State of the art in the task-based assessment of medical imaging systems

Kyle J. Myers*
Puente Solutions LLC, 3302 E Tonto Dr., Phoenix, AZ 85044

ABSTRACT

Image Science provides a framework for the task-based assessment of image quality. This framework has been used to support the evaluation of medical imaging system hardware, iterative reconstruction algorithms and other image processing methods, and display devices by academia, industry, and US FDA. Since the earliest instances of the SPIE Medical Imaging Symposium the conference has served as an essential venue for presentations and discussions related to the objective assessment of image quality, featuring first disclosures of new models for physiological backgrounds and pathologies, tools for simulating medical imaging systems, models for the human and Bayesian observer, and methods for computing task-based figures of merit. This paper highlights recent advances in the objective or task-based assessment of image quality through the use of computational models and methods, and points to new initiatives intended to develop resources to move this important field forward.

Keywords: computational models, virtual clinical trials, reader studies, model observers

1. INTRODUCTION

The objective assessment of image quality is essential for determining how well an imaging system supports making decisions or drawing inferences from the resulting images. A better imaging system is one that allows for better inferences regarding the object *on average*. Comparisons of example images can be indicative, but not sufficient. Task-based assessment is a statistical methodology.

The defining elements of task-based assessment of image quality were laid out comprehensively by Barrett and Myers.[1] These elements include defining the task in terms of the inferences one wishes to make about the object, and the associated figure of merit. One must specify the particular set of objects (patients in a medical imaging context), imaging systems, observers – whatever entity ingests image data and renders inferences – and statistical tools that allow for meaningful comparisons across imaging system evaluations. Much progress is being made in the development of tools and resources that facilitate the application of task-based assessment methodologies for more realistic imaging tasks. Especially true in this regard is the harnessing of modern computational power through *in silico* modeling of all elements of the imaging chain, bringing us ever closer to the possibility of realizing rigorous, predictive task-based assessment studies of medical imaging systems without humans in the loop either as patients in a clinical study or as image readers. Section 2 provides some examples of recent advances in the essential components for task-based assessment of medical imaging systems. These examples are intended to give an indication of the many recent, excellent contributions to these efforts and are in no way a comprehensive review of the field. This reality is the happy result of the fact that there are so many oustanding groups and contributors to this growing field. Section 3 describes efforts that pull these components together. Concluding remarks are offered in Section 4.

2. RECENT ADVANCES IN TASK-BASED ASSESSMENT

2.1 Models for patients

Early studies of task-based image quality in medical imaging made use of highly stylized object models, typically in which signals were exactly known and backgrounds were uniform. These studies included theoretical investigations[2] and those making use of simplistic digital[3] or physical[4] imaging phantoms with ideal, human observers, or both.[5] It soon became clear that more complex models for the object were needed to avoid conclusions regarding system quality that did not generalize well to more complex imaging scenarios. As a result, investigators turned to statistical models for

Medical Imaging 2023: Physics of Medical Imaging, edited by Lifeng Yu, Rebecca Fahrig, John M. Sabol, Proc. of SPIE Vol. 12463, 1246302
 · 1605-7422 · doi: 10.1117/12.2662392

backgrounds and signals that were defined statistically.[6-8] As highlighted next, we've come a very long way since those early efforts to develop models for signals and backgrounds with greater complexity.

Procedural, analytical, or parameterized models are one major area of significant technical advances in recent years. For example, the digital breast model of Graff et al.[9] makes use of a set of procedures to generate the major anatomical structures of the female breast, including fat and glandular tissues, the ductal tree, vasculature, and ligaments. The model allows for patient characteristics such as breast shape, volume, density, and compressed thickness to be controlled. Source code for the model is freely available.[10] Models for masses[11] and microcalcifications[12] based on procedure models allow for the insertion of pathologies in such digital breast objects, resulting in a family of digital breasts with simulated lesions of varying size and shape, and with even the possibility of modeling tumor growth over time.[13]

Hybrid models combine clinical backgrounds with edited clinical lesions to create larger datasets. These models typically leverage the much larger numbers of normal images at hand, and create a new, larger set of abnormal images by inserting lesions that were digitally harvested from abnormal images. These models have been demonstrated in the domain of the raw/detected data[14-16] as well as in the domain of reconstructed images.[17] Because such procedures make use of real clinical backgrounds and pathologies, they have the advantage of clinical realism, so long as the lesion insertion process does not result in noticeable artifacts. However, as these methods start with image domain data, they are less amenable to evaluations of imaging system hardware. They are more likely to be useful in the evaluation of competing reconstruction methods, image processing tools, display devices, and data augmentation for training of artificial intelligence and deep learning (AI/DL) algorithms.

