Translating computational innovations into reality: focus on the users!

Elizabeth A. Krupinski*[a]
[a]Emory University Department of Radiology & Imaging Sciences,
1364 Clifton Rd NE, Atlanta, GA, USA 30322

ABSTRACT

Computer-aided/AI-driven tools are increasing being developed for use with digital pathology images. Whether a given scheme makes it into clinical use depends on a multitude of factors, perhaps most importantly whether it has an impact on a clinician's decision-making process thus on patient care and outcomes. To have a positive impact, clinical decision support tools must be well-integrated into routine clinical workflows and thus require assessment from a human factors perspective that includes attention to ways these tools impact users' perceptual and cognitive information processing mechanisms. Methods from implementation sciences (the scientific study of methods to promote systematic uptake of research findings and other evidence-based practices into routine practice to improve the quality and effectiveness of health services) can also be used to prepare users for and then assess the impact of introducing these tools into the clinical workflow. This paper will provide an overview of these perspectives, drawing on the history of medical image perception research in radiology and the growing application of these principles and methods in pathology.

Keywords: image perception, human factors, implementation science, observer performance

1. INTRODUCTION

Medical imaging is at an important crossroad. Computer-based (artificial intelligence (AI), deep learning (DL), machine learning (ML)) tools are increasing being developed for use with practically every type of medical data, especially images (e.g., radiology, pathology, dermatology, ophthalmology). The tasks these computer-based analysis tools are designed to do vary considerably and include but are not limited to lesion detection and classification, organ/feature segmentation, outcomes and survival prediction, feature/lesion measurement (size, count, volume), report generation and analysis, image quality enhancement, more efficient image acquisition, and workflow analysis. The goals vary as well but ultimately focus on improving healthcare outcomes through improved data analytics that the healthcare system (e.g., providers, technologists, physicists, schedulers, administrators) can utilize to improve the efficiency and efficacy of diagnoses, treatments and outcomes. Currently however, there are far more analytic schemes being developed than there are being implemented into clinical practice. What can we do to help accelerate clinical use of these tools to realize their full potential and really impact patient care?

The past 30 years has seen dramatic changes in radiology and pathology as advances and improvements in imaging acquisition, analysis, display and storage have occurred. Additionally, public expectations in response to these changes have changed, contributing to referring clinicians and patients expecting and often demanding expert interpretation of images and other medical data not only in major urban areas, but also in areas that are rural and medically underserved. One consequence of the demand for imaging and sub-specialty interpretation is that radiologists and pathologists more than ever are expected to provide service 24/7, requiring providers to be on-call after hours and on weekends. This has led to the development of protocols and software to enable bidirectional communication between physicians, technologists, imaging managers and patients. This is where AI and related tools can also have an impact in terms of improving the efficiencies of accessing and adding to electronic medical records, peer review interfaces, and dictation systems that eliminate manual interfaces (e.g., paper-based tools, non-voice activated/controlled dictation systems) and other tools that are not well suited to increased work demands.

In many respects radiology paved the way pathology with respect to going digital earlier, thus opening the door to AI, DL and ML development and use earlier. Radiology also has a longer history of conducting observer performance studies than does pathology, but that leads to a bit of a dilemma. Once of the core gold standards for radiology is pathology, assuming that the pathologist can provide the definitive "true" answer.

*ekrupin@emory.edu

Medical Imaging 2023: Digital and Computational Pathology, edited by John E. Tomaszewski,
Aaron D. Ward, Proc. of SPIE Vol. 12471, 124710S · © 2023 SPIE
1605-7422 · doi: 10.1117/12.2663067

