytopathology (50). Prevalence effects have wide implications. However, these studies would be implausibly long for diagnostician observers. Even if one curated thousands of medical images for a study at 2% prevalence, using alternative study designs to reduce the burden to study participants (51), recruiting multiple radiologists to read all those images would be difficult. The basic science of the interventions must be established in the laboratory before replications can be attempted in clinical settings.

Challenges and Solutions for Transdisciplinary Research

Today, multiple barriers hinder collaborations between perception researchers and diagnosticians. Medical image perception is an inherently transdisciplinary enterprise, requiring collaborations between diagnosticians, medical physicists, computer scientists, biomedical informaticians, engineers, and psychologists. Historically, MDs such as William Tuddenham and Harold Kundel (52,53) championed applying principles from psychology and other disciplines to advance medical image research (54). These collaborations were easier to initiate in an era when more perceptual psychologists were embedded in radiology departments with diagnosticians and physicists. These psychophysical researchers could become intimately familiar with clinical issues and more easily find collaborators. Together, physicists and psychologists exposed diagnosticians to medical image perception research and attracted radiologists to the subfield. Unfortunately, coexistence of multidisciplinary researchers within medical centers is less common now. Nevertheless, perceptual and cognitive challenges in medicine still require collaborations between diagnosticians and perception researchers, including cognitive psychologists, vision and oculomotor scientists, cognitive neuroscientists, and so forth.

Transdisciplinary teams face many challenges (55). The Think Tank identified specific challenges for medical image perception collaborations and research and proposed several solutions to address these challenges. A substantial challenge for both collaboration and participation in reader studies is competition for diagnosticians' limited time. Collaborations should be designed to respect the diagnostician's schedule, with projects developed around common interests of both parties. Incentives for diagnosticians' participation should include grant funding, institutional support (eg, career path recognition), and authorship on publications. Additional incentives should include providing continuing medical education (CME) credit and/or CME in risk management for participating in cognitive research on medical image perception and decision making. To facilitate CME credit, an educational module should be created for participants, with objectives such as education on fundamental perceptual or cognitive concepts relating to clinical practice (eg, human vision, attentional mechanisms, sources of error, cognitive heuristics). Such CME models would also require that researchers share their results with participants on completion of the study. Specialty organizations like the American College of Radiology and the College of American Pathologists could enable collaborations by incorporating perception research into their quality assurance programs. The colleges could build up a "library of errors," images that tend to lead to incorrect responses, to be used both as a teaching tool and resource for perception researchers.

Fundamental knowledge of relevant concepts in cognition and perception should be incorporated into diagnosticians' training to foster the value of collaborations. These concepts are included in the noninterpretive skills section of the American Board of Radiology Core and Certifying examination, although there is typically little formal instruction on cognition and perception during residency and fellowship. CME credit, discussed above, should be tied to this content. It would also be helpful to revive the practice of providing training in the basics of medical image perception for panelists on relevant National Institutes of Health study sections. Joint funding opportunities with NCI, National Institute of Biomedical Imaging and Bioengineering, and the Agency for Healthcare Research and Quality could support collaborative research by including protected time for diagnosticians committed to such research and clinical mentoring for basic research scientists.

Academic researchers, especially those not based in a medical setting, have limited access to diagnosticians, who tend to be extremely busy. As a solution, "pop-up" perception laboratories at diagnostician conferences have been successful in recent years (56). A perception laboratory at a medical conference allows researchers access to a larger population of experts at a time when those experts are not occupied with their clinical work. Sustaining and strengthening these efforts will require partnering with national and regional medical conferences (eg, College of American Pathologists, American College of Radiology, American Society for Clinical Pathology, American Society of Cytopathology, Radiological Society of North America [RSNA], Society of Nuclear Medicine and Molecular Imaging, Association for Medical Ultrasound). Industry could collaborate with perception researchers to create a model "reading room of the future" at conferences and invite diagnosticians to different workstations that promote research outcomes. The next frontier will be platforms that allow medical image perception research to occur online. The development of such platforms will benefit from targeted funding and industry collaborations.

Beyond the recruitment issue, university medical centers and NCI-designated cancer centers should be encouraged to make medical image perception research a priority and to financially sustain perception laboratories, perhaps through NIH infrastructure grants. Collaboration often develops serendipitously, and serendipity depends on proximity. For perception researchers not embedded within medical centers, opportunities to network or promote research through invited talks are limited. One solution would be to create a matchmaking resource to identify perception scientists, diagnosticians, conferences, private practices, and medical departments who are open to collaborative research.

