POINT/COUNTERPOINT

Suggestions for topics suitable for these Point/Counterpoint debates should be addressed to Colin G. Orton, Professor Emeritus, Wayne State University, Detroit: ortonc@comcast.net. Persons participating in Point/Counterpoint discussions are selected for their knowledge and communicative skill. Their positions for or against a proposition may or may not reflect their personal opinions or the positions of their employers.

Machine learning will transform radiology significantly within the next 5 years

Ge Wang, Ph.D.
Biomedical Imaging Center, Center for Biotechnology & Interdisciplinary Studies, Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
(Tel: 518 276 3726; E-mail: wangg6@rpi.edu)

Mannudeep Kalra, M.D.
Division of Thoracic and Cardiovascular Imaging, MGH Webster Center for Quality and Safety, Department of Radiology, Massachusetts General Hospital, Boston, MA 02114, USA
(Tel: 617 643 4583; E-mail: mkalra@mgh.harvard.edu)

Colin G. Orton, Ph.D., Moderator

(Received 1 March 2017; accepted for publication 2 March 2017; published 20 April 2017)

[https://doi.org/10.1002/mp.12204]

OVERVIEW

With impressive progress in machine learning, there has been increasingly more interest in its relevance to medical physics, which involves both medical imaging and radiation treatment planning. However, because it is still generally unclear how to identify unique niches, utilize big data, and optimize neural networks, machine learning is yet to have a major impact on medical physics practice. Nevertheless, there are optimistic opinions that machine learning will have a major impact on medical physics and radiology within the next 5 years. This is the premise debated in this month's Point/Counterpoint.

Arguing for the Proposition is Ge Wang, Ph.D. Dr. Wang received his Ph.D. in Electrical & Computer Engineering from the University at Buffalo in 1992. Upon graduation, he assumed a junior faculty position at the Mallinckrodt Institute of Radiology, Washington University School of Medicine. In 1997, he joined the University of Iowa as Associate Professor, then in 2006 Virginia Tech as Pritchard Professor, and in 2013 he moved to his current position as Clark & Crossan Endowed Chair Professor at the Rensselaer Polytechnic Institute, where he leads the Biomedical Imaging Center. He has made original contributions to spiral/helical cone-beam/multislice CT, bioluminescence tomography, interior tomography, energy-sensitive CT, and multimodality fusion, and has authored or co-authored over 400 peer-reviewed journal papers. Dr. Wang is a member of the *Medical Physics* Board of Associate Editors and is a Fellow of the AAPM, AIMBE, OSA, SPIE, IEEE, and AAAS.

Arguing against the Proposition is Mannudeep Kalra, M.D. Dr. Kalra completed his diagnostic radiology residency in India in 1999. He subsequently became a research fellow at the Massachusetts General Hospital (MGH) and then at Emory University Hospital. In 2005, he started a clinical fellowship in the thoracic and cardiac imaging sections and then joined as an attending radiologist at MGH. He has authored or co-authored over 300 peer-reviewed journal articles and book chapters. He has co-edited five textbooks and special journal issues in radiology. Currently, he is Associate Professor of Radiology with the Harvard Medical School and Director of the Webster Center for Quality and Safety. His interests include imaging technology assessment, radiation dose optimization, and deep learning applications in radiology.

FOR THE PROPOSITION: Ge Wang, Ph.D.

Opening statement

I view machine learning as a truly disruptive technology, or more accurately a *paradigm shift*,[1] and believe that it has *transformative* potential in the medical physics field, valid for both medical imaging and treatment planning. Clearly, the interest in machine learning seems much greater than that in compressed sensing, as evidenced by my quick PubMed search for the title to contain "*machine learning*" and "*compressed sensing*", respectively. The number of hits for "*machine learning*" has increased from 151 to 450 over the past 5 years, while the number for "*compressed sensing*" has only gone from 84 to 102. Given the successes of machine learning in other areas, I have little doubt that machine intelligence will reshape medical physics, and more generally radiology, and we should immediately make major efforts toward this direction.