Numerous tools are available to screen and detect cancer, but tissue biopsy examination by pathologists is still the gold standard for the definitive diagnosis of cancer. This examination is typically performed by cutting thin slices of tissue followed by examination under the microscope, but this examination is far from an exact science and subject to significant variability. It is evident that clinical pathology variability and error is just as high in pathology as it is in radiology. For example, kappa values for assessing tumor grade in breast cancer is ~0.50.[1] A recent analysis on behalf of the International Ki67 working group (IKWG) showed high discordance rates (5-40%) between pathologists.[2] This is important clinically as Ki67 cutoff values of 20% are FDA recommended as a companion diagnostic for Abemaciclib. A recent ring study also documented poor concordance of 18 pathologists reading 170 breast cancer biopsies stained for HER2.[3] Using a 4-point scale, they found only 26% concordance between 0 and 1+ compared with 58% concordance between 2+ and 3+ expression. These data clearly show subjectivity in pathology assessment and highlight the need for tools to assess and control subjectivity among pathologists even with light microscopy.

Advances in technology have led to digital scanning of tissue sections and screen-based examination of digital whole slide images (WSI), with increasing evidence, support, and use for primary diagnoses due to high diagnostic concordance rates.[4,5] Major concerns, however, include loss of quality during image acquisition, artifacts introduced by image handling, compression, and storage,[6,7] and determining the best and most efficient viewing strategies.[8] WSIs have led to significant advances in AI tools for image segmentation and analysis[9-11] but even with explainable AI schemes "black boxes" are often trained by computers and engineers with minimal input from pathologists. There is significant concern regarding AI "trustworthiness",[12-14] particularly in difficult cases and in cases where there is an admix of tumor and normal elements. There are few if any provisions for pathologists to understand the basis of the outputs provided, forcing them to seek resolution in situations where there is discrepancy between their perceptions and those of the AI.

One of the first steps to reduce diagnostic variability, improve training methods, and better integrate decision aids such as AI into the clinical routine is to understand the perceptual and cognitive factors underlying medical decision making.[15-17] Radiology studies have used eye-tracking technology for over 50 years to characterize search strategies, the development of expertise, and causes of error and variability as radiologists diagnose radiographic images (hardcopy film and digital softcopy).[18-20] Current efforts in AI development incorporate human observers and eye-tracking to inform steps such as automated image segmentation (Figure 1).[21-23] WSI makes it possible to conduct similar studies in pathology, but *they have not included important comparisons to light microscopy* since eye-tracking to date has not been readily feasible with traditional light microscopes,[24-32] although one study videotaped pathologists viewing glass slides[33] and we have been investigating tools to capture search patterns using light microscopes (Figure 2).

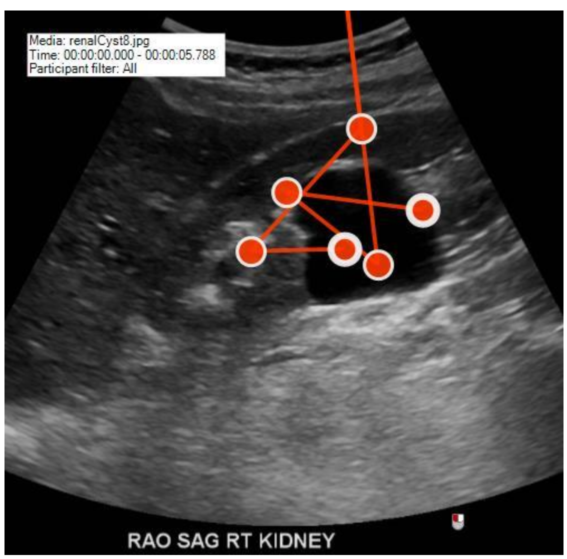

Figure 1. Using eye tracking to identify lesion edges for image annotation and segmentation.

