

John Halamka, MD, MS;
Paul Cerrato, MA

Harvard Medical School, Boston, Mass, and New England Healthcare Exchange Network (Dr. Halamka); Beth Israel Deaconess Medical Center, New York, NY, and Warwick, NY (Mr. Cerrato; affiliated independent medical journalist). Dr. Halamka and Mr. Cerrato are coauthors of *Realizing the Promise of Precision Medicine* and *The Transformative Power of Mobile Medicine*.

✉ cerrato@aol.com,
pcerrato@optonline.net

The authors reported no potential conflict of interest relevant to this article.

An FP's guide to AI-enabled clinical decision support

To better understand the capabilities and challenges of artificial intelligence and machine learning, we look at the role they can play in screening for retinopathy and colon cancer.

PRACTICE RECOMMENDATIONS

› *Encourage patients with diabetes who are unwilling to have a regular eye exam to have an artificial intelligence-based retinal scan that can detect retinopathy.* **(B)**

› *Consider using a machine learning-based algorithm to help evaluate the risk of colorectal cancer in patients who are resistant to screening colonoscopy.* **(B)**

› *Question the effectiveness of any artificial intelligence-based software algorithm that has not been validated by at least 2 independent data sets derived from clinical parameters.* **(B)**

Strength of recommendation (SOR)

- (A)** Good-quality patient-oriented evidence
- (B)** Inconsistent or limited-quality patient-oriented evidence
- (C)** Consensus, usual practice, opinion, disease-oriented evidence, case series

Computer technology and artificial intelligence (AI) have come a long way in several decades:

- Between 1971 and 1996, access to the Medline database was primarily limited to university libraries and other institutions; in 1997, the database became universally available online as PubMed.¹
- In 2004, the President of the United States issued an executive order that launched a 10-year plan to put electronic health records (EHRs) in place nationwide; EHRs are now employed in nearly 9 of 10 (85.9%) medical offices.²

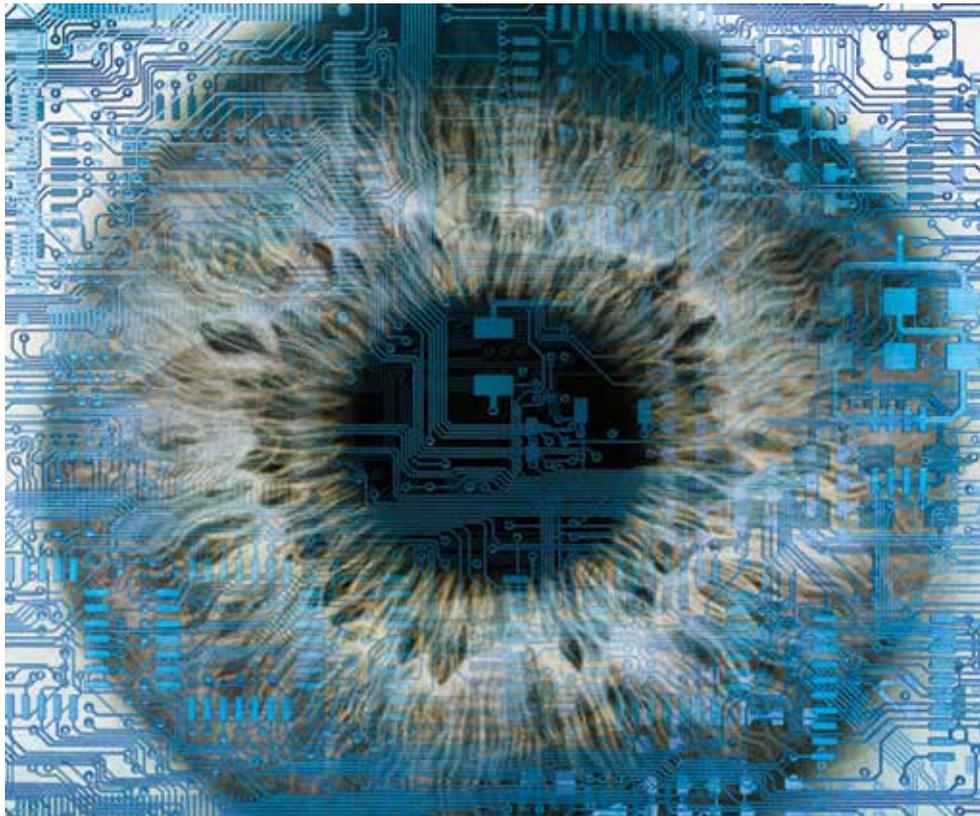
Over time, numerous online resources sprouted as well, including DxPlain, UpToDate, and Clinical Key, to name a few. These digital tools were impressive for their time, but many of them are now considered “old-school” AI-enabled clinical decision support.

