Will Artificial Intelligence Replace the Human Echocardiographer?:

Clinical Considerations

Partho P. Sengupta, MD, DM and **Donald A. Adjeroh, PhD**

Division of Cardiology, West Virginia University Heart and Vascular Institute, School of Medicine (P.P.S.) and Lane Department of Computer Science and Electrical Engineering (D.A.A.), West Virginia University, Morgantown.

Keywords

An older population with an increased prevalence of cardiovascular disease and an aging workforce are engendering a state of healthcare crisis in cardiology.[1] Most cardiologists now face an unprecedented time crunch as they rush through their appointments to perform and interpret more and more procedures. The need to multitask creates exhaustion leading to burnout and frequent reporting errors.[2] The recent interest in using artificial intelligence techniques, such as machine learning, may offer a solution to reduce physician workload including repetitive and tedious tasks involved in diagnosing and analyzing patient data and imaging. To this end, the study by Zhang and colleagues[3] in this issue of *Circulation* adds to the growing enthusiasm for developing a machine learning algorithm that automates several facets of echocardiography measurement and interpretation.

Zhang and colleagues used a deep learning model that has enjoyed spectacular success in addressing computer vision problems including image classification, face recognition, robot navigation, and driverless cars to name a few. Although traditional machine learning workflow includes an initial stage of feature engineering and selection from the data for classification, deep learning methods attempt to learn the important features directly from the raw image data (with minimal preprocessing). Zhang and colleagues applied an algorithm that has triumphed in image recognition tasks and reported a 96% accuracy for distinguishing between broad echocardiographic view classes (eg, parasternal long axis from short axis, or an apical view) and an 84% accuracy overall (including partially obscured views). These results are consistent with a recent study that applied deep learning with convolutional neural networks for view classification of echocardiograms.[4] However, Zhang et al notably used a deeper architecture with more layers (18 versus 11), considered a larger

Correspondence: Partho P. Sengupta, MD, DM, Heart and Vascular Institute, West Virginia University, 1 Medical Center Drive, Morgantown, WV 26506. partho.sengupta@wvumedicine.org.

number of echocardiography view classes (23 versus 15 views), and applied their technique to a larger data set (14 035 versus 267 echocardiographic studies).

The authors also reported an overall metric of accuracy of image segmentation ranging from 72% to 90% for image segmentation. Although deep learning has been explored previously for segmenting the left ventricle,[5] the work by Zhang et al was much more extensive with additional cardiac segments beyond the left ventricle and a larger data set, involving millions of images from 14 035 studies. Moreover, the authors succeeded in going a step beyond simple classification and segmentation by providing an algorithm for automated quantification of cardiac structure and function. The comparison with manually recorded measurements, however, showed wide limits of agreements emphasizing the potential real-world variability of echocardiography measurements. Independent verification from a core laboratory or the use of a gold standard like cardiac magnetic resonance was not reported and could perhaps help us understand the accuracy and precision of the new technology. Nevertheless, automated view detection, segmentation, and measurements are important advances that have been made increasingly available in recent years.[6]

Finally, the authors assessed the ability to perform disease detection essentially based on the results of the automated quantification. With the output restricted to 2 classes, disease and control, authors were able to identify 3 cardiovascular diseases with impressive results: area under the receiver operating characteristic curve of 0.93 for hypertrophic cardiomyopathy, 0.87 for cardiac amyloidosis, and 0.85 for pulmonary arterial hypertension. These remarkable results expand the observations made in several previous studies where machine learning was touted to have abilities in detecting different cardiac phenotypes.[7–10]

Transthoracic echocardiographic imaging is the most commonly performed noninvasive cardiac procedure. Image quality, however, varies substantially between patients and is also operator dependent, which increases interobserver variability. Consequently, despite the wide utilization, errors in echocardiography quantification and interpretation are prevalent and often associated with conflicting interpretations in echocardiography reports.[11] In the majority of instances, such errors may be linked to physician fatigue, impaired attention, memory, and executive function that decrease reader recall and attention to detail. There is little doubt that machine learning techniques like the one illustrated by Zhang and colleagues[3] would go a long way for better data organization, improving workflows, and reducing echocardiography measurement and cognitive errors. Moreover, these technologies may offer solutions for limited training opportunities and the lack of expert supervision. Undoubtedly, deep learning has the potential of improving clinical workflow and diagnostic efficacy in echocardiography. It is not surprising, as the figure shows, that there has been a recent surge in interest in the application of artificial intelligence techniques in medicine and, in particular, in echocardiography. The figure shows that machine learning and deep learning techniques are rapidly emerging as the preferred approach over traditional methods that use feature engineering. However, the question that is looming currently is whether deep learning techniques could eventually supersede an echocardiographer? This question pleads for an understanding of the current state of deep learning and artificial intelligence technologies in relation to human intelligence. The strengths and limitations of the current deep learning techniques have been recently debated.[12] First, the bio-inspired neural