In response to Think Tank deliberations, NCI created a networking and collaboration website for medical image perception research (https://ncihub.org/groups/medicalimageperception/overview). Members of the site can create a profile, listing their name, affiliation, email, and a bio detailing their area of expertise or interest and a proposed project or potential research contribution, such as participating in reader studies, inviting researchers to give talks at their institutions, providing image datasets, generating research ideas and providing feedback, or becoming investigators.

Collaboration is more likely if diagnosticians know about perception research. Although the primary purpose of a pop-up laboratory, such as the NCI Perception Lab at RSNA, is to recruit research participants, these endeavors can also expose diagnosticians to perceptual research and help spark conversations between researchers and diagnosticians. Clinical research projects could be encouraged, perhaps by grant supplements, to include perception research as a secondary aim. Perceptual researchers should be encouraged to include outreach to clinical audiences in their dissemination plans.

Accessing medical image datasets can be a challenge for perception researchers. Curating image datasets is often time-consuming for diagnosticians and requires financial resources to undertake and sustain. Sharing images across institutions is difficult because of privacy and confidentiality risks and liability issues. Even image sets from large government-funded trials can be challenging to access and may need additional annotation for perception research. One way to address these problems would be to develop standardized templates for research collaboration and sharing agreements. Funding agencies could promote the use of such templates, much as they now encourage PIs to make datasets publicly available.

Existing image repositories (eg, NCI's The Cancer Imaging Archive or the Alzheimer's Disease Neuroimaging Initiative) may reduce the difficulties associated with accessing clinical trial image sets, but they often lack needed clinical context, appropriate annotations, diversity of pathology, and an efficient interface for navigating and searching for relevant images. The NCI PRISM project (57) is creating an infrastructure to integrate radiology and pathology images, which will include imaging features assessed by humans and by algorithms along with clinical context. Obtaining human annotations can be a challenge because of the effort and coordination required to carry this out. To address these challenges, NCI's Cancer Imaging Program is funding the hosting of annotated diagnostic imaging from phase II and III clinical trials conducted by NCI networks on The Cancer Imaging Archive. The Crowds Cure Cancer project recruits radiologists attending RSNA to identify lesions and lesion characteristics and to provide details regarding image quality. This professional crowdsourcing approach is a model for annotating other medical image repositories. Involving perception scientists in project design could help to ensure that variables important to perception studies are coded as well. As AI research faces similar issues around annotation, this might be a good area for collaboration between perception science, AI, and medical imaging (58).

Within the cognitive science research community, there has long been an interest in medical decision making. However, there has been only modest collaboration between basic researchers with expertise in areas such as human judgment and decision making and researchers with interests in medical image perception. Diagnosticians are often confronted with classic decision-making problems (eg, integrating abundant and various forms of information), which have been extensively researched but not in the context of medical imaging. Closer interaction with judgment and decision research communities could offer insight to further advance medical imaging research.

Symposiums and workshops should be held at cognitive science conferences to engage wider audiences (ie, Society for Mathematical Psychology, the Society of Medical Decision Making, Psychonomic Society, Vision Sciences Society). As a model, consider the outreach efforts conducted by the International Society for Optics and Photonics to educate researchers on the diagnostician's perspective and challenges. This effort includes demonstrations of diagnosticians interpreting case s, giving researchers insight into how diagnosticians read images.

Strong connections already exist between perception, cognitive science, and computer science. Computational vision is a robust research field that substantially overlaps with perception research and computer science. As AI solutions to imaging problems proliferate, collaborations between clinicians, perceptual, decision, and computer scientists should be encouraged.

Elevating the Profile of Medical Image Perception and Cognition Research

Creating a network of invested stakeholders (eg, diagnosticians, academics, biomedical informaticians, government and regulatory agencies, medical centers, industry, and conferences) is critical to support an awareness program promoting education, collaboration, and research. Here we identify key essential elements for achieving this goal.

Diagnostician advocates are needed to lend visibility and credibility and to reinforce the value of perception research to the medical community. This is important not just to facilitate the research itself but also to encourage adoption of research-based solutions to diagnostician challenges. One model for this network is the Alliance for Digital Pathology, a regulatory science initiative that brings together diagnosticians, researchers, government, and industry to identify key aims to advance digital pathology and provide an infrastructure for collaborative projects (59).