In the past 2 to 3 years, innovative clinicians and technologists have pushed medicine into a new era that takes advantage of machine learning (ML)-enhanced diagnostic aids, software systems that predict disease progression, and advanced clinical pathways to help individualize treatment. Enthusiastic early adopters believe these resources are transforming patient care—although skeptics remain unconvinced, cautioning that they have yet to prove their worth in everyday clinical practice.

In this review, we first analyze the strengths and weaknesses of evidence supporting these tools, then propose a potential role for them in family medicine.

Machine learning takes on retinopathy

The term “artificial intelligence” has been with us for longer than a half century.³ In the broadest sense, AI refers to any



A software platform has been developed that is capable of diagnosing or screening for diabetic retinopathy without the involvement of an experienced ophthalmologist.

computer system capable of automating a process usually performed manually by humans. But the latest innovations in AI take advantage of a subset of AI called “machine learning”: the ability of software systems to learn new functionality or insights on their own, without additional programming from human data engineers. Case in point: A software platform has been developed that is capable of diagnosing or screening for diabetic retinopathy without the involvement of an experienced ophthalmologist.

The landmark study that started clinicians and health care executives thinking seriously about the potential role of ML in medical practice was spearheaded by Varun Gulshan, PhD, at Google, and associates from several medical schools.⁴ Gulshan used an artificial neural network designed to mimic the functions of the human nervous system to analyze more than 128,000 retinal images, looking for evidence of diabetic retinopathy. (See “Deciphering artificial neural networks,” page 490, for an explanation of how such networks function.⁵) The algorithm they employed was compared with the diagnostic skills of several board-certified ophthalmologists.

Using an area-under-the-receiver operating curve (AUROC) as a metric, and choos-

ing an operating point for high specificity, the algorithm generated sensitivity of 87% and 90.3% and specificity of 98.1% and 98.5% for 2 validation data sets for detecting referable retinopathy, as defined by a panel of at least 7 ophthalmologists. When AUROC was set for high sensitivity, the algorithm generated sensitivity of 97.5% and 96.1% and specificity of 93.4% and 93.9% for the 2 data sets.

These results are impressive, but the researchers used a retrospective approach in their analysis. A prospective analysis would provide stronger evidence.

That shortcoming was addressed by a pivotal clinical trial that convinced the US Food and Drug Administration (FDA) to approve the technology. Michael Abramoff, MD, PhD, at the University of Iowa Department of Ophthalmology and Visual Sciences and his associates⁶ conducted a prospective study that compared the gold standard for detecting retinopathy, the Fundus Photograph Reading Center (of the University of Wisconsin School of Medicine and Public Health), to an ML-based algorithm, the commercialized IDx-DR. The IDx-DR is a software system that is used in combination with a fundal camera to capture retinal images. The researchers found that “the AI system exceeded all pre-



INSTANT POLL

Are you (or other members of your organization) taking steps to learn more about how AI can benefit your practice?

- Yes, we are doing preliminary investigations now.
- Yes, we are aggressively looking into ways this technology can help us to improve care.
- No, we have taken no steps to investigate AI.

mledge.com/familymedicine

>
A large-scale validation study performed on data from Kaiser Permanente Northwest found that it is possible to estimate a person's risk of colorectal cancer by using age, gender, and complete blood count.

specified superiority endpoints at sensitivity of 87.2% ... [and] specificity of 90.7% ...”