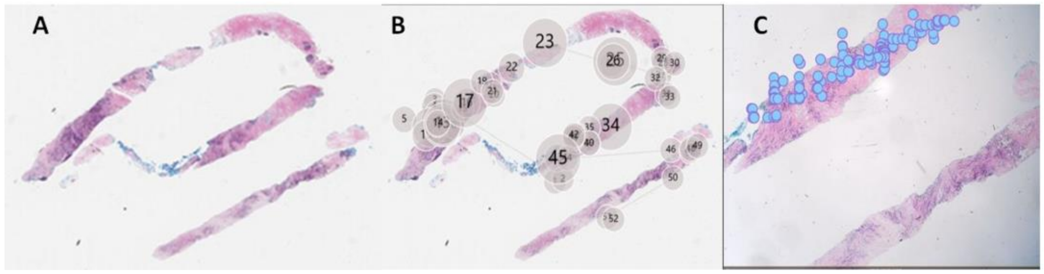

Figure 2. A) Original WSI. B) Search pattern of pathologist examining WSI on computer monitor using eye tracking technology. C) Search pattern of the same pathologist viewing glass slide under light microscope.

1.1 Focus on the task

Creating an AI scheme that can achieve levels of performance equivalent to or even better than its intended users is only half the challenge. The ultimate goal is to translate tools into clinical use so it can aid clinical decision making, improve efficiency, and impact patient care. This requires a very different approach and set of skills. Implementation sciences, human factors, and an understanding of the perceptual and cognitive processes involved in clinical decision making are key to helping ensure the successful translation of AI into clinical use. A good first step to figuring out what types of tools would be of benefit in a particular clinical setting for a given set of medical data and task is to sit down with the stakeholders – the potential users – watch what they do in their daily routine and talk with them to identify their pain points. Where do they think computer-based assistance would help with their daily routine and in what ways? For example, in radiology it is necessary to identify vertebrae numerically and by type (thoracic, lumbar etc.) so they can be readily identified in the report (Figure 3). This task does not take much skill and only a couple of minutes to annotate on an image but by the end of a day of reading cases a radiologist has likely spent 30-60 minutes doing a tedious, repetitive task a computer could readily do automatically. Having an tool that automatically identifies and adds labels to the images relieves the radiologist of the burden and tedium of a task that does not require their advanced skills and gives them the additional 30-60 minutes to read additional cases or engage in other relevant tasks. In pathology a similarly tedious task that is subject to inter and intra-observer variation and takes valuable time that could be better spent is counting (Figure 4) cell nuclei (e.g., Ki-67 images).

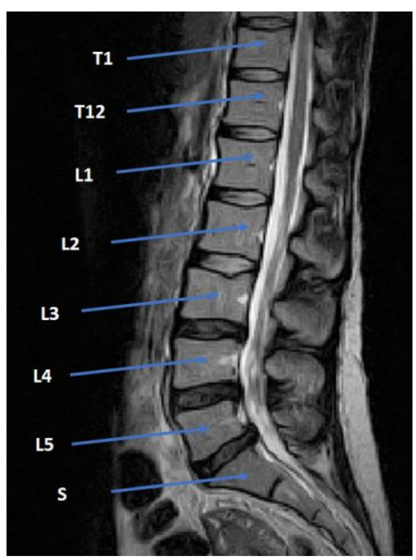

Figure 3. Example of a spine image with vertebra identified and numbered.

Once the goal has been established for the development process that does not mean the early stakeholders are no longer relevant. They should be consulted through the development process to ensure algorithm/scheme development stays on course with the original intent.

1.2 Implementation science (IS)

Interventions such as the introduction of AI tools into clinical practice that are poorly or not at all implemented cannot have the expected health benefits they were designed to have. Even when a tool is exquisitely designed and passes all the technical hurdles of validation and achieving a clinically acceptable level of performance it still may not yield expected outcomes or benefits. The real test is similar to the dust test – run your finger over a piece of furniture and if it comes back dusty you know it has not been used in quite a while. In AI, check the user logs and determine whether a tool has been used – by whom, how often, and for what. Even more importantly and even harder to assess in real world situations is whether the tools has achieved its intended outcomes – improved diagnoses, more efficient diagnoses, better or more appropriate treatment initiated, reduced patient length of stay, longer survival etc.