Medical image perception research would benefit from a more robust presence in diagnostician-focused journals. Although there are specialty journals that routinely publish medical image perception research, publication in high-impact radiology and pathology journals is uneven. Some of this may be the result of a lack of editorial infrastructure (eg, the absence of associate editors with expertise in the area makes it harder to obtain reliable peer review of submissions). The basic perceptual and cognitive literature would be enriched by review and opinion pieces by diagnosticians describing the perceptual and cognitive questions that arise in clinical work. Similarly, it would be helpful to have reviews of the relevant perceptual and cognitive literature in widely read clinical journals.

Future Goals and Strategies to Evaluate Progress

The overall goal of advancing medical image perception research is to improve patient care by reducing diagnostic errors, improving machine interfaces, and mitigating diagnosticians' challenges. Research should yield best practices that can be incorporated into every aspect of medical image perception, including training, device and algorithm development, and clinical practice. How will we know if this initiative is bearing fruit?

Outreach can be tracked via the number of funding announcements, number of new principal investigators recruited into the field, and grants with transdisciplinary teams including radiologists or pathologists. Match-making and new collaborations can be tracked, in part, via the website mentioned earlier. Medical image perception research outcomes can be tracked in industry modifications to workstations or AI, reductions in errors, reductions in the time diagnosticians spend per case, and improvement in diagnostician attrition. Ultimately, these outcomes should be modeled to measure the overall improvement in potential lives saved and/or quality of adjusted life years gained. The baseline trajectory for the field should begin with 2014, the year that NCI's Vision Science Problems in Medical Imaging Workshop was held. We can then compare progress to a counterfactual trajectory, extrapolated from prior trends, and assess what is needed to rectify and/or further improve the trajectories (see Figure 2).

Summary of Think Tank Deliberations

NCI's "Cognition and Medical Image Perception Thank Tank" brought together diagnosticians, perception researchers, and representatives from government agencies and the medical imaging community to deliberate and identify courses of action to advance medical image perception research. Attendees pinpointed key challenges facing diagnosticians that could be addressed by appropriately designed perception research. Challenges include information overload, data integration from multiple sources, and AI. Advancing medical image perception research requires active multidisciplinary participation from diagnosticians, perception researchers, and the medical imaging community. Together, these fields can improve patient care by addressing the many diverse challenges facing diagnosticians.

Funding

Funds from the National Cancer Institute were used to convene experts for the Think Tank that informed the manuscript's scientific scope.

Notes

Role of the funder: National Cancer Institute (NCI) staff had roles in the design of the Think Tank, contributed to the writing of the manuscript and the decision to submit the manuscript for publication. Moreover, the manuscript was reviewed and approved for submission through the manuscript clearance processes of both the NCI Division of Cancer Control and Population Sciences and the National Center for Complementary and Integrative Health.

Acknowledgements: We would like to thank Paige Green for her leadership and support, and Elizabeth Krupinski for comments on the manuscript.

Disclosures: MT: None. GB: Has received payment from genae, a contract research organization engaged for work on a project