Intelligence is essentially the capability to extract knowledge that allows comprehension and prediction, which can be in most cases performed *computationally*. When data are becoming diversified and explosive in either medical imaging or radiotherapy, the classic methods cannot model and utilize huge data effectively and efficiently. It seems that big data and deep learning promise numerous opportunities for medical physicists. Instead of trying to enumerating all the possibilities, without loss of generality, let me discuss this *transformative* approach as related to two *transforms*: the Radon transform (from an underlying image to its projections) and radiation treatment planning (from a source distribution to therapeutic beam profiles).

This year is for the first centenary celebration of the Radon transform, which is fundamental to not only CT but also other tomographic modalities. In practice, Radon data are never ideal; for example, in x-ray imaging, projections are compromised by source spot size, beam hardening, detector imperfection, geometric mismatch, patient motion, metal artifacts, photon fluctuations, and other factors. Over the past decades, excellent analytic and iterative reconstruction methods have been developed. However, the assumed data model is only approximate and compromises image quality; for example, it is challenging to convert photon-counting data into linear integrals, especially when radiation dose is low. In this aspect, image quality can be potentially improved via a deep neural network. This is to perform the Radon transform via machine learning, a freshly new way to recharge the existing reconstruction algorithms for more quantitative results.

The optimization of a therapeutic plan needs to ensure tumor killing while sparing healthy/sensitive tissues for the best prognosis.[2–7] In this context, there is a critical need for a high-quality predictive model which integrates a huge amount of heterogeneous data via machine learning,[8,9] including electronic health records, tomographic and therapeutic images, and genomic profiles. Tomographic images can be improved via machine learning to reduce metal artifacts,[10] estimate an attenuation background,[11] target tumors,[12] and so on. Hence, the potential of machine learning ought to be significant for radiotherapy. We expect that the ultimate therapeutic system will be able to reconstruct images and design plans with high confidence, and keep learning from huge, distributed, and living data sources.

AGAINST THE PROPOSITION: Mannudeep Kalra, M.D.

Opening statement

If a big company's AI makes a mistake, it might get sued for a billion-dollars! A skeptical radiologist might opine that malpractice is the most important issue with artificial intelligence (AI) replacing him over the next few years. This concern is accentuated by unexpected and undetectable behavior of "*black box*" deep learning techniques in a multidimensional feature space. Errors with AI can result when confounding factors are correlated with pathologic entities in the training datasets rather than true signs of diseases. In fact, applications so far have been limited to low level, narrow task-specific pursuits such as detection of pulmonary nodules, rather than the spectra of abnormalities found in the real clinical environment with unbound, unstructured inputs from multiple scanners, entities, and institutions. This might take more than a few years to validate and then gain acceptance. Yet, patients tend to trust the decisions of physicians but question the diagnoses made by a machine.[13]

Doubts have been raised on aspirations of AI to unseat human radiologists.[14,15] If neural networks have high generalization performance, why should adversarial negatives of regular examples confound them?[15] While human vision could not delineate subtle changes engineered in test images, blind spots in neural networks led to several misclassifications including a dog labeled as an ostrich!

Deep neural networks expressed 99% overconfidence for classifying unrecognizable images such as labeling of a red crayon as a syringe.[14] "*Fooling*" of these networks raises questions about their true generalization capabilities in face of tremendous biological and physical variations in patients and imaging modalities. Dr. Bryan observed that variations between intersite and multivendor measurements limited AI applications for cerebral blood flow imaging techniques in Alzheimer Disease.[16] Such variations can be robustly normalized in human vision but need considerable advances in deep learning to avoid dangers from underappreciated and underrepresented statistical errors.

Alternative use of large, nonhomogeneous data with flexible learning algorithms is challenged by the general lack of annotated imaging data for training. Manual segmentation is severely limited by human resources and inability to demarcate diffuse or heterogeneous abnormalities. Multiple instance learning can overcome certain aspects of weakly labeled image database[17,18] but still requires standardized labels of specific diagnoses, which are often not available. Natural language processing to parse radiology reports requires a full syntactic parser trained on radiology reports. This is immature, partially due to lack of integration between radiology findings and clinical, pathology and laboratory results from claims and electronic medical records databases.[19,20]

Other challenges include the cost of producing labeled datasets not confined to single diagnostic entities and that of time-consuming, intensive computation requiring in depth know-how of graphics processing units, and systematic rigorous cross-validation for clinical acceptance of machine learning. Radiology extends beyond medical physics to interpretation of radiology findings and correlation with clinical context. AI can help medical physics but its ability to replace radiologists in the context of interpretation of radiology findings and correlation with clinical and laboratory findings is unlikely within the next 5 years.