Implementation science is the rigorous scientific study of methods and strategies that facilitate the adoption of research and/or new technologies into regular use. It helps provide a systematic approach to understanding outcomes and processes, demonstrates value of a program/intervention/technology, identifies challenges and successes, facilitates utilization of data to address barriers and challenges, and guide implementation strategies to improve outcomes. Dissemination & implementation (D&I) research aims to accelerate timely translation evidence-based research findings to practice & policy by designing studies to better understand *how* interventions, practices, and innovations.

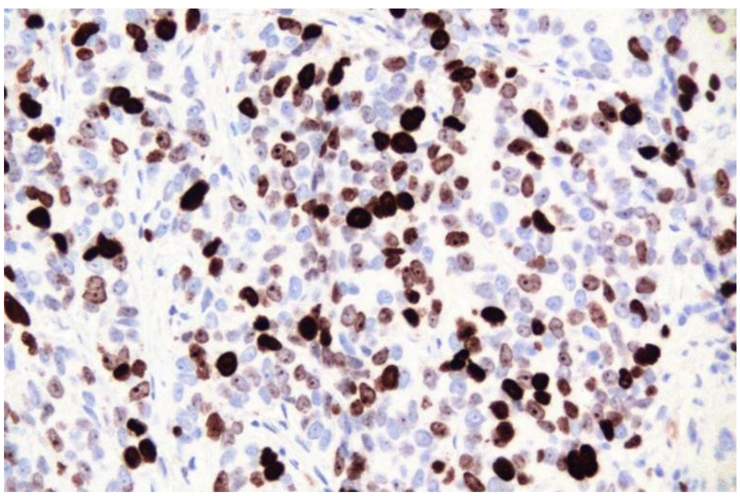

Figure 4. Example of a Ki-67 pathology image in which the task is to count cell nuclei.

There are 6 key steps to follow when designing an IS study.

1. Refine the Question
2. Determine the Study Design
3. Define the Implementation Science outcomes
4. Select the Implementation Science Framework[34]
5. Develop Data Collection Tools and/or Measures
6. Select the Implementation Strategies

Some common outcomes of interest to consider include the following. Also listed for each topic area are possible facilitators and/or barriers to implementation (F/B), and potential implementation strategies (ST). These can be tailored to a given evaluation as a function of the tools/intervention under investigation, the goals, and who the stakeholders are.

- *Acceptability:* Perception among stakeholders that tool is acceptable or satisfactory. F/B: Readiness, self-efficacy for change; perceptions complexity tool within organization that would affect implementation. ST:

Conduct educational meetings and educational outreach visits on benefits intervention to organization; distribute materials.

- *Adoption*: Initial decision to "take on" program within organization. F/B: Identify key leader or champion for new program. ST: Recruit, designate, train for leadership. Identify and prepare champions.
- *Appropriateness:* Fit and compatibility service within existing organization or environment to reach target population. F/B: Adequate demand new intervention; perceptions about complexity program and fit into current workflows. ST: Increase demand; prepare consumers to be active participants; trainings and structural design to fit into workflows.
- *Costs:* Costs implementing program or cost-effectiveness program. F/B: Costs of new program for needed additional resources. ST: Access new funding; promote adaptability within local environment.
- *Feasibility:* Indicator of fit of program in workflows particular setting. F/B: Adequate time and resources needed for implementation. ST: Organize implementation team meetings to provide dedicated time to implementation efforts; promote adaptability within environment for workload changes.
- *Fidelity:* Measure whether steps of program delivered as intended & planned. F/B: Adequate resources, staffing, time; staff knowledge of intervention and steps required for implementation. ST: Promote adaptability within local environment; audit and provide feedback; use data experts; provide ongoing consultation.
- *Penetration:* Integration service into routine practice. F/B: Continued dedication to changes needed for implementation; continued demand for service. ST: Provide ongoing consultation; intervene to increase uptake.
- *Sustainability:* Measure long-term integration and sustained usability of service within organization, environment and/or target population. F/B: Sustained interest and demand for continued implementation of new innovations. ST: Involve users in long term implementation efforts; involve executive boards for long term planning.