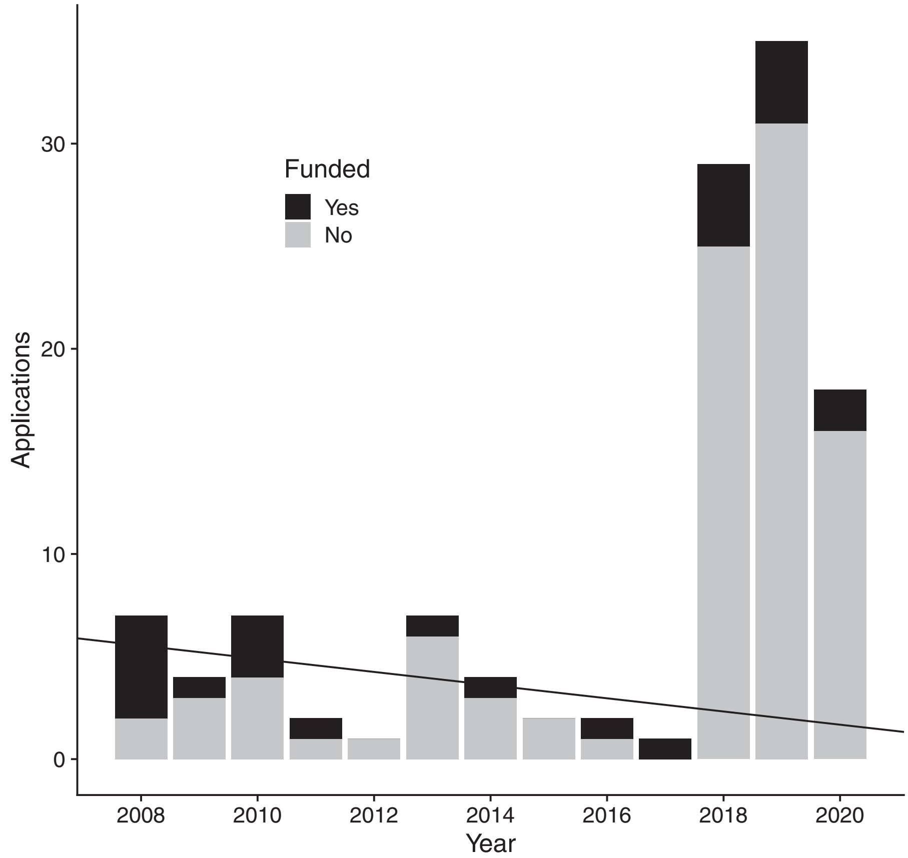

Figure 2. Funding trends for medical image perception and cognition research at the National Institutes of Health, 2008-2020. **Black lines** denotes the linear trend based on 2008-2014 (ie, before the Vision Science Problems in Medical Imaging Workshop was held), extrapolated to 2020.

sponsored by the Hammanatsu Corporation. AC: None. PC: None. TD: None. ME: None. AF: None. BDG: None. MG: Multiple patents over the past 30 years that were licensed as a bundle through the University of Chicago Technology Transfer offices. Includes royalties from Hologic, GE. Scientific advisor to Qlarity Imaging. SMH: None. TSH: None. YVJ: None. BK: None. SMC: None. SM: None. LN: None. JS: Patent title: "System and Method To Quantify Tumor-Infiltrating Lymphocytes (TILs) For Clinical Pathology." FS: None. SES: On the Board of Review for the GE-AUR Radiology Research Fellowship Program and received travel reimbursements to attend 1 Board and Mentoring meeting per year. BS: None. LS: None. ES: None. MT: None. JST: National Science Foundation paid to Vanderbilt University. ALVD: None. AMV: 1. Grant support from Siemens Healthineers and from the UT MD Anderson Cancer Center-Siemens Alliance. 2. Grant support from the Institutional Research Grant Program of the University of Texas MD Anderson Cancer Center (UTMDACC). 3. Grant support from the UTMDACC Cancer Center Support Grant (CCSG) Radiation Oncology and Cancer Imaging Program. 4. Grant support from the UTMDACC Radiation Oncology Strategic Initiatives (ROSI) Program. DW: None. JMW: None.

Author contributions: MT: Conceptualization, Visualization, Writing—original draft, Writing—review and editing. GB: Writing—review and editing. AC: Writing—review and editing. PC: Writing—review and editing. TD: Writing—review and editing. ME: Writing—review and editing. AF: Writing—review and editing. BDG: Writing—review and editing. MG: Writing—review and editing. SMH: Writing—review and editing. TSH: Conceptualization, Funding acquisition, Project administration, Visualization, Writing—review and editing. YVJ: Writing—review and editing. BK: Writing—review and editing. SMC: Writing—review and editing. SM: Writing—review and editing. LN: Writing—review and editing. JS: Writing—review and editing. FS: Writing—review and editing. SES: Writing—review and editing. BS: Writing—review and editing. LS: Writing—review and editing. ES: Writing—review and editing. MT: Writing—review and editing. JST: Writing—review and editing. ALVD: Writing—review and editing. AMV: Writing—review and editing. DW: Writing—review and editing. JMW: Writing—review and editing.

Data Availability

Data and code used to generate Figure 2 are available via the Open Science Framework, at https://osf.io/cpzyk.