The FDA clearance statement for this technology⁷ limits its use, emphasizing that it is intended only as a screening tool, not a stand-alone diagnostic system. Because IDx-DR is being used in primary care, the FDA states that patients who have a positive result should be referred to an eye care professional. The technology is contraindicated in patients who have a history of laser treatment, surgery, or injection in the eye or who have any of the following: persistent vision loss, blurred vision, floaters, previously diagnosed macular edema, severe nonproliferative retinopathy, proliferative retinopathy, radiation retinopathy, and retinal vein occlusion. It is also not intended for pregnant patients because their eye disease often progresses rapidly.

Additional caveats to keep in mind when evaluating this new technology include that, although the software can help detect retinopathy, it does not address other key issues for this patient population, including cataracts and glaucoma. The cost of the new technology also requires attention: Software must be used in conjunction with a specific retinal camera, the Topcon TRC-NW400, which is expensive (new, as much as \$20,000).

Speaking of cost: Health care providers and insurers still question whether implementing AI-enabled systems is cost-effective. It is too early to say definitively how AI and machine learning will have an impact on health care expenditures, because the most promising technological systems have yet to be fully implemented in hospitals and medical practices nationwide. Projections by Forbes suggest that private investment in health care AI will reach \$6.6 billion by 2021; on a more confident note, an Accenture analysis predicts that the best possible application of AI might save the health care sector \$150 billion annually by 2026.⁸

What role might this diabetic retinopathy technology play in family medicine? Physicians are constantly advising patients who have diabetes about the need to have a regular ophthalmic examination to check for early signs of retinopathy—advice that is often ignored. The American Academy of Oph-

thalmology points out that “6 out of 10 people with diabetes skip a sight-saving exam.”⁹ When a patient is screened with this type of device and found to be at high risk of eye disease, however, the advice to see an eye-care specialist might carry more weight.

Screening colonoscopy: Improving patient incentives

No responsible physician doubts the value of screening colonoscopy in patients 50 years and older, but many patients have yet to realize that the procedure just might save their life. Is there a way to incentivize resistant patients to have a colonoscopy performed? An ML-based software system that only requires access to a few readily available parameters might be the needed impetus for many patients.

A large-scale validation study performed on data from Kaiser Permanente Northwest found that it is possible to estimate a person's risk of colorectal cancer by using age, gender, and complete blood count.¹⁰ This retrospective investigation analyzed more than 17,000 Kaiser Permanente patients, including 900 who already had colorectal cancer. The analysis generated a risk score for patients who did not have the malignancy to gauge their likelihood of developing it. The algorithms were more sensitive for detecting tumors of the cecum and ascending colon, and less sensitive for detection of tumors of the transverse and sigmoid colon and rectum.

To provide more definitive evidence to support the value of the software platform, a prospective study was subsequently conducted on more than 79,000 patients who had initially declined to undergo colorectal screening. The platform, called ColonFlag, was used to detect 688 patients at highest risk, who were then offered screening colonoscopy. In this subgroup, 254 agreed to the procedure; ColonFlag identified 19 malignancies (7.5%) among patients within the Maccabi Health System (Israel), and 15 more in patients outside that health system.¹¹ (In the United States, the same program is known as LGI Flag and has been cleared by the FDA.)

Although ColonFlag has the potential to reduce the incidence of colorectal cancer,

Deciphering artificial neural networks

The promise of health care information technology relies heavily on statistical methods and software constructs, including logistic regression, random forest modeling, clustering, and neural networks. The machine learning-enabled image analysis used to detect diabetic retinopathy and to differentiate a malignant melanoma and a normal mole is based on neural networking.

