Clinical Implications and Challenges of Artificial Intelligence and Deep Learning

William W. Stead, MD

Artificial intelligence (AI) and deep learning are entering the mainstream of clinical medicine. For example, in December 2016, Gulshan et al. reported development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. An accompanying editorial by Wong and Bressler pointed out limits of the study, the need for further validation of the algorithm in different populations, and unresolved challenges (eg, incorporating the algorithm into clinical work flows and convincing clinicians and patients to “trust a ‘black box’”). Sixteen months later, the Food and Drug Administration (FDA) permitted marketing of the first medical device to use AI to detect diabetic retinopathy. FDA reduced the risk of releasing the device by limiting the indication for use to screening adults who do not have visual symptoms for greater than mild retinopathy, to refer them to an eye care specialist.

This issue of JAMA contains 2 Viewpoints on deep learning in health care. Hinton explains the technology underlying AI and deep learning, using clinical examples. AI is the general term for imitating human intelligence with computer systems. Early AI systems represented human reasoning with symbolic logic. As computer processing and storage became more powerful, researchers developed machine-learning techniques to imitate the way the human brain learns. The first machine learning continued to rely on human experts to label the data the system trained on (eg, the diagnosis) and to identify the significant features (eg, findings). Machine learning weighted the features from the data. With continued advances in computational power and with larger data sets, researchers began to develop deep learning techniques. The first deep learning algorithms were “supervised” in that human experts continued to label the training data, and the deep learning algorithms learned the features and weights directly from the data. The retinopathy screening algorithms are an example of supervised deep learning. Hinton describes continuing development of new deep learning techniques, including ones that are completely unsupervised. He also points out that it is not feasible to see the features learned by deep learning to explain how the system reaches a conclusion.

Naylor identifies 7 factors driving adoption of AI and deep learning in health care: (1) the strengths of digital imaging over human interpretation; (2) the digitization of health-related records and data sharing; (3) the adaptability of deep learning to analysis of heterogeneous data sets; (4) the capacity of deep learning for hypothesis generation in research; (5) the promise of deep learning to streamline clinical workflows and empower patients; (6) the rapid-diffusion open-source and proprietary deep learning programs; and (7) of the adequacy of today’s basic deep learning technology to deliver improved performance as data sets get larger. Factors 3, 4, and 6 are specific to deep learning; the other factors apply to other AI techniques as well.

Artificial intelligence is a family of technical techniques in the same way the radiologic imaging tool kit includes flat images, computed tomography scans, and functional imaging such as magnetic resonance imaging. Advances in computational technology, computer science, informatics, and statistics improve existing techniques and make new techniques possible. The addition of deep learning to the AI family of techniques represents an advance similar in magnitude to the addition of the computed tomography scanner to the radiology tool kit. Each AI technique has strengths and weaknesses. Symbolic logic is self-explaining but difficult to scale. For example, knowledge engineers extract the logic by interviewing or observing human experts. Statistical techniques such as supervised deep learning scale, but are subject to bias in the training data, and the reasoning cannot be explained. Since deep learning systems are trained on data from the past, they are not prepared to reason in the way humans do about conditions that have not been seen before. In the future, unsupervised deep learning may reduce this gap between human intelligence and AI.

The potential applications of AI in health care present a range of computational difficulty. Narrow tasks, in which the context is predefined, are relatively easy. Image-processing tasks such as recognizing the border of an organ to suggest where to cut off a scan, or highlighting a suspicious area in an image for the radiologist or pathologist, are examples of narrow tasks. Image analysis and diagnostic prediction tasks such as the diabetic retinopathy example are broader and harder, but doable with today’s technology. Very broad data analysis and pattern prediction tasks such as analyzing heterogeneous data sets from diverse sources to suggest novel associations are feasible today because the purpose is limited to hypothesis generation. Thinking in the way humans do—reasoning, for example, from a few observations to suggest a novel scientific framework as Einstein did with the theory of relativity—is beyond technology on the horizon.

Clinicians should view the output of AI programs or devices as statistical predictions. They should maintain an index of suspicion that the prediction may be wrong, just as they...
do when the laboratory reports a high potassium value and the physician orders an electrocardiogram to see if the T wave is peaked before addressing the laboratory value. Clinicians should judge whether the prediction applies to the patient who is in front of them and decide if additional data or expertise is needed to inform that decision. As Verghese and colleagues’ advised, “clinicians should seek a partnership where the machine predicts, and the human explains and decides on action.” Physicians are trained to reason and act under uncertainty. AI does not reduce the importance of this professional competency. On the contrary, AI is best suited to augmenting human intelligence, bringing more data into focus to support human intelligence. Clinicians should take the time to refresh their understanding of statistics and learn about measures such as calibration that are applicable to prediction to prepare to partner with AI.

Clinicians, AI researchers, and developers of AI applications and devices should work together to accelerate progress and to limit adverse consequences of applying AI in health care. Rigorous translation pipelines will be needed to support their work. The first step in the pipeline involves hypothesis generation, such as applying an AI technique to predict a subpopulation at high risk for an adverse outcome. The second step is in silico replication to validate the predictive algorithm with data from other populations and settings. The third step creates a prototype intervention combining the predictive algorithm with information and decision support for health care teams and patients, as with the diabetic retinopathy screening device. The fourth step tests the feasibility and utility of the intervention in a pilot study in a health care setting. The fifth step documents efficacy under clinical trial conditions. The final steps document effectiveness when the AI application is deployed in practice, followed by implementation as standard of care and improvement in patient outcomes. For low-risk interventions, it may be appropriate to skip efficacy trials and proceed directly from usability to effectiveness.

Artificial intelligence and deep learning are entering the mainstream of clinical medicine. This technology can augment human intelligence to improve decision making and operational processes. Physicians need to actively engage to adapt their practice and to shape the technology.

**REFERENCES**


**ARTICLE INFORMATION**

**Author Affiliation:** McKesson Foundation, Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee.

**Corresponding Author:** William W. Stead, MD, Department of Biomedical Informatics, Vanderbilt University Medical Center, 1161 21st Ave S, D1300 Medical Center N, Nashville, TN 37232-2104 (bill.stead@vumc.org).

**Conflict of Interest Disclosures:** The author has completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Stead reports receiving travel reimbursement, honoraria, or both for serving on the Journal Oversight Committee for JAMA; as chair of the technical advisory committee of the Center for Medical Interoperability; on the planning committee for the Computing Research Association; on panels for the Gordon & Betty Moore Foundation; on committees for the Centers for Disease Control and Prevention; on committees for the National Academies of Sciences, Engineering, and Medicine; and on a strategic planning panel for the National Library of Medicine. He also reports receiving stock grants and director fees from HealthStream and is listed as coinventor of an electronic health record product licensed by Vanderbilt to Informatics Corporation of America and on an electronic health record product licensed to McKesson.

**REFERENCES**