References

1. ALQahtani DA, Rotgans JI, Mamede S, Mahzari MM, Al-Ghamdi GA, Schmidt HG. Factors underlying suboptimal diagnostic performance in physicians under time pressure. *Med Educ*. 2018;52(12):1288–1298. doi:10.1111/medu.13686.
2. Gunderman RB. Information overload. *J Am Coll Radiol*. 2006;3(7):495–497. doi: 10.1016/j.jacr.2006.03.021.
3. Ratwani RM, Wang E, Fong A, Cooper CJ. A human factors approach to understanding the types and sources of interruptions in radiology reading rooms. *J Am Coll Radiol*. 2016;13(9):1102–1105. doi:10.1016/j.jacr.2016.02.017.
4. Shanafelt TD, West CP, Sinsky C, et al. Changes in burnout and satisfaction with work-life integration in physicians and the general US Working Population between 2011 and 2017. *Mayo Clin Proc*. 2019;94(9):1681–1694. doi: 10.1016/j.mayocp.2018.10.023.

5. Panagioti M, Geraghty K, Johnson J, et al. Association between physician burnout and patient safety, professionalism, and patient satisfaction. *JAMA Intern Med*. 2018;178(10):1317–1331. doi:10.1001/jamainternmed.2018.3713.
6. Shanafelt T, Goh J, Sinsky C. The business case for investing in physician well-being. *JAMA Intern Med*. 2017;177(12):1826–1832. doi:10.1001/jamaintern med.2017.4340.
7. Wallace JE, Lemaire JB, Ghali WA. Physician wellness: a missing quality indicator. *The Lancet*. 2009;374(9702):1714–1721. doi:10.1016/s0140-6736(09)61424-0.
8. Berlin L. Accuracy of diagnostic procedures: has it improved over the past five decades? *AJR Am J Roentgenol*. 2007;188(5):1173–1178. doi:10.2214/ajr.06.1270.
9. Haider MA, Yao X, Loblaw A, Finelli A. Multiparametric magnetic resonance imaging in the diagnosis of prostate cancer: a systematic review. *Clin Oncol (R Coll Radiol)*. 2016;28(9):550–567. doi:10.1016/j.clon.2016.05.003.
10. Sivaraman A, Sanchez-Salas R, Ahmed HU, et al. Clinical utility of transperineal template-guided mapping biopsy of the prostate after negative magnetic resonance imaging–guided transrectal biopsy. *Urol Oncol*. 2015;33(7):329.e7–329.e11. doi:10.1016/j.urolonc.2015.04.005.
11. Raab SS, Grzybicki DM, Janosky JE, et al. Clinical impact and frequency of anatomic pathology errors in cancer diagnoses. *Cancer*. 2005;104(10):2205–2213. doi:10.1002/cncr.21431.
12. Singh H, Sethi S, Raber M, Petersen LA. Errors in cancer diagnosis: current understanding and future directions. *J Clin Oncol*. 2007;25(31):5009–5018. doi:10.1200/jco.2007.13.2142.
13. Smith MJ. *Error and Variation in Diagnostic Radiology*. Springfield, IL: Charles. C. Thomas; 1967.
14. Andriole KP, Wolfe JM, Khorasani R, et al. Optimizing analysis, visualization, and navigation of large image data sets: one 5000-section CT scan can ruin your whole day. *Radiology*. 2011;259(2):346–362. doi:10.1148/radiol.11091276.
15. Andriole KP, Ruckdeschel TG, Flynn MJ, et al. ACR–AAPM–SIIM practice guideline for digital radiography. *J Digit Imaging*. 2013;26(1):26–37. doi:10.1007/s10278-012-9523-1.
16. Goo JM, Choi J-Y, Im J-G, et al. Effect of monitor luminance and ambient light on observer performance in soft-copy reading of digital chest radiographs. *Radiology*. 2004;232(3):762–766. doi:10.1148/radiol.2323030628.
17. Krupinski EA, Johnson J, Roehrig H, Nafziger J, Lubin J. On-axis and off-axis viewing of images on CRT displays and LCDs. *Acad Radiol*. 2005;12(8):957–964. doi:10.1016/j.acra.2005.04.015.
18. Gallas BD, Chan H-P, D'Orsi CJ, et al. Evaluating imaging and computer-aided detection and diagnosis devices at the FDA. *Acad Radiol*. 2012;19(4):463–477. doi:10.1016/j.acra.2011.12.016.
19. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging—"how-to" guide and critical reflection. *Insights Imaging*. 2020;11(1):91. doi:10.1186/s13244-020-00887-2.