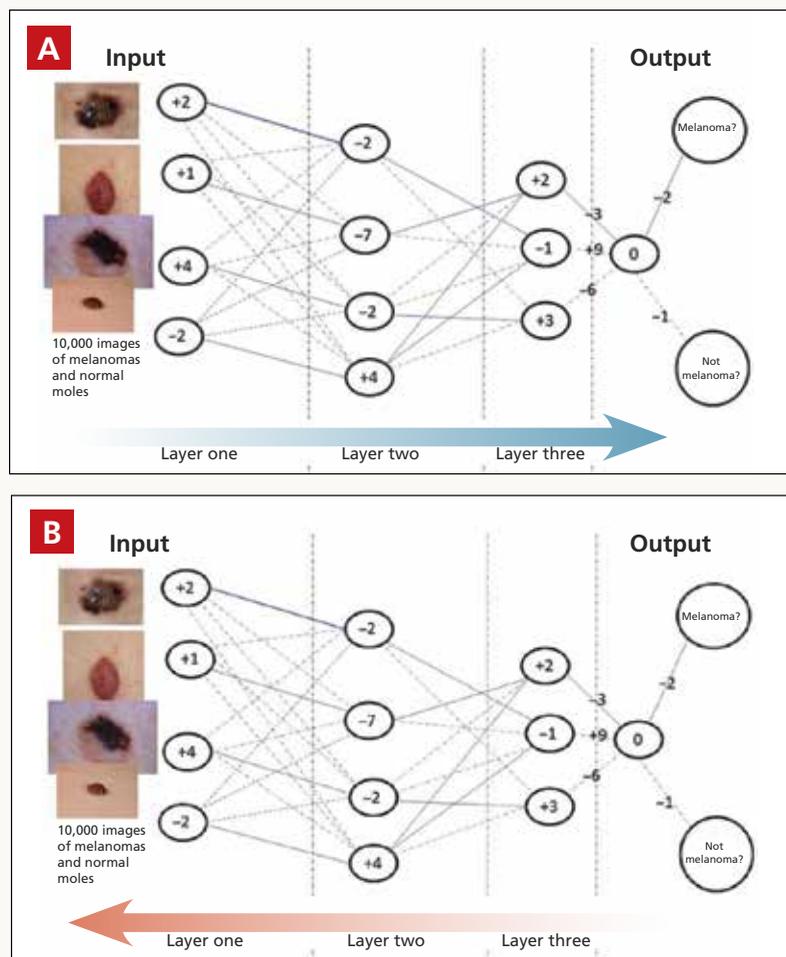
As we discussed in the body of this article, these networks mimic the nervous system, in that they comprise computer-generated “neurons,” or nodes, and are connected by “synapses” (FIGURE⁵). When a node in Layer 1 is excited by pixels coming from a scanned image, it sends on that excitement, represented by a numerical value, to a second set of nodes in Layer 2, which, in turns, sends signals to the next layer—and so on.

Eventually, the software’s interpretation of the pixels of the image reaches the output layer of the network, generating a negative or positive diagnosis. The initial process results in many interpretations, which are corrected by a backward analytic process called backpropagation. The video tutorials mentioned in the main text provide a more detailed explanation of neural networking.

FIGURE

How does a neural network operate?

A neural network designed to distinguish melanoma from a normal mole scans tens of thousands of images to teach itself how to recognize small differences between normal and abnormal skin growths (A). During the process of differentiating normal tissue and abnormal tissue, a neural network makes many mistakes. Backpropagation (B) looks back at these mistakes to help the program readjust its algorithms and improve its accuracy.



Source: Cerrato P, Halamka J. *The Transformative Power of Mobile Medicine*.⁵ Reproduced with permission of the authors and publisher.

other evidence-based screening modalities are highlighted in US Preventive Services Task Force guidelines, including the guaiac-based fecal occult blood test and the fecal immunochemical test.¹²

Beyond screening to applications in managing disease

The complex etiology of sepsis makes the condition difficult to treat. That complexity has also led to disagreement on the best

course of management. Using an ML algorithm called an “Artificial Intelligence Clinician,” Komorowski and associates¹³ extracted data from a large data set from 2 nonoverlapping intensive care unit databases collected from US adults. The researchers’ analysis suggested a list of 48 variables that likely influence sepsis outcomes, including:

- demographics,
- Elixhauser premonitory status,
- vital signs,
- clinical laboratory data,
- intravenous fluids given, and
- vasopressors administered.