Rebuttal: Ge Wang, Ph.D.

I agree with most of what my opponent has said. In principle, all the challenges can be met over time, but how soon will machine learning plays a significant role in hospitals and clinics? I would imagine that it could be as soon as within the next 5 years. This view is based on my resonance to the prophecy that the singularity of artificial intelligence is near.[21] Many of us share a feeling that the scientific advancement is at an accelerated rate due to the combinatory effect of knowledge and tools, as demonstrated by the data fitting into the Moore's law as well as the surge of machine learning and high-performance computing (including quantum computing) research. Now, machine intelligence has competed with, or already outperformed, humans in a number of tasks, such as chess playing, image classification, and speech recognition. Hence, the efforts along this direction are well justified in medical imaging, therapeutic planning, and beyond.

Two examples are supportive of my optimism. The first is the software *Master* (an upgrade of *AlphaGo*) that recently defeated the world's best Go players in several dozen games in a row. Not long ago, when *AlphaGo* won over Lee Sedol 4-1, Ke Jie watched and claimed that "*it can't beat me*." Recently, however, Ke lost three games to *Master*. The team *DeepMind* behind *Master* is actively working on machine learning methods for other applications, including healthcare, and so are many other teams including ours. In January, 2017, Nature reported that a machine learning algorithm developed at Stanford performed on par with 21 board-certified dermatologists in the diagnosis of skin cancer.[22] Their single neural network, which was trained on a dataset of 129,450 clinical images consisting of 2032 different diseases, clearly showed potential for highly variable tasks across many fine-grained object categories. They pointed out that "*Outfitted with deep neural networks, mobile devices can potentially extend the reach of dermatologists outside of the clinic*." I feel confident that machine learning would impact radiology similarly and quite soon.

Rebuttal: Mannudeep Kalra, M.D.

Dr. Wang makes several claims in favor of machine learning. Without being a procrastinator, I cite the following contrary arguments. First, the opinion in Forbes on accidents involving self-driving cars raised the probability of carmakers getting sued for hefty fines.[23] Such libels will stifle progress, and our legal system is underprepared for a fully autonomous AI driver or a machine radiologist in the current context. While car accident lawsuits are the most common type of personal injury claims, medical malpractice suits are among the most complex ones!

Second, the FDA approved computer-aided diagnosis (CAD) for mammography in 1998 based on its comparable performance with, or outperformance of, human observers. Over the past two decades, CAD programs remain relegated to being a second reader without any exception in clinical radiology practices around the world! Are these not the very same programs that AI hopes to improvise based on training data labeled by human observers?

Third, applications of AI in the physics domain (such as image reconstruction and equipment calibration) might be ripe opportunities but, in clinical practice, AI will face challenges to experts at work on creating the most intelligent deep learning algorithms. Such nontrivial challenges stem from the lack of decent theories and labeled datasets, and validation of AI algorithms against dissenting human radiologists and its acceptance among radiologists, ordering physicians, and the patients. Like existing CAD programs, AI algorithms should also undergo a prolonged phase of enquiry and verification in clinical practice, a task without trims or short cuts. As a second reader, AI will learn from humans while helping them in return to take better care of their patients.

Finally, AI fares pretty well on "low hanging" targets of sharply defined skin cancers in colorful 2D photographs[22] but will face challenges from 3D gray scale, fuzzy radiology images where lesions are often subtle or diffuse, differentials are wider, and artifacts masquerade.

CONFLICTS OF INTEREST

We have no conflicts of interest to disclose.