20. Han HS, Magliocco AM. Molecular testing and the pathologist's role in clinical trials of breast cancer. *Clin Breast Cancer*. 2016;16(3):166–179. doi:10.1016/j.clbc.2016.02.016.
21. Lippi G, Plebani M. Integrated diagnostics. *Biochem Med (Zagreb)*. 2020;30(1):010501–010530. doi:10.11613/bm.2020.010501.
22. Bhargavan M, Kaye AH, Forman HP, Sunshine JH. Workload of radiologists in United States in 2006-2007 and trends since 1991-1992. *Radiology*. 2009;252(2):458–467. doi:10.1148/radiol.2522081895.
23. Waite S, Kolla S, Jeudy J, et al. Tired in the reading room: the influence of fatigue in radiology. *J Am Coll Radiol*. 2017;14(2):191–197. doi:10.1016/j.jacr.2016.10.009.
24. Krupinski EA, Berbaum KS, Caldwell RT, Schartz KM, Kim J. Long radiology workdays reduce detection and accommodation accuracy. *J Am Coll Radiol*. 2010;7(9):698–704. doi:10.1016/j.jacr.2010.03.004.
25. Nast C. A.I. versus M.D. *The New Yorker*. https://www.newyorker.com/magazine/2017/04/03/ai-versus-md. 2017. Accessed September 22, 2021.
26. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature*. 2020;577(7788):89–94. doi:10.1038/s41586-019-1799-6.
27. Lehman CD, Wellman RD, Buist DSM, Kerlikowske K, Tosteson ANA, Miglioretti DL; Breast Cancer Surveillance Consortium. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Intern Med*. 2015;175(11):1828–1837. doi:10.1001/jamainternmed.2015.5231.
28. Taplin SH, Rutter CM, Lehman CD. Testing the effect of computer-assisted detection on interpretive performance in screening mammography. *Am J Roentgenol*. 2006;187(6):1475–1482. doi:10.2214/ajr.05.0940.
29. Drew T, Cunningham C, Wolfe JM. When and why might a computer-aided detection (CAD) system interfere with visual search? An eye-tracking study. *Acad Radiol*. 2012;19(10):1260–1267. doi:10.1016/j.acra.2012.05.013.
30. Nishikawa RM, Schmidt RA, Linver MN, Edwards AV, Papaioannou J, Stull MA. Clinically missed cancer: how effectively can radiologists use computer-aided detection? *Am J Roentgenol*. 2012;198(3):708–716. doi:10.2214/ajr.11.6423.
31. Char DS, Shah NH, Magnus D. Implementing machine learning in health care — addressing ethical challenges. *N Engl J Med*. 2018;378(11):981–983. doi:10.1056/nejmp1714229.
32. Hu L, Bell D, Antani S, et al. An observational study of deep learning and automated evaluation of cervical images for cancer screening. *J Natl Cancer Inst*. 2019;111(9):923–932. doi:10.1093/jnci/djy225.
33. Gromet M. Comparison of computer-aided detection to double reading of screening mammograms: review of 231,221 mammograms. *Am J Roentgenol*. 2008;190(4):854–859. doi:10.2214/ajr.07.2812.
34. Rozenblum R, Rodriguez-Monguio R, Volk LA, et al. Using a machine learning system to identify and prevent medication prescribing errors: a clinical and cost analysis evaluation. *Jt Comm J Qual Patient Saf*. 2020;46(1):3–10. doi:10.1016/j.jcjq.2019.09.008.
35. Krupinski E, Bronkalla M, Folio L, et al. Advancing the diagnostic cockpit of the future: an opportunity to improve diagnostic accuracy and efficiency. *Acad Radiol*. 2019;26(4):579–581. doi:10.1016/j.acra.2018.11.017.
36. Kuang C. How the dumb design of a WWII plane led to the Macintosh. *Wired*. https://www.wired.com/story/how-dumb-design-wwii-plane-led-macintosh/. Accessed September 22, 2021.
37. Gachon J, Beaulieu P, Sei JF, et al. First prospective study of the recognition process of melanoma in dermatological practice. *Arch Dermatol*. 2005;141(4):434–438. doi:10.1001/archderm.141.4.434.
38. Xu B, Rourke L, Robinson JK, Tanaka JW. Training melanoma detection in photographs using the perceptual expertise training approach. *Appl Cognit Psychol*. 2016;30(5):750–756. doi:10.1002/acp.3250.
