EDITORIAL

Deep Learning Algorithms for Detection of Lymph Node Metastases From Breast Cancer
Helping Artificial Intelligence Be Seen

Jeffrey Alan Golden, MD

Author Video Interview and JAMA Report Video

Related article page 2199

Artificial intelligence (AI), the theory and development of computer systems able to perform tasks that normally require human intelligence, is creeping into almost every facet of modern life. Familiar examples include computer chess games, speech recognition, intelligent routing in content delivery networks, and autonomous driving cars. In the financial sector, AI is routinely used for fraud detection, algorithmic trading, and chatbots (ie, computer programs that appear to conduct conversations via auditory or textual methods, such as with online virtual assistants).

Health care has been slower to adopt AI but the pace of implementation is accelerating at an impressive rate. In 2014, the acquisition of AI startups in health care was about $600 million; in 2021, it is anticipated to be $6.6 billion or a 40% compound annual growth rate.[1] One reason health care is ripe for AI is "big data": the health care industry has rich data sets that are ideal for AI given the requirement for large test sets of data with which the computer can "learn."

Most computer modeling enhancements in health care, particularly in the image analysis field, have focused on feature engineering, essentially asking a computer to evaluate explicit features specified by experts. This permits the algorithms to detect abnormalities or predict specified lesions. In contrast, deep learning is a form of AI that includes machine learning techniques that perform iterative optimization strategies that are based on pixel-by-pixel evaluation of the data from images.[2]

The promise of AI in health care is the delivery of improved quality and safety of care and the potential to democratize expertise. For example, in a study by Esteva et al,[3] the authors compared the ability of a deep convolutional neural network (CNN) to discriminate the most common skin cancers including malignant melanoma. They compared and demonstrated at least equivalence in the performance of their algorithm against 21 board-certified dermatologists in evaluating biopsy-proven clinical images. In this example, AI was used to discriminate whether skin lesions were malignant. The authors suggested that mobile devices, like smartphones, could be deployed with similar algorithms, permitting potentially low-cost universal access to vital diagnostic care anywhere in the world.

In another study, Gulshan et al[4] applied a deep CNN approach to a test set of more than 128 000 retinal fundus images from adult patients with diabetes to identify referable diabetic retinopathy. The algorithm developed had very high sensitivity and specificity for detecting referable diabetic retinopathy and macular edema.[4] This study established a clear path toward use of AI not to replace physicians, but rather to perform simple, cost-effective, and widely available examinations and analyses that could help identify at-risk patients who require referral for specialty care while reassuring other patients that potential retinal manifestations of their diabetes are not present or are stable.

Radiology, having converted to digital images more than 25 years ago, is well-positioned to deploy AI for diagnostics. Several studies have shown considerable opportunity to support radiologists in evaluating a variety of scan types including mammography for breast lesions, computed tomographic scans for pulmonary nodules and infections, and magnetic resonance images for brain tumors including the molecular classification of brain tumors.[5-9]

In contrast to radiology, pathology has been late to adopt digital imaging and thus computer-assisted diagnostic technologies. In part, this is the result of practical and financial obstacles. With conversion to digital images, radiology eliminated film, chemicals, developers, and storage of the films. Radiology departments also solved problems related to loss of films and transport of films to where they are needed, for example, in operating rooms, emergency departments, and intensive care units. Unforeseen at the time, although anticipated by some, was the inherent value within these images for greater learning using computers to improve the quality, safety, and efficiency of radiologists.

Many, if not most, of the practical benefits realized by radiology would not be achieved with pathology digitization. An anatomic pathology workflow that includes digital pathology will not reduce or remove the need to produce and ultimately store glass slides of pathology specimens. Instead of any reductions, digital pathology will require additional workflows, personnel, equipment, and importantly storage of data (it is estimated that digital pathology images are at least 10 times larger files than radiology images), all on top of an already financially and operationally stressed health care system. Certainly, the adoption of digital pathology will bring some advantages, particularly in areas such as rapid teleconsultations with experts and in quality and safety. Nonetheless, widespread adoption of digital pathology will require a defined value proposition that has been slow to materialize.

Another challenge to deploying digital pathology was recently addressed. In April 2017, Philips received US Food and Drug Administration clearance for its Philips IntelliSite Pathology Solution to be used for primary pathology diagnostics. This device is used for scanning glass pathology slides and for reviewing these slides on computer monitors. Furthermore, the Philips IntelliSite Pathology Solution has been established as a predicate device that could pave the way for a host of other whole-slide scanners available today to use a 510(k) process for approval rather than a premarket analysis. Many new Food and Drug Administration–approved scanners for primary diagnosis are expected to become available in the coming years.