Atlas models derive families of patients through the segmentation of clinical image sets. The XCAT Phantom Program developed by investigators of the Duke University's Center for Virtual Imaging Trials is a prominent example of this approach.[18] Because these phantoms are derived from clinical images, they are necessarily limited by the imaging hardware used to acquire the underlying images. Research investigators are working to develop approaches for inserting sub-resolution structures and textures from other imaging modalities or via modeling methods.

A very recent entry into the toolbox for creating models of patients is the use of learning models based on Generative Adversarial Networks (GANs). Anastasio's team at UIUC have contributed several important papers demonstrating this approach.[19-20] The Grand Challenge on "Deep Generative Modeling for Learning Medical Imaging Statistics" hosted by UIUC and the US FDA, and hosted by the American Association of Physicists in Medicine (AAPM), will provide the community with a state-of-the-art understanding of these models' ability to reproduce visually similar images that also preserve task-based relevant statistical properties.[21] The beauty of this challenge is that, because the models will be trained on a set of digital breast models with known statistical properties,[9] the GAN model entries will be able to be evaluated against a gold standard in object space. If such GANs can be shown to preserve diagnostically relevant information in a task-based sense, they will open new avenues for task-based assessments.

2.2 Models for imaging systems

Solid and steady progress has been made on the expansion of realistic physics models that incorporate recent innovations in imaging hardware, including geometries with multiple and/or moving sources, new detector materials and mechanisms, and so on. As computing power and storage has grown, so has our ability to model the many physical interactions that comprise the imaging process with greater realism and precision. This area of endeavor has long been the bedrock of the Physics Conference at the SPIE Medical Imaging Symposium.

Research into the potential for GANs to synthesize images is a more recent area of interest. GANs might be able to learn the conversion from one kind of acquisition system to another.[22] GANs are also being investigated for their potential to learn the forward imaging model so as to produce realistic images from digital body phantoms.[23] Much work remains to be done to understand the validity of these approaches in terms of their ability to support the task-based evaluation of imaging systems.

2.3 Models for readers/observers

Broadly speaking, observer models follow two tracks. The first is the study of human perception in order to better understand the ability of humans to extract information from images. From such studies, models for human readers are being developed for use in the evaluation of imaging tools that are intended to create or display data for human interpretation, particularly image reconstruction algorithms, image processing methods, and novel display or visualization systems. The state-of-the-science with respect to modeling human perception includes the elucidation of the human observer's template for more types of tasks,[24] the creation and validation of models for human interpretation that incorporate foveation for visual search in 3D medical images,[25] and the prediction of human performance in classification tasks via deep learning models.[26,27]

The second track in the modeling of observers is pursued by those who seek to model the Ideal or Bayesian Observer (IO). By definition the IO captures all information in the detected data, without adding noise or uncertainty. The difference in performance between the IO and the human observer for a particular task offers a window into the opportunity for improved inferences from those images via computer-aided diagnosis, or perhaps in the future computer diagnosis. For the purpose of this paper, however, the importance of the IO is in its ability to facilitate the evaluation of image acquisition systems on an absolute scale. A search of the SPIE Digital Library will demonstrate that over many decades, presentations at the SPIE Medical Imaging Symposium have grown our ability to compute and understand IO performance and that of its close cousin, the optimal linear or Hotelling observer, for tasks of increasing range and complexity. Moreover, it has long been known that neural networks can approximate a Bayes optimal discriminant.[28] While this was true in theory, in practice the computing requirements did not match the computing power at hand, leaving us still limited with respect to the task-based evaluations we could execute for the IO. That reality is rapidly changing. The SPIE 2020 Medical Imaging Symposium was a breakout year for papers that presented approaches to approximating the ideal observer.[29-31]