39. Mermelstein H, Guzman E, Rabinowitz T, Krupinski E, Hilty D. The application of technology to health: the evolution of telephone to telemedicine and telepsychiatry: a historical review and look at human factors. *J Technol Behav Sci*. 2017;2(1):5–20. doi:10.1007/s41347-017-0010-x.
40. Wolfe JM. Use-inspired basic research in medical image perception. *Cogn Res Princ Implic*. 2016;1(1):17. doi:10.1186/s41235-016-0019-2.
41. Stokes DR. *Pasteur's Quadrant: Basic Science and Technological Innovation*. Washington, DC: Brookings Institution Press; 1997.
42. Krupinski EA. The important role of task-based model observers and related techniques in medical imaging. *J Nucl Cardiol*. 2021;28(2):638–640. doi:10.1007/s12350-019-01769-x.
43. Whiting P, Rutjes AWS, Reitsma JB, Glas AS, Bossuyt PMM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy. *Ann Intern Med*. 2004;140(3):189–202. doi:10.7326/0003-4819-140-3-200402030-00010.
44. Wolfe JM, Horowitz TS, Van Wert MJ, Kenner NM, Place SS, Kibbi N. Low target prevalence is a stubborn source of errors in visual search tasks. *J Exp Psychol Gen*. 2007;136(4):623–638. doi:10.1037/0096-3445.136.4.623.
45. Mitroff SR, Biggs AT. The ultra-rare-item effect. *Psychol Sci*. 2014;25(1):284–289. doi:10.1177/0956797613504221.
46. Wolfe JM, Van Wert MJ. Varying target prevalence reveals two dissociable decision criteria in visual search. *Curr Biol*. 2010;20(2):121–124. doi:10.1016/j.cub.2009.11.066.
47. Ishibashi K, Kita S, Wolfe JM. The effects of local prevalence and explicit expectations on search termination times. *Atten Percept Psychophys*. 2012;74(1):115–123. doi:10.3758/s13414-011-0225-4.
48. Horowitz TS. Prevalence in visual search: from the clinic to the lab and back again. *Jpn Psychol Res*. 2017;59(2):65–108. doi:10.1111/jpr.12153.
49. Evans KK, Birdwell RL, Wolfe JM. If you don't find it often, you often don't find it: why some cancers are missed in breast cancer screening. *PLoS One*. 2013;8(5):e64366. doi:10.1371/journal.pone.0064366.
50. Evans KK, Tambouret RH, Evered A, Wilbur DC, Wolfe JM. Prevalence of abnormalities influences cytologists' error rates in screening for cervical cancer. *Arch Pathol Lab Med*. 2011;135(12):1557–1560. doi:10.5858/arpa.2010-0739-oa.
51. Gallas BD, Chen W, Cole E, et al. Impact of prevalence and case distribution in lab-based diagnostic imaging studies. *J Med Imaging (Bellingham)*. 2019;6(1):015501. doi:10.1117/1.jmi.6.1.015501.
52. Kundel HL. History of research in medical image perception. *J Am Coll Radiol*. 2006;3(6):402–408. doi:10.1016/j.jacr.2006.02.023.
53. Tuddenham WJ. Visual search, image organization, and reader error in roentgen diagnosis. *Radiology*. 1962;78(5):694–704. doi:10.1148/78.5.694.
54. Tuddenham WJ. Roentgen image perception–a personal survey of the problem. *Radiol Clin North Am*. 1969;7(3):499–501.
55. Hall KL, Vogel AL, Huang GC, et al. The science of team science: a review of the empirical evidence and research gaps on collaboration in science. *Am Psychol*. 2018;73(4):532–548. doi:10.1037/amp0000319.
56. Toomey RJ, McEntee MF, Rainford LA. The pop-up research centre – challenges and opportunities. *Radiography*. 2019;25:S19–S24. doi:10.1016/j.radi.2019.05.009.
57. Sharma A, Tarbox L, Kurc T, et al. PRISM: a platform for imaging in precision medicine. *J Clin Oncol Clin Cancer Inform*. 2020;4:491–499. doi:10.1200/cci.20.00001.
58. Dudgeon SN, Wen S, Hanna MG, et al. A pathologist-annotated dataset for validating artificial intelligence: a project description and pilot study. arXiv; 2020;12:45. https://arxiv.org/abs/2010.06995. Accessed January 26, 2021.
59. Marble H, Huang R, Dudgeon S, et al. A regulatory science initiative to harmonize and standardize digital pathology and machine learning processes to speed up clinical innovation to patients. *J Pathol Inform*. 2020;11(1):22. doi:10.4103/jpi.jpi_27_20.