network design used currently for deep learning remotely resembles the activities of a human brain. Structurally, they resemble at best the outer layers of the retina or the visual cortex where images are just sensed or represented. These layers of neurons can hardly do what our brains are capable of, such as the reasoning and knowledge components represented in the more open terrains of general intelligence, that form a foundation of pervasive common-sense knowledge in humans. There are several other innate virtues of human cognition. Human learning is distinguished by its richness and its efficiency; a child learns rapidly to ambulate after a few falls. Moreover, humans can use a learned model to orchestrate action sequences that maximize future reward. However, a deep learning algorithm would require thousand-to-millions of attempts at learning before it can accurately relegate an image! Thus deep learning, although widely publicized for its amazing performance, is actually quite shallow in its intelligence. Ironically enough, the deep in deep learning refers to the architecture of the neural network and not the concept of learning as has been well highlighted in a recent treatise by Gary Marcus.[12]

From the foregoing account, it is clear that complex decisions like determining the appropriateness of a test or the mechanism to weigh and extract information from study in the clinical context would be quite arduous for current computational algorithms. A physician uses his tacit and codified knowledge of cardiac physiology to target measurements in end-systole and end-diastole and integrates these measurements in a mental model. A deep learning technique may not be programmed to extrapolate the knowledge of cardiac physiology, but rather depend on mundane features that may not make intuitive sense beyond the data sets. It is astounding how expeditiously physicians learn incipient pathologies and disease presentations. We develop a hierarchical pattern of new knowledge representation based on association with previous knowledge. This allows us to quickly identify the uniqueness of a new case. Deep learning techniques currently cannot relegate such cognizance hierarchically; most correlations between sets of features are flat or nonhierarchical, as if in a simple, unstructured list, with every feature often treated with similar priority.[12] Thus, unlike human intelligence, deep learning cannot yet integrate prior knowledge effectively (at least, in the same way as a human), and in the absence of large data sets, it would be difficult to train the network to correctly identify rare anomalies. That said, it is worth considering that the most appropriate approach used by machine leaning techniques to reach their decisions may not always follow the same steps that are often used by humans. Clearly, this is an area where cross talk between new cognitive, unsupervised, deep learning, and reinforcement learning approaches may be especially paramount for future development.

In summary, the work presented by Zhang and colleagues represents a ladder in the right direction; however, it is still a long journey ahead. In the current time, the best use of the technology would be to free up time for physicians from repetitive low-level and uneventful activities like measurements, data preparation, standardization, and quality control to more direct time spent in higher calibers of interpretation, patient care, and medical decision making. This will perhaps allow physicians to be more interactive and experiential in answering diagnostic questions, and communicating the findings optimally to teams, a key ingredient of good patient care. Moreover, such strategies can embolden the patient-doctor relationships and overcome any engineered approaches that carry the risks of dehumanizing

care. A well-functioning patient-physician encounter is an essential part of healing, in particular, for chronic disorders where the skills of physicians and diagnostic tests can influence patients' objective and subjective measures of well-being.[13,14] Perhaps by efficaciously integrating the automated measurements and interpretations, the human physician will continue to do what machines are incapable of (or at least, not yet good at), bring higher wisdom in personalized care, foster patient engagement, and motivate patients for much needed chronic lifestyle behavior changes.

Disclosures

REFERENCES

1. Narang A, Sinha SS, Rajagopalan B, Ijioma NN, Jayaram N, Kithcart AP, Tanguturi VK, Cullen MW. The supply and demand of the cardiovascular workforce: striking the right balance. J Am Coll Cardiol 2016;68:1680–1689. doi: 10.1016/j.jacc.2016.06.070 [PubMed: 27712782]

2. Michel JB, Sangha DM, Erwin JP, 3rd. Burnout among cardiologists. Am J Cardiol 2017;119:938–940. doi: 10.1016/j.amjcard.2016.11.052 [PubMed: 28065490]

3. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, Lassen MH, Fan E, Aras MA, Jordan C, Fleischmann KE, Melisko M, Qasim A, Shah SJ, Bajcsy R, Deo RC. Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy. Circulation 2018;138:1623–1635. doi: 10.1161/CIRCULATIONAHA.118.034338 [PubMed: 30354459]

4. Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classification of echocardiograms using deep learning. npj Digit Med 2018;1:6. [PubMed: 30828647]

5. Smistad E, Ostvik A, Haugen BO, Lovstakken L. 2D left ventricle segmentation using deep learning. IEEE Int Ultrason Symp IUS 2017;4–7. doi: 10.1109/ULTSYM.2017.8092812. https://ieeexplore.ieee.org/document/8092812. Accessed Aug 14, 2018.