Komorowski and co-workers concluded that “... mortality was lowest in patients for whom clinicians’ actual doses matched the AI decisions. Our model provides individualized and clinically interpretable treatment decisions for sepsis that could improve patient outcomes.”

A randomized clinical trial has found that an ML program that uses only 6 common clinical markers—blood pressure, heart rate, temperature, respiratory rate, peripheral capillary oxygen saturation (SpO₂), and age—can improve clinical outcomes in patients with severe sepsis.¹⁴ The alerts generated by the algorithm were used to guide treatment. Average length of stay was 13 days in controls, compared with 10.3 days in those evaluated with the ML algorithm. The algorithm was also associated with a 12.4% drop in in-hospital mortality.

Addressing challenges, tapping resources

Advances in the management of diabetic retinopathy, colorectal cancer, and sepsis are the tip of the AI iceberg. There are now ML programs to distinguish melanoma from benign nevi; to improve insulin dosing for patients with type 1 diabetes; to predict which hospital patients are most likely to end up in the intensive care unit; and to mitigate the opioid epidemic.

An ML Web page on the *JAMA* Network (<https://sites.jamanetwork.com/machine-learning/>) features a long list of published

research studies, reviews, and opinion papers suggesting that the future of medicine is closely tied to innovative developments in this area. This Web page also addresses the potential use of ML in detecting lymph node metastases in breast cancer, the need to temper AI with human intelligence, the role of AI in clinical decision support, and more.

The *JAMA* Network also discusses a few of the challenges that still need to be overcome in developing ML tools for clinical medicine—challenges that you will want to be cognizant of as you evaluate new research in the field.

■ **Black-box dilemma.** A challenge that technologists face as they introduce new programs that have the potential to improve diagnosis, treatment, and prognosis is a phenomenon called the “black-box dilemma,” which refers to the complex data science, advanced statistics, and mathematical equations that underpin ML algorithms. These complexities make it difficult to explain the mechanism of action upon which software is based, which, in turn, makes many clinicians skeptical about its worth.

For example, the neural networks that are the backbone of the retinopathy algorithm discussed earlier might seem like voodoo science to those unfamiliar with the technology. It’s fortunate that several technology-savvy physicians have mastered these digital tools and have the teaching skills to explain them in plain-English tutorials. One such tutorial, “Understanding How Machine Learning Works,” is posted on the *JAMA* Network (<https://sites.jamanetwork.com/machine-learning/#multimedia>). A more basic explanation was included in a recent Public Broadcasting System “Nova” episode, viewable at www.youtube.com/watch?v=xS2G0oolHpo.

■ **Limited analysis.** Another problem that plagues many ML-based algorithms is that they have been tested on only a single data set. (Typically, a data set refers to a collection of clinical parameters from a patient population.) For example, researchers developing an algorithm might collect their data from a single health care system.

Several investigators have addressed this shortcoming by testing their software on



A randomized clinical trial has found that an ML program that uses only 6 common clinical markers can improve clinical outcomes in patients with severe sepsis.

2 completely independent patient populations. Banda and colleagues¹⁵ recently developed a software platform to improve the detection rate in familial hypercholesterolemia, a significant cause of premature cardiovascular disease and death that affects approximately 1 of every 250 people. Despite the urgency of identifying the disorder and providing potentially lifesaving treatment, only 10% of patients receive an accurate diagnosis.¹⁶ Banda and colleagues developed a deep-learning algorithm that is far more effective than the traditional screening approach now in use.

To address the generalizability of the algorithm, it was tested on EHR data from 2 independent health care systems: Stanford Health Care and Geisinger Health System. In Stanford patients, the positive predictive value of the algorithm was 88%, with a sensitivity of 75%; it identified 84% of affected patients at the highest probability threshold. In Geisinger patients, the classifier generated a positive predictive value of 85%.