2. SUMMARY

The future of healthcare clearly involves computer-based AI, DL and ML tools throughout the enterprise serving many different roles for various stakeholders. Imaging informatics grew out of and will continue to shape the future of radiology and pathology. Technology development and deployment are critical to improve patient care, health outcomes, and the efficacy and efficiency with which our healthcare systems achieve these goals, but it cannot take place without considering how it will be accepted and integrated in routine daily use by all stakeholders. User-centered methods, human factors, perception and cognition, implementation science, and related research frameworks should be used to help ensure successful translation of these tools into clinical use and can provide metrics with which success can be measured objectively.

References

[1] Van Seijen, M., Jozwiak, K., Pinder, S.E., Hall, A., Krishnamurthy, S., Thomas, J.S.J., Collins, L.C., Bijron, J., Bart, J., Cohen, D., Ng, W., Bouybayoune, I., Stobart, H., Hudecek, J., Schaapveld, M., Thompson, A., Lips, E.H., Wesseling, J., the Grand Challenge PRECISION Team, "Variability in grading of ductal carcinoma in situ among an international group of pathologists," J. Pathol. Clin. Res. 7, 233-242 (2021)

[2] Acs, B., Leung, S.C.Y., Kidwell, K.M., Arun, I., Augulis, R., Badve, S.S., Bai, Y., Bane, A.L., Bartlett, J.M.S., Bayani, J., Bigras, G., Blank, A., Buikema, H., Chang, M.C., Dietz, R.L., Didson, A., Fineberg, S., Focke, C.M., Gao, D., Gown, A.M., Gutierrez, C., Hartman, J., Kos, Z., Laekholm, A.V., et al. On behalf of the International Ki67 Breast Cancer Working Group of the Breast International Group and North American Breast Cancer Group (BIG-NABCG), "Systematically higher Ki67 scores on core biopsy samples compared to corresponding resection specimen in breast cancer: a multi-operator and multi-institutional study," Modern Pathol. 35, 1362-1369 (2022)

[3] Fernandez, A.I., Liu, M., Bellizzi, A., Brock, J., Fadfare, O., Hanley, K., Harigopal, M., Jorns, J.M., Kuba, M.G., Ly, A., Podoll, M., Rabe, K., Sanders, M.A., Singh, K., Snir, O.L., Soong, T.R., Wei, S., Wen, H., Wong, S., Yoon, E., Pusztai, L., Resisenbichler, E., Rimm, D.L., "Examination of low ERB2 protein expression in breast cancer tissue," JAMA Oncol. 8, 1-4 (2022)

[4] Evans, A.J., Bauer, T.W., Bui, M.M., Cornish, T.C., Duncan, H., Glassy, E.F., Hipp, J., McGee, R.S., Murphy, D., Myers, C., O'Neill, D.G., Parwani, A.V., Rampy, A., Salama, M.E., Pantanowitz, L., "US Food and Drug Administration approval of whole slide imaging for primary diagnosis: a key milestone is reached and new questions are raised," Arch. Pathol. Lab. Med. 142, 1383-1387 (2018)