2.4 Statistical methods

Rigorous science requires error bars. Task-based assessments require the incorporation of all sources of uncertainty in the evaluation process that contribute to uncertainty in the figure of merit. These sources include the variability in the underlying patients being imaged, measurement noise from the imaging system, and uncertainty coming from the readers. In the past 10+ years our community has come to appreciate that Multi-Reader, Multi-Case (MRMC) analysis is essential for testing the significance of differences in competing imaging modalities for detection and discrimination tasks. MRMC methods give the total uncertainty in estimates of ROC-based metrics stemming from the range of case difficulty, reader skill and mindset, and their interactions. That's not particularly new. What is new is the development of "split-plot" methodologies that allow for the analysis of more general study designs than the fully-crossed design in which every reader reads every case for each modality. Split-plot designs can be used to reduce the number of reads per reader and the number of total reads. Furthermore, for the same number of total reads, the split-plot design has been found to be more statistically efficient.[32-35]

There has been considerable progress in recent years in methods for collecting better data from human readers. Data from human observers may be needed to serve as the reference standard for circumstances in which other forms of truth are not available. A prominent example of this is in pathology applications. In order to assess a new digital pathology imaging device, including the many efforts to develop AI/DL methods to assist in pathology tasks, it is common to make use of experts for the "truthing" of pathology data. New tools are emerging for estimating truth from panels of expert readers, including improved approaches to reader training to reduce reader variability, improved collection interfaces, and statistical analysis tools.[36,37] Similarly, better approaches are emerging for collecting data from humans who are serving as "study" readers in a task-based assessment. These also include improved training methods and data collection interfaces that facilitate the collection of more precise reader data on a finer measurement scale.[33] US FDA/CDRH shares examples of reader training materials and data collection interfaces that facilitate the collection of improved reader data, along with powerful statistical software tools for reader or algorithm study design and data analysis.[38,39] The inclusion of iMRMC statistical tools in the FDA/CDRH Regulatory science tools catalog is further indication of the FDA's interest and support for the development and use of such tools in medical imaging device development and review.

The role of clinicians in task-based assessments of image quality was highlighted in a recent consensus paper from the Society of Nuclear Medicine and Molecular Imaging.[40] That group noted that clinicians are important partners in these investigations, playing key roles as study readers and expert truthers. They also assist in the selection of appropriate clinical tasks and patient populations.

There has also been significant recent progress related to the assessment of imaging systems for tasks related to the estimation of quantitative imaging biomarkers (QIBs). RSNA's Quantitative Imaging Biomarker Alliance (QIBA) published a series of papers in 2015 that laid out a comprehensive framework for QIBs, intending to standardize the terminology,[41] methods for evaluating imaging biomarkers,[42] and methods for comparing imaging biomarkers.[43] Now comes the publication of a set of papers in Academic Radiology that expands QIBA's earlier work to multiparametric quantitative imaging biomarkers.[44-49]

3. PUTTING IT ALL TOGETHER

With so many advances in models as building blocks for performing task-based assessment of medical imaging devices, we now have the opportunity to evaluate medical imaging systems entirely *in silico* and demonstrate comparable results to actual clinical studies. A notable example is the US FDA/CDRH's Virtual Imaging Clinical Trial for Regulatory Evaluation or VICTRE.[50] That work set out to replicate an actual clinical trial performed by a company seeking marketing authorization in the US for a digital breast tomosynthesis system as a replacement for digital mammography. The VICTRE investigators utilized *in silico* models for the study subjects, the imaging systems under comparison, and the image readers. The difference in the estimated ROC-based figures of merit for the two systems was consistent with that of the actual trial and estimated at a fraction of the cost. The tools underlying the VICTRE study are publicly available and included in the US FDA/CDRH's regulatory science tool catalog.[51]

The development of a metaverse for intelligent healthcare[52] will incorporate all aspects of task-based image quality assessment: digital models for patients, or so-called digital twins, with patient-specific inserted diseases that are scanned *in silico* to determine the best imaging protocol for that patient. Human and model/AI observers will be essential to the task-based evaluations that select the best imaging protocol as well as the interpretation of the actual data that comes from the imaging optimization step. To achieve this futuristic vision, our community is working together on many collaborative efforts related to data and software sharing. An outstanding and very recent example is the Medical Imaging Data Resource Center's (MIDRC)[53] effort to develop an open, curated image data commons and machine intelligence computational capabilities. The open data commons will serve many purposes, and for task-based assessment applications these include the development and validation of models for patients and scanners as well as models for image readers. The sequestered data being set aside by MIDRC is truly intended for task-based assessment of AI/DL algorithms. To support MIDRC investigators and the wider community, MIDRC is developing a decision tree to assist users in selecting task-based performance metrics.[54] This latter effort will encourage the use of consistent metrics and our ability to compare the results across studies. Communities of practice and collaboration like that of MIDRC are essential.