The future of these technologies

AI and ML are not panaceas that will revolutionize medicine in the near future. Likewise, the digital tools discussed in this article are not going to solve multiple complex medical problems addressed during a single office visit. But physicians who ignore mounting evidence that supports these emerging technologies will be left behind by more forward-thinking colleagues.

A recent commentary in *Gastroenterology*¹⁷ sums up the situation best: “It is now too conservative to suggest that CADE [computer-assisted detection] and CADx [computer-assisted diagnosis] carry the potential to revolutionize colonoscopy. The artificial intelligence revolution has already begun.” **JFP**

CORRESPONDENCE

Paul Cerrato, MA, cerrato@aol.com, pcerrato@optonline.net.
John Halamka, MD, MS, john.halamka@bilh.org.

References

- Lindberg DA. Internet access to National Library of Medicine. *Eff Clin Pract.* 2000;3:256-260.
- National Center for Health Statistics, Centers for Disease Control and Prevention. Electronic medical records/electronic health records (EMRs/EHRs). www.cdc.gov/nchs/fastats/electronic-medical-records.htm. Updated March 31, 2017. Accessed October 1, 2019.
- Smith C, McGuire B, Huang T, et al. The history of artificial intelligence. University of Washington. <https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf>. Published December 2006. Accessed October 1, 2019.
- 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;316:2402-2410.
- Cerrato P, Halamka J. *The Transformative Power of Mobile Medicine*. Cambridge, MA: Academic Press; 2019.
- Abramoff MD, Lavin PT, Birch M, et al. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med.* 2018; 1:39.
- US Food and Drug Administration. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. Press release. www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye. Published April 11, 2018. Accessed October 1, 2019.
- AI and healthcare: a giant opportunity. Forbes Web site. www.forbes.com/sites/insights-intelai/2019/02/11/ai-and-healthcare-a-giant-opportunity/#5906c4014c68. Published February 11, 2019. Accessed October 25, 2019.
- Boyd K. Six out of 10 people with diabetes skip a sight-saving exam. American Academy of Ophthalmology Website. <https://www.aao.org/eye-health/news/sixty-percent-skip-diabetic-eye-exams>. Published November 1, 2016. Accessed October 25, 2019.
- Hornbrook MC, Goshen R, Choman E, et al. Early colorectal cancer detected by machine learning model using gender, age, and complete blood count data. *Dig Dis Sci.* 2017;62:2719-2727.
- Goshen R, Choman E, Ran A, et al. Computer-assisted flagging of individuals at high risk of colorectal cancer in a large health maintenance organization using the ColonFlag test. *JCO Clin Cancer Inform.* 2018;2:1-8.
- US Preventive Services Task Force. Final recommendation statement: colorectal cancer: screening. www.uspreventiveservices.org/Page/Document/RecommendationStatementFinal/colorectal-cancer-screening2#tab. Published May 2019. Accessed October 1, 2019.
- Komorowski M, Celi LA, Badawi O, et al. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med.* 2018;24:1716-1720.
- Shimabukuro DW, Barton CW, Feldman MD, et al. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. *BMJ Open Respir Res.* 2017;4:e000234.
- Banda J, Sarraju A, Abbasi F, et al. Finding missed cases of familial hypercholesterolemia in health systems using machine learning. *NPJ Digit Med.* 2019;2:23.
- What is familial hypercholesterolemia? FH Foundation Web site. https://thefhfoundation.org/familial-hypercholesterolemia/what-is-familial-hypercholesterolemia?gclid=EAIaIQobChMIx8_XubvJ5QIV2IVaBR0wWgy8EAAYASAAEgIYnVD_BwE. Accessed November 1, 2019.
- Byrne MF, Shahidi N, Rex DK. Will computer-aided detection and diagnosis revolutionize colonoscopy? *Gastroenterology.* 2017;153:1460-1464.E1.

➤ The best possible application of AI might save the health care sector \$150 billion annually by 2026, according to an economic analysis.



Visit us @ mledge.com/familymedicine