[5] Evans, A.J., Brown, R.W., Bui, M.M., Chlipala, E.A., Lacchetti, C., Milner, D.A., Pantanowitz, L., Parwani, A.V., Reid, K., Riben, M.W., Reuter, V.E., Stephens, L., Stewart, R.L., Thomas, N.E., "Validating whole slide imaging systems for diagnostic purposes in pathology: guideline update from the College of American Pathology in collaboration with the American Society for Clinical Pathology and the Association for Pathology Informatics," Arch. Pathol. Lab. Med. 146, 440-450 (2022)
[6] Randell, R., Ambepitiya, T., Mello-Thoms, C., Ruddle, R.A., Brettle, D., Thomas, R.G., Treanor, D., "Effect of display resolution on time to diagnosis with virtual pathology slides in a systematic search task," J. Dig. Imaging 28, 68-76 (2015)
[7] Zarella, M.D., Bowman, D., Aeffner, F., Farahani, N., Xthona, A., Absar, S.F., Parwani, A., Bui, M., Hartman, D.J., "A practical guide to whole slide imaging: a white paper from the Digital Pathology Association," Arch. Pathol. Lab. Med. 143, 222-234 (2019)
[8] Hanna, M.G., Reuter, V.E., Hameed, M.R., Tan, L.K., Chiang, S., Sigel, C., Hollmann, T., Giri, D., Samboy, J., Moradel, C., Rosado, A., Otilano, J.R., England, C., Corsale, L., Stamelos, E., Yagi, Y., Schuffler, P.J., Fuchs, T,, Klimstra, D.S., Sirintrapun, S.J., "Whole slide imaging equivalency and efficiency study: experience at a large academic medical center," Modern Pathol. 32, 916-928 (2019)
[9] Tizhoosh, H.R., Pantanowitz, L., "Artificial intelligence and digital pathology: challenges and opportunities," J. Pathol. Inform. 9, 38 (2018)
[10] Rahka, E.A., Toss, M., Shilino, S., Gamble, P., Jaroensri, R., Mermel, C.H., Chen, P.H.C., "Current and future applications of artificial intelligence in pathology: a clinical perspective," J. Clin. Pathol. 74, 409-414 (2021)
[11] Baxi, V., Edwards, R., Montalto, M., Saha, S., "Digital pathology and artificial intelligence in translational medicine and clinical practice," Modern Pathol. 35, 23-32 (2021)
[12] Alvarado, R., "Should we replace radiologists with deep learning? Pigeons, error and trust in medical AI," Bioethics 36, 121-133 (2022)
[13] Drogt, J., Milota, M., Vos, S., Bredenoord, A., Jongsma, K.. "Integrating artificial intelligence in pathology: a qualitative interview study of users' experiences," Modern Pathol. 35, 1540-1550 (2022)
[14] Von Eschenbach, W,J., "Transparency and the black box problem: why we do not trust AI," Philosophy & Technol. 34, 1607-1622 (2021)
[15] Krupinski, E.A., "Perceptual factors in reading medical images," In: Samei, E., Krupinski, E.A. (eds). The Handbook of Medical Image Perception and Techniques, New York, NY: Cambridge University Press; 2018. Pp. 95-106.
[16] Manning, D., "Cognitive factors in reading medical images: thinking processes in interpreting medical images," In: Samei, E., Krupinski, E.A. (eds). The Handbook of Medical Image Perception and Techniques. New York, NY: Cambridge University Press; 2018. Pp. 107-120.
[17] Pantanowitz, L., Mello-Thoms, C., Krupinski, E.A., "Perception issues in pathology," In: Samei, E., Krupinski, E.A. (eds). The Handbook of Medical Image Perception and Techniques. New York, NY: Cambridge University Press; 2018. Pp. 495-505.
[18] Kundel ,H.L., Nodine, C.F., "A short history of image perception in medical radiology," In: Samei, E., Krupinski, E,A, (eds). The Handbook of Medical Image Perception and Techniques. New York, NY: Cambridge University Press; 2018. Pp. 11-22.
[19] Van der Gijp, A., Ravensloot, C.J., Jarodzka, H., van der Shaik, J.P.J., ten Cate, T.J., "How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology," Adv. Health Sci. Educ. 22, 765-787 (2017)
[20] Krupinski, E.A., "The importance of perception in imaging: past and future," Sem. Nuc. Med. 41, 392-400 (2011)
[21] Karagyris, A., Kashyap, S., Lourentzou, I., Wu, J.T., Sharma, A., Tong, M., Abedin, S., Beymer, D., Mukherjee, V., Krupinski, E.A., Moradi, M., "Creation and validation of a chest x-ray dataset with eye-tracking and report dictation for AI development," Scientific Data 8, 92 (2021)
[22] Stember, J.N., Celik, H., Gutman, D., Swinburne, N., Young, R., Eskreis-Winkler, S., Holodny, A., Jambawalikar, S., Wood, B.J., Chang, P.D., Krupinski, E.A., Bagci, U., "Integrating eye tracking and speech recognition accurately annotates MR brain images for deep learning: proof of principle," Radiol. AI. 3, e200047 (2020)
[23] Stember, J.N., Celik, H., Krupinski, E., Chang, P.D., Mutasa, S., Wood, B.J., Lignelli, A., Moonis, G., Schwartz, L.H., Jambawalikar, S., Bagci, U., "Eye tracking for deep learning segmentation using convolutional neural networks." J. Dig. Imag. 32, 597-604 (2019)
[24] Tiersma, E.S.M., Peters, A.A.W., Mooij, H.A., Fleuren, G.J., "Visualising scanning patterns of pathologists in the grading of cervical intraepithelial neoplasia," J. Clin. Pathol. 56, 677-680 (2003)