4. CONCLUDING REMARKS

The language and methodology of "task-based image quality assessment" has become commonplace, with a great many papers making use of this phrase in their titles and many talks at the SPIE Medical Imaging Symposium on this topic. There truly has been tremendous progress on all aspects and across imaging modalities and pipeline elements. Future efforts like that of the VICTRE study and ones like it will continue to bring down the cost of *in silico* task-based assessments as simulation models for each of the components of such studies are developed, validated, and shared. Open science in this spirit will reduce the redundancy in small efforts while increasing the likelihood of robustness and generalizability of the evaluations that are performed in this manner, leading to improved imaging systems – improved derived inferences – for patients.

REFERENCES

[1] Barrett, H.H. and Myers, K.J., Foundations of Image Science, Wiley and Sons, New York, (2004).
[2] Wagner, R.F., Brown, D.G., Metz, C.E., "On the multiplex advantage of coded source/aperture photon imaging," Proc. SPIE 314, 72-76 (1981).
[3] Burgess, A.E., Wagner, R.F., Jennings, R.J., and Barlow, H.B., "Efficiency of Human Visual Signal Discrimination," Science 214(2), 93-94 (1981).
[4] Gagne, R.M., Gallas, B.D., and Myers, K.J., "Toward objective and quantitative evaluation of imaging systems using images of phantoms," Med. Phys. 33(1) 83-95 (2006).
[5] Tsui, B.M.W., Metz, C.E., Atkins, F.B., Starr, S.J., and Beck, R.N., "A Comparison of Optimum Detector Spatial Resolution in Nuclear Imaging based on Statistical Theory and on Observer Performance," Phys. Med. Biol. 23(4) 654-676 (1978).
[6] Barrett, H. H., Rolland, J. P., Wagner, R. F. and Myers, K. J., "Detection and discrimination of known signals in inhomogeneous, random backgrounds," Proc. SPIE 1090, 176-182 (1989).
[7] Burgess, A. E., "Statistically defined backgrounds: Performance of a modified non-prewhitening observer model," J. Opt. Soc. Am. A 11, 1237-1242 (1994).
[8] Bochud, F. O., Abbey, C. K. and Eckstein, M. P., "Statistical texture synthesis of mammographic images with clustered lumpy backgrounds," Opt. Express 4, 193-199 (1999).
[9] Graff, C.G. "A new, open-source, multi-modality digital breast phantom," Proc. SPIE 9783 978309 (2016).
[10] "GitHub - DIDSR/breastPhantom: A multi-modality anthropomorphic digital breast phantom," (accessed 28 February 2023).
[11] de Sisternes, L., Brankov, J.G., Zysk, A.M., Schmidt, R.A., Nishikawa, R.M., Wernick, M.N., "A computational model to generate simulated three-dimensional breast masses," Med. Phys. 42(2) 1098-1118 (2015).
[12] Bakic, P.R., Barufaldi, B., Higginbotham, D., Weinstein, S., Avanaki, A., Espig, K., et al. "Virtual clinical trial of lesion detection in digital mammography and digital breast tomosynthesis," Proc. SPIE 1057306. (2018).
[13] Sengupta, A., Sharma, D., Badano, A., "Computational model of tumor growth for in silico trials," Proc. SPIE 115954S (2021).
[14] Chen, B., Leng., S., Yu, L., Ma, C., McCollough, C. "Lesion insertion in the projection domain: Methods and initial results," Med. Phys. 42(12) 7034-7042 (2016).
[15] Ghanian, Z., Pezeshk, A., Petrick, N., Sahiner, B. "Computational insertion of microcalcification clusters on mammograms: reader differentiation from native clusters and computer-aided detection comparison," J. Med. Imag. 5(4) 044502 (2018).
[16] Dilger, S.K.N., Leng, S., Chen, B., Carter, R.E., Favazza, C.P., Fletcher, J.G., McCollough, C.H., Yu, L., "Localization of liver lesions in abdominal CT imaging: II. Mathematical model observer performance correlates with human observer performance for localization of liver lesions in abdominal CT imaging," Phys. Med. Biol. 64 105012 (2019).
[17] Pezeshk, A., Sahiner, B., Zeng, R., Wunderlich, A., Chen, W., Petrick, N. "Seamless Insertion of Pulmonary Nodules in Chest CT Images," IEEE TBE 62(12) 2812-2827 (2015).
[18] "XCAT Phantom Program - CVIT - Center for Virtual Imaging Trials (duke.edu)," (accessed 3 March 2023).
[19] Zhou, W., Bhadra, S., Brooks F.J., Li, H. Anastasio, M.A., "Learning stochastic object models from medical imaging measurements by use of advanced ambient generative adversarial networks," J. Med. Imag. 9(1) 015503 (2020).
[20] Kelkar, V., Gotsis, D.S., Brooks, F.J., KC, P., Myers, K.J., Zeng, R., Anastasio, M.A. "Assessing the ability of generative adversarial networks to learn canonical medical image statistics, IEEE TMI Early Access Article (2023).
[21] "Deep Generative Modeling for Learning Medical Image Statistics (DGM-Image Challenge) (aapm.org)," (accessed 3 March 2023).
[22] Ghanian, Z., Badal, A., Cha, K., Farhangi, M.M., Petrick, N., Sahiner, P. "Mammographic Image Conversion Between Source and Target Acquisition Systems Using cGAN," In: Liu, M., Yan, P., Lian, C., Cao, X. (eds) Machine Learning in Medical Imaging, LNIP 12436 (2020).
[23] Russ T, Goerttler S, Schnurr A-K, et al. "Synthesis of CT images from digital body phantoms using CycleGAN," International Journal of Computer Assisted Radiology and Surgery 14(10) 1741-1750 (2019).
[24] Abbey, C.K., Samuelson, F.W., Zeng, R., Boone, J.M., Eckstein, M.P., Myers, K.J, "Human observer templates for lesion discrimination tasks," Proc of SPIE Medical Imaging, vol 113160U, (2020).