The emergence of AI in health care, the reduced costs of digital data, and the availability of usable digital images are now in alignment for digital pathology to succeed. In this issue of *JAMA*, Ehteshami Bejnordi and colleagues[10] report the results of an investigation developed in response to an international contest to have a machine detect sentinel lymph node metastases of breast cancer. The CAMELYON16 grand challenge (Cancer Metastases in Lymph Nodes Challenge) was organized in collaboration with the Institute of Electrical and Electronics Engineers' International Symposium on Biomedical Imaging.[11] Two hundred and seventy hematoxylin and eosin–stained whole-slide images of sentinel lymph nodes (110 with and 160 without nodal metastases) validated with immunohistochemical staining, were provided to the 390 entrants to build their algorithms along with an independent test set of 129 whole-slide images (49 with and 80 without nodal metastases), for which the actual diagnosis (ie, "ground truth") was blinded. A total of 23 teams submitted 32 methods for evaluation, nearly 80% using a deep CNN method.[10] While several teams submitted methods other than deep learning–based algorithms, the deep CNN methods performed significantly better. At least the top 5 algorithms performed as well as pathologists, if not slightly better, with several caveats.

In the study by Ehteshami Bejnordi et al,[10] 2 approaches were taken in comparing the performance of the algorithms generated by automated deep learning systems to evaluation by pathologists for detection of nodal metastases in the whole-slide images. The first method involved a panel of 11 pathologists with varying degrees of expertise in breast pathology, who were given 2 hours to review all 129 test slides, less than 1 minute per slide. Two problems exist with this method; this is an unrealistically short period of time for evaluation of 129 slides and, in routine practice, pathologists are unlikely to review 129 consecutive sentinel nodes and they will request additional sections or special stains in questionable cases. The second method was to ask one pathologist to review all cases without a limit of time. This person took 30 hours to review all cases and, not surprisingly, scored better than the group with limited time. While no perfect control exists, it is clear the comparisons made in this study have limitations.

Confidence in the algorithms comes from their ability to detect metastases. The top algorithms performed better than the 11 pathologists with time constraint at identifying micrometastases (tumor cell cluster diameter, 0.2 to <2 mm) (area under the curve for the best algorithm, 0.994; 95% CI, 0.983-0.999 vs mean area under the curve for the 11 pathologists, 0.810; range, 0.738-0.884; $P < .001$) but were not statistically different when compared with the pathologist with unlimited time (area under the curve, 0.943).[10] Although micrometastases are currently being evaluated for their clinical importance, the fact that the algorithms detected these abnormalities at the same rate or better than pathologists is exciting. However, it is unclear that the algorithms were equally effective at detecting all types of breast cancer. While the algorithms were effective in detecting invasive ductal carcinoma, lobular carcinoma is notoriously difficult to identify and easily missed. The algorithms were better than the 11 pathologists with time limits at detecting noninvasive ductal carcinoma. However, the data for comparison with the one pathologist without time limits suggest that, given adequate time, the pathologist does as well or better than the algorithms for detecting noninvasive ductal carcinoma.

The CAMELYON16 challenge highlights a significant opportunity for AI in pathology, namely assisting pathologists with screening for lesions in histopathologic sections. Image analysis has also aided screening in other areas of pathology, cytology being an excellent example.[12] Another area of pathology in which AI is emerging as a potential supplement to the pathologist is to identify high- and low-risk lesions. As early as 2006, work using computerized morphometry determined that valid tools could be built to grade dysplasia in Barrett esophagus and predicting progression to adenocarcinoma.[13] Subsequently, work using morphology feature extraction demonstrated that the stroma was an important indicator of survival in neuroblastoma and breast cancer.[14,15]

Yu et al[16] recently designed a fully automated informatics pipeline to extract objective quantitative image features and build classifiers to distinguish lung cancer with different survival outcomes. Using 2186 whole-slide images from The Cancer Genome Atlas database as a training set and another 294 images from a tissue microarray as a test set, the authors showed that these automated classifier algorithms could reliably differentiate longer-term survivors from short-term survivors in a cohort with stage I adenocarcinoma. Their deep learning method was able to better predict clinical outcome than either clinical stage or histopathologic grade. Notably, their method achieved similar results with squamous cell carcinoma of the lung. Recognizing this type of method will need to be independently validated using unbiased images not selected from The Cancer Genome Atlas, the results still indicate AI can be used to supplement pathologic diagnoses to provide prognostic information.