[25] Krupinski, E.A., Tillack, A.A., Richter, L., Henderson, J.T., Bhattacharyya, A.K., Scott, K.M., Graham, A.R., Descour, M.R., Davis, J.R., Weinstein, R.S., "Eye-movement study and human performance using telepathology virtual slides. Implications for medical education and differences with experience," Hum. Path. 37, 1543-1556 (2006)
[26] Krupinski, E.A., Graham, A.R., Weinstein, R.S., "Characterizing the development of visual search expertise in pathology residents viewing whole slide images," Hum. Path. 44, 357-364 (2013)
[27] Mello-Thoms, C., Mello, C.A.B., Medvedeva, O., Castine, M., Legowski, E., Gardner, G., Tseytlin, E., Crowley, R., "Perceptual analysis of the reading of dermatopathology virtual slides by pathology residents," Arch. Pathol. Lab. Med. 136, 551-562 (2012)
[28] Jaarsma, T., Jarodzka, H., Nap, M., van Merrienboer, J.J.G., Boshuzien, H.P.A., "Expertise in digital pathology: combining the visual and cognitive perspective," Adv. Health Sci. Educ. 20, 1089-1106 (2015)
[29] Brunye, T.T., Mercan, E., Weaver, D.L., Gilmore, J.G., "Accuracy is in the eyes of the pathologist: the visual interpretive process and diagnostic accuracy with digital whole slide images," J. Biomed. Inform. 66, 171-179 (2017)
[30] Drew, T., Lavelle, M., Kerr, K.F., Shuchard, H., Brunye, T.T., Weaver, D.L., Elmore, J.G., "More scanning, but not zooming, is associated with diagnostic accuracy in evaluating digital breast pathology slides," J. Vision 21, 1-17 (2021)
[31] Ghezloo, F., Wang, P.C., Kerr, K.F., Brunye, T., Drew, T., Chang, O.H., Reisch, L.M., Shapiro, L.G., Elmore, J.G., "An analysis of pathologists' viewing processes as they diagnose whole slide digital images," J. Pathol. Inform. 13, 100104 (2022)
[32] Mercan, E., Shapiro, L.G., Brunye, T.T., Weaver, D.L., Elmore, J.G., "Characterizing diagnostic search patterns in digital breast pathology: scanners and drillers," J. Dig. Imag. 31, 32-41 (2018)
[33] Crowley, R.S., Naus, G.J., Stewart, J., Friedman, C.P., "Development of visual diagnostic expertise in pathology: an information-processing study," JAMIA 10, 39-51 (2003)
[34] Tabak, R.G., Khoong, E.C., Chambers, D.A., Brownson, R.C., "Bridging research and practice: models for disseminationand implementation research," Am. J. Prev. Med. 43, 337-350 (2012)