6. Knackstedt C, Bekkers SC, Schummers G, Schreckenberg M, Muraru D, Badano LP, Franke A, Bavishi C, Omar AM, Sengupta PP. Fully automated versus standard tracking of left ventricular ejection fraction and longitudinal strain: the FAST-EFs multicenter study. J Am Coll Cardiol 2015;66:1456–1466. doi: 10.1016/j.jacc.2015.07.052 [PubMed: 26403342]

7. Narula S, Shameer K, Salem Omar AM, Dudley JT, Sengupta PP. Machine-learning algorithms to automate morphological and functional assessments in 2D echocardiography. J Am Coll Cardiol 2016;68:2287–2295. doi: 10.1016/j.jacc.2016.08.062 [PubMed: 27884247]

8. Omar AMS, Narula S, Abdel Rahman MA, Pedrizzetti G, Raslan H, Rifaie O, Narula J, Sengupta PP. Precision phenotyping in heart failure and pattern clustering of ultrasound data for the assessment of diastolic dysfunction. JACC Cardiovasc Imaging 2017;10:1291–1303. doi: 10.1016/j.jcmg.2016.10.012 [PubMed: 28109936]

9. Sengupta PP, Huang YM, Bansal M, Ashrafi A, Fisher M, Shameer K, Gall W, Dudley JT. Cognitive machine-learning algorithm for cardiac imaging; a pilot study for differentiating constrictive pericarditis from restrictive cardiomyopathy. Circ Cardiovasc Imaging 2016;9:e004330. doi: 10.1161/CIRCIMAGING.115.004330 [PubMed: 27266599]

10. Sanchez-Martinez S, Duchateau N, Erdei T, Fraser AG, Bijnens BH, Pi-ella G. Characterization of myocardial motion patterns by unsupervised multiple kernel learning. Med Image Anal 2017;35:70–82. doi: 10.1016/j.media.2016.06.007 [PubMed: 27322071]

11. Spencer KT, Arling B, Sevenster M, DeCara JM, Lang RM, Ward RP, O'Connor AM, Patel AR. Identifying errors and inconsistencies in real time while using facilitated echocardiographic reporting. J Am Soc Echocardiogr 2015;28:88–92.e1. doi: 10.1016/j.echo.2014.09.005 [PubMed: 25441328]

12. Marcus G Deep learning: a critical appraisal. arXiv:1801.00631 2018;1–27. http://arxiv.org/abs/1801.00631. Accessed Aug 14, 2018.

13. Canale S Del, Louis DZ, Maio V, Wang X, Rossi G, Hojat M, Gonnella JS. The relationship between physician empathy and disease complications: an empirical study of primary care physicians and their diabetic patients in Parma, Italy. Acad Med 2012;87:1243–1249. [PubMed: 22836852]

14. Howard ZD, Noble VE, Marill KA, Sajed D, Rodrigues M, Bertuzzi B, Liteplo AS. Bedside ultrasound maximizes patient satisfaction. J Emerg Med 2014;46:46–53. doi: 10.1016/j.jemermed.2013.05.044 [PubMed: 23942153]
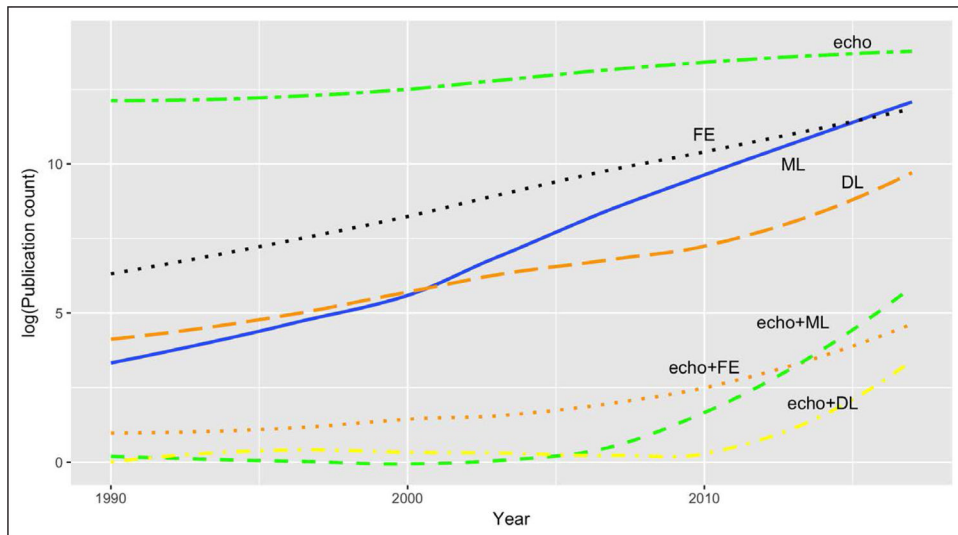
Figure. Trends in artificial intelligence techniques in echocardiography based on publications in PubMed.

echo indicates "echocardiography" or "echocardiogram" or "cardiac ultrasound"; FE, "feature extraction" or "feature engineering"; ML, "machine learning"; DL, "deep learning"; "+," logical AND. PubMed publication counts obtained using the EUtilsSummary() function in the R package "RISmed" (https://CRAN.R-project.org/package=RISmed).