A major unresolved issue is how AI will be implemented in routine clinical practice. Numerous intertwined issues will have to be addressed to overcome several significant obstacles. The first will be creating the value proposition in pathology. Given that digital pathology is likely to be more costly owing to additional workflow, including personnel, AI must demonstrate improved efficiency, quality, and safety. The study by Ehteshami Bejnordi et al[10] begins to address efficiency and possibly quality; however, the comparator groups make it difficult to evaluate utility in a routine clinical practice not to mention the applicability of the specific task. Further benefit will inevitably come from the use of AI with digital images and multiple other orthogonal data sets, for example, genomic data and radiologic images, to further enhance the value of data utilization for the health care system. Addressing these issues will be the first and immediate barrier to broad implementation.

Cost is the second barrier. This could be directly addressed if the value is clearly demonstrated and results in government and third-party payers developing reimbursement strategies for the use of AI in pathology. Even though some reimbursement codes exist for computational analyses, they are not widely used and often are rejected. With national health care reimbursement trends moving to quality and safety metrics for value-based care rather than fee for service, the recognition of AI as part of reimbursement strategies that reward value-based care would provide important incentives to develop and implement validated algorithms.

Third, education will be the greatest challenge and will require the longest period to address. AI and other computational methods must be integrated into all of our training programs. Future generations of pathologists must be comfortable and facile using digital images and other data in combination with computer algorithms in their daily practice; optimistically, it will take 5 to 10 years to build such a workforce, and that is only if the process begins today.

AI as an idea has become a major element of the health care landscape, and AI as a reality can potentially provide value in many sectors. Recent examples with skin lesions, diabetic retinopathy, and radiology detection have highlighted the value proposition AI provides to aid clinicians to improve quality, safety, diagnosis, and democratization of care. Radiologists can now read imaging studies from anywhere in the world at their home institution/office, bringing expert care to parts of the world that previously had limited expertise. Pathology has the opportunity to do the same with digital imaging and AI permits rapid and accurate local care. AI may be just what pathology has been waiting for. While still requiring evaluation within a normal surgical pathology workflow, deep learning has the opportunity to assist pathologists by improving the efficiency of their work, standardizing quality, and providing better prognostication. Like electron microscopy, immunohistochemistry, and molecular diagnostics ahead of AI, there is little risk of pathologists being replaced. Although their workflow is likely to change, the contributions of pathologists to patient care will continue to be critically important.

ARTICLE INFORMATION

Author Affiliation: Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts.

Corresponding Author: Jeffrey Alan Golden, MD, Department of Pathology, Brigham and Women's Hospital, 75 Francis St, Boston, MA 02115 (jagolden@bwh.harvard.edu).

Conflict of Interest Disclosures: The author has completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

REFERENCES

1. Accenture. Artificial intelligence is the future of growth. https://www.accenture.com/us-en/insight-artificial-intelligence-future-growth. Accessed October 3, 2017.

2. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444.

3. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-118.

4. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402-2410.

5. Yao J, Dwyer A, Summers RM, Mollura DJ. Computer-aided diagnosis of pulmonary infections using texture analysis and support vector machine classification. *Acad Radiol*. 2011;18(3):306-314.

6. Wang J, Yang X, Cai H, Tan W, Jin C, Li L. Discrimination of breast cancer with microcalcifications on mammography by deep learning. *Sci Rep*. 2016;6:27327.

7. Rajkomar A, Lingam S, Taylor AG, Blum M, Mongan J. High-throughput classification of radiographs using deep convolutional neural networks. *J Digit Imaging*. 2017;30(1):95-101.

8. Cheng JZ, Ni D, Chou YH, et al. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. *Sci Rep*. 2016;6:24454.

9. Korfiatis P, Kline TL, Coufalova L, et al. MRI texture features as biomarkers to predict MGMT methylation status in glioblastomas. *Med Phys*. 2016;43(6):2835-2844.

10. Ehteshami Bejnordi B, Veta M, van Diest PJ, et al; CAMELYON16 Consortium. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*. doi:10.1001/jama.2017.14585

11. CAMELYON16. https://camelyon16.grand-challenge.org/. Accessed October 2, 2017.

12. Rosenthal DL. Computerized scanning devices for Pap smear screening: current status and critical review. *Clin Lab Med*. 1997;17(2):263-284.

13. Sabo E, Beck AH, Montgomery EA, et al. Computerized morphometry as an aid in determining the grade of dysplasia and progression to adenocarcinoma in Barrett's esophagus. *Lab Invest*. 2006;86(12):1261-1271.

14. Sertel O, Kong J, Shimada H, Catalyurek UV, Saltz JH, Gurcan MN. Computer-aided prognosis of neuroblastoma on whole-slide images: classification of stromal development. *Pattern Recognit*. 2009;42(6):1093-1103.

15. Beck AH, Sangoi AR, Leung S, et al. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci Transl Med*. 2011;3(108):108ra113.

16. Yu KH, Zhang C, Berry GJ, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun*. 2016;7:12474.