REFERENCES

1. Wang G. Perspective on deep imaging. *IEEE Access*. 2016;4:8914–8924.
2. Kazhdan M, McNutt T, Taylor R, Wu B, Simari P. Comment on "A planning quality evaluation tool for prostate adaptive IMRT based on machine learning". *Med Phys*. 2011;38:2820.
3. Zhu X, Ge Y, Li T, Thongphiew D, Yin F-F, Wu QJ. A planning quality evaluation tool for prostate adaptive IMRT based on machine learning. *Med Phys*. 2011;38:719–726.
4. Arimura H, Jin Z, Shioyama Y, Nakamura K, Magome T, Sasaki M. Automated method for extraction of lung tumors using a machine learning classifier with knowledge of radiation oncologists on data sets of planning CT and FDG-PET/CT images. *Conf Proc IEEE Eng Med Biol Soc*. 2013;2013:2988–2991.
5. Guidi G, Maffei N, Meduri B, et al. A machine learning tool for re-planning and adaptive RT: a multicenter cohort investigation. *Physica Med*. 2016;32:1659–1666.
6. Buus S, Rylander S, Hokland SL, Bentzen L. Learning curve of MRI-based planning for high-dose-rate brachytherapy for prostate cancer. *Brachyther*. 2016;15:426–434.
7. Ikushima K, Arimura H, Jin Z. Computer-assisted framework for machine-learning-based delineation of GTV regions on datasets of planning CT and PET/CT images. *J Radiat Res*. 2017;58:123–134.

8. El Naqa I, Bradley JD, Lindsay PE, Hope AJ, Deasy JO. Predicting radiotherapy outcomes using statistical learning techniques. *Phys Med Biol*. 2009;54:S9–S30.
9. Bibault JE, Giraud P, Burgun A. Big data and machine learning in radiation oncology: state of the art and future prospects. *Cancer Lett*. 2016;382:110–117.
10. Gjesteby L, De Man B, Yanyan J, et al. Metal artifact reduction in CT: where are we after four decades? *IEEE Access*. 2016;4:5826–5849.
11. Nie D, Cao X, Gao Y, Wang L, Shen D. Estimating CT image from MRI data using 3D fully convolutional networks. In: Carneiro G, ed. *Deep Learning and Data Labeling for Medical Applications*. Springer International Publishing; 2016:170–178.
12. Hu P, Wu F, Peng J, Liang P, Kong D. Automatic 3D liver segmentation based on deep learning and globally optimized surface evolution. *Phys Med Biol*. 2016;61:8676–8698.
13. Chockley K, Emanuel E. The end of radiology? Three threats to the future practice of radiology. *J Am Coll Radiol*. 2016;13:1415–1420.
14. Nguyen A, Yosinski J, Clune J. Deep neural networks are easily fooled: high confidence predictions for unrecognizable images. In: *Computer Vision and Pattern Recognition (presented in Computer Vision and Pattern Recognition (CVPR) 2015)*. Boston, MA: IEEE; 2015 (Accessed on 1.21.2017 at http://www.evolvingai.org/files/DNNsEasilyFooled_cvpr15.pdf)
15. Szegedy C, Zaremba W, Sutskever I, et al. *Intriguing properties of neural networks*. arXiv preprint arXiv 2013:1312.6199.
16. Bryan NB. Machine learning applied to Alzheimer disease. *Radiol*. 2016;281:665–668.
17. Melendez J, van Ginneken B, Maduskar P, et al. A novel multiple-instance learning-based approach to computer-aided detection of tuberculosis on chest X-rays. *IEEE Trans. Med. Imag*. 2015;34:179–192.
18. Sørensen L, Nielsen M, Lo P, Ashraf H, Pedersen JH, de Bruijne M. Texture-based analysis of COPD: a data-driven approach. *IEEE Trans Med Imag*. 2012;31:70–78.
19. Rink B, Roberts K, Harabagiu S, et al. Extracting actionable findings of appendicitis from radiology reports using natural language processing. *AMIA Jt Summits Transl Sci Proc*. 2013;2013:221–238.
20. Crown WH. Potential application of machine learning in health outcomes research and some statistical cautions. *Value in Health*. 2015;18:137–140.
21. Kurzweil R. *The Singularity is Near: When Humans Transcend Biology*. New York: Viking; 2005.
22. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115–118.
23. Ben-Shahar O. *Should carmakers be liable when a self-driving car crashes?* Forbes, September 22, 2016. http://www.forbes.com/sites/omribenshahar/2016/09/22/should-carmakers-be-liable-when-a-self-driving-car-crashes/#67c8dcec1f40. Accessed February 3, 2017.