Welcome to the second issue of the *Journal of Medical Imaging* (JMI) for the 2024 year!

With a few months to settle in, I've discovered many of the complexities and intricacies that are facing our communities. While considering this position, I did not fully appreciate the transformative challenges that we are facing in terms of AI ethics and nonprofit academic publishing models. Nevertheless, we have a tremendous community that I am continually learning to appreciate more deeply for their passion, insight, and creativity.

Aside from the relatively dry job of assigning handling editors, I have the unique opportunity to preview a wide variety of fantastic work that I would not have had the pleasure of reading if I were simply browsing the headlines and flipping for articles that were specifically in my field. Frankly, being EIC is like being able to attend a little bit of the SPIE Medical Imaging conference every day. It's fun! Thank you for all for being here.

In 2024's Volume 11 Issue 1, we presented fascinating results, and I would like to focus on one today: our most recent presentation of an SPIE-originated "Challenge" in JMI. Today, I am celebrating the work of Heiselman et al.[1] for the Image-to-Physical Liver Registration Sparse Data Challenge and for bringing SPIE and JMI into this conversation. Their article is an important link between surgical navigation and image processing. I'm impressed with the millimeter-level accuracy that the teams were able to achieve. More important, from my perspective, is that they've established a state of the art that is community driven.

I have been involved in challenges since my academic career began just over a decade ago. [In the interest of full disclosure, I am active in Medical Image Computing Computer-Assisted Intervention (MICCAI) Special Interest Group (SIG) on Challenges.] These challenges can serve many purposes, from the flagship Grand Challenges and Kaggle Challenges that bring thousands of people together to solve data science problems often with substantial cash prizes, to the very small ones, with a few dozen people in a room and 3-D printed trophies, focusing on a problem that hasn't been examined in depth. All too often we're trapped in our own echo chambers or bubbles and our discoveries become difficult to translate beyond our research team. Challenges help us get outside of our comfort zone. I will be the first to admit that there are plenty of issues with challenges regarding gaming, circular reasoning, unintentional information leakage, and problematic metrics (especially over a long period in time). Yet, for all their flaws and problems, challenges bring people together.

Now we may worry about Goodhart's Law, "When a measure becomes a target, it ceases to be a good measure,"[2] ... and I do. However, with this target, we can move forward. As long as we are honest and clear about the level of validation/data interrogation, these challenges allow us to jumpstart the discussion and expand our community. Of course, emerging innovation needs to be validated and reproduced beyond challenges.

My primary goal with JMI and, in fact, most of my academic career, is to bring people together and allow creative ideas to shine. These challenges are a prime example of what is possible and the best of what we can be. I look forward to seeing more of these challenges organized within the SPIE Medical Imaging community and being shared through JMI.

Yet, my enthusiasm is not completely unrestrained. There are a couple of caveats that we've learned over the past decade, which I would like to share.

First, it's critical to understand how the data were acquired and how the study was designed. To aid in assessment, many of us have worked together to create BIAS reporting standards for challenges.[3] This framework allows preregistration of challenges much like a clinical trial to provide transparency for interpretation and validation. I particularly welcome engagement in challenges that seek to rigorously self-assess.

Second, there are substantial differences in how one would assess an algorithm for different areas. The "Metrics Reloaded" community has come together recently and published an article describing many of these challenges.[4] Perception and observer assessment are a key area of JMI. In this issue, Drukker et al.[5] created a "metrology resource to assist researchers in evaluating the performance of their medical image analysis ML algorithms." Meta-analysis of challenge designs themselves is a science that is just beginning to be understood as a science. I especially welcome continued innovation toward understanding the metrics and the observers that we should use in these areas and challenges in our community.

Finally, many ranking designs are unstable when we assess multiple possible combinations of metrics and submissions. This instability is a fundamental problem with maximal and minimal statistics focusing on outlier distributions. As we rank challenges, it is important to understand how distributional assumptions impact the selection of "winners." The statistics and biostatistics community has a lot to offer in this field. One of many possible areas is in the "challengeR" type of framework, for example.[6] I invite authors to consider the stability of their results.

I welcome suggestions for references and approaches to share with the community as we continue on this journey of exploration together. I wish you many exciting discoveries and Eureka moments as you enjoy this issue.

Warm regards,
Bennett Landman
JMI Editor-in-Chief

P.S. The photo accompanying this letter is from my recent trip to Capitol Hill with AIMBE to discuss the positive impacts of science on our Middle-Tennessee community. I hope that we can continue to connect through SPIE and beyond.

References

1. J. S. Heiselman et al., "The image-to-physical liver registration sparse data challenge: comparison of state-of-the-art using a common dataset," *J. Med. Imaging* **11**(1), 015001 (2024).
2. M. Strathern, "'Improving ratings': audit in the British University system," *Eur. Rev.* **5**(3), 305–321 (1997).
3. L. Maier-Hein, "BIAS: Transparent reporting of biomedical image analysis challenges," *Med. Image Anal.* **66**, 101796 (2020).
4. L. Maier-Hein et al., "Metrics reloaded: recommendations for image analysis validation," *Nat. Methods* **21**(2), 195–212 (2024).
5. K. Drukker et al., "MIDRC-MetricTree: a decision tree-based tool for recommending performance metrics in artificial intelligence-assisted medical image analysis," *J. Med. Imaging* **11**(2), 024504 (2024).
6. M. Wiesenfarth et al., "Methods and open-source toolkit for analyzing and visualizing challenge results," *Sci. Rep.* **11**(1), 1–15 (2021).