[25] Lago, M.A., Abbey, C.K., Eckstein, M.P., "Foveated Model Observers for Visual Search in 3D Medical Images," IEEE TMI 40 (3) 1021-1031 (2021).
[26] Gong J. Med. Imag., (2020).
[27] Fan, F., Ahn, S., De Man, B., Wangerin, K.A., Wollenweber, S.D., Abbey, C.K., Kinahan, P.E., "Deep learning-based model observers that replicate human observers for PET imaging," Proc. SPIE 11316E (2020).
[28] Ruck, D. W., Rogers, S.K., Kabrisky, M., Oxley, M.E., Suter, B., "The multilayer perceptron as an approximation to a Bayes optimal discriminant function," IEEE Trans. Neural Networks 1, 296--298 (1990).
[29] Granstedt, J.L. Zhou, W., Anastasio, M.A., "Learning efficient channels with a dual loss autoencoder," Proc. SPIE 113160C (2020).
[30] Zhou, W., Anastasio, M.A., "Markov-Chain Monte Carlo approximation of the Ideal Observer using generative adversarial networks," Proc. SPIE 113160D (2020).
[31] Lorente, I., Abbey, C.K., Brankov, J.G., "Deep learning based model observer by U-Net," Proc. SPIE 113160F (2020).
[32] Obuchowski, N.A., Gallas, B.D., Hillis, S.L. "Multi-reader ROC Studies with Split-plot Designs: A Comparison of Statistical Methods," Acad. Radiol. 19 1508-1517 (2012).
[33] Gallas, B.D., Chen, W., Cole, E., Ochs, R., Petrick, N., Pisano, E.D., Sahiner, B., Samuelson, F.W., Myers, K.J., "Impact of prevalence and case distribution in lab-based diagnostic imaging studies," J. Med. Imag. 6(1) 015501 (2019).
[34] Chen, W., Gong., Q., Gallas, B.D. "Paired split-plot designs of multireader multicase studies," J. Med. Imag. 5(3) 031410 (2018).
[35] Huang, E.P., Shih, J.H., "Assigning readers to cases in imaging studies using balanced incomplete block designs," Stat. Methods Med. Res. 30(10):2288-2312 (2021).
[36] Elfer, K.N., Barcia, V., Gallas, B.D., "Analysis of a pilot study collecting pathologist annotations for validating machine learning algorithms," Proc SPIE 12467-24 (2023).
[37] Garcia, V., Elfer, K.N., Peeters, D.J.E., Ehinger, A., Werness, B., Ly, A., Li, X., Hanna, M.G., Blenman, K.R.M., Salgado, R., "Development of Training Materials for Pathologists to Provide Machine Learning Validation Data of Tumor-Infiltrating Lymphocytes in Breast Cancer," Cancers 14, 2467 (2022).
[38] NCI Hub - Group: eeDAP studies ~ Wiki: HTT Data Collection Training (cancer.gov), (accessed 3 March 2023).
[39] GitHub - DIDSR/iMRMC: iMRMC: Software to do multi-reader multi-case analysis of reader studies, (accessed 3 March 2023).
[40] Jha, A.K., Myers, K.J., Obuchowski, N.A., Liu, Z., Rahman, A., Saboury, B., Rahmim, A., Siegel, B.A., "Objective Task-Based Evaluation of Artificial Intelligence-Based Medical Imaging Methods: Framework, Strategies, and Role of the Physician," PET Clinics 16 493-511 (2021).
[41] Kessler, L.G., Barnhart, H.X., Buckler, A.J., Choudhury, K.R., Kondratovich, M.V., Toledano, A., Guimaraes, A., Filice, R., Zhang, Z., Sullivan, D.C., "The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions," Stat. Meth. Med. Res., 24(1): 9-26 (2015).
[42] Raunig, D.L., McShane, L.M., Pennello, G., Gatsonis, C., Carson, P.L., Voyvodic, J.T., Wahl, R.L., et al., "Quantitative imaging biomarkers: A review of statistical methods for technical performance assessment," Stat. Meth. Med. Res. 24(1): 27-67 (2015).
[43] Obuchowski, N.A., Reeves, A.P., Huang, E.P., et al. "Quantitative imaging biomarkers: A review of statistical methods for computer algorithm comparisons," Stat. Meth. Med. Res. 24(1):68-106 (2015).
[44] Obuchowski, N.A., Hall, T.J., "Introduction to Multiparametric QIB Series," Acad. Radiol. 30(2) 145-146 (2023).
[45] Obuchowski, N.A., Huang, E., deSouza, N.M., et al., "A Framework for Evaluating the Technical Performance of Multiparameter Quantitative Imaging Biomarkers (mp-QIBs) Acad. Radiol. 30(2) 147-158 (2023).
[46] Raunig, D.L., Pennello, G.A., Delfino, J.G., et al., "Multiparametric Quantitative Imaging Biomarker as a Multivariate Descriptor of Health: A Roadmap," Acad. Radiol. 30(2) 159-182 (2023).
[47] Delfino, J.G., Pennello, G.A., Barnhart, H.X., et al., "Multiparametric Quantitative Imaging Biomarkers for Phenotype Classification: A Framework for Development and Validation," Acad. Radiol. 30(2) 183-195 (2023).
[48] Huang, E.P., Pennello, G.A., deSouza, N.M., et al., "Multiparametric Quantitative Imaging in Risk Prediction: Recommendations for Data Acquisition, Technical Performance Assessment, and Model Development and Validation," Acad. Radiol. 30(2) 196-214 (2023).
[49] Wang, X., Pennello, G.A., deSouza, N.M., et al., "Multiparametric Data-driven Imaging Markers: Guidelines for Development, Application and Reporting of Model Outputs in Radiomics," Acad. Radiol. 30(2) 215-229 (2023).

[50] Badano, A., Graff, C.G., Badal, A., Sharma, D., Zeng, R., Samuelson, F.W., Glick, S., Myers, K.J., "Evaluation of Digital Breast Tomosynthesis as Replacement of Full-Field Digital Mammography Using an In Silico Imaging Trial," JAMA Netw Open 1(7):e185474 (2018).
[51] GitHub - DIDSR/VICTRE: Virtual Imaging Clinical Trial for Regulatory Evaluation, (accessed 3 March 2023).
[52] Wang, G., Badal, A., Jia, X., Maltz, J.S., Mueller, K., Myers, K.J., Niu, C., Vannier, M., Yan, P., Yu, Z., Zeng, R., "Development of Metaverse for Intelligent Healthcare," Nat. Mach. Intell. 4, 922-929 (2022).
[53] www.midrc.org (accessed 3 March 2023).
[54] Drukker, K., Sahiner, B., Hu, T., Kim, G.H., Whitney, H.M., Baughan, N., Myers, K.J., Giger, M.L., McNitt-Gray, M. "Assistance tools for the evaluation of machine learning algorithm performance: the decision tree based tools developed by the Medical Imaging and Data Resource Center (MIDRC) Technology Development Project (TDP) 3c effort," Proc. SPIE 12467-12 (2023).

---

*Submit correspondence to: drkylejmyers@gmail.com