REVIEW

THE AMERICAN JOURNAL of MEDICINE®

Artificial Intelligence in Medical Practice: The Question to the Answer?

D. Douglas Miller, MD, CM FACP,[a] Eric W. Brown, PhD[b]

[a]New York Medical College, Valhalla; [b]Foundational Innovations, IBM Watson Health, Yorktown Heights, NY.

ABSTRACT

Computer science advances and ultra-fast computing speeds find artificial intelligence (AI) broadly benefitting modern society—forecasting weather, recognizing faces, detecting fraud, and deciphering genomics. AI's future role in medical practice remains an unanswered question. Machines (computers) learn to detect patterns not decipherable using biostatistics by processing massive datasets (big data) through layered mathematical models (algorithms). Correcting algorithm mistakes (training) adds to AI predictive model confidence. AI is being successfully applied for image analysis in radiology, pathology, and dermatology, with diagnostic speed exceeding, and accuracy paralleling, medical experts. While diagnostic confidence never reaches 100%, combining machines *plus* physicians reliably enhances system performance. Cognitive programs are impacting medical practice by applying natural language processing to read the rapidly expanding scientific literature and collate years of diverse electronic medical records. In this and other ways, AI may optimize the care trajectory of chronic disease patients, suggest precision therapies for complex illnesses, reduce medical errors, and improve subject enrollment into clinical trials.
 • The American Journal of Medicine (2018) 131, 129–133

KEYWORDS: Analytics; Artificial intelligence; Big data; Chronic disease; Deep learning; Electronic medical record; Machine learning; Medical imaging; Natural language processing; Neural networks; Precision medicine

In 1936, mathematician Alan Turing published *On Computable Numbers, With an Application to the Entscheidungsproblem*, a paper later dubbed "*the founding document of the computer age.*"[1] Turing's life was reprised in the 2014 film, *The Imitation Game*. Attempting to solve the *Entscheidungsproblem*, Turing and his Princeton colleague, Alonzo Church, used calculus to define the concept of "*effective calculability.*" Such intelligent human problem-solving became the basis of computational models called algorithms.

In 1943, neurophysiologist Warren McCulloch and mathematician Walter Pitts modeled brain neuronal interactions using a simple neural network made of electrical circuits. The first computer research with artificial neural networks was done in the 1950s by Nathanial Rochester at International Business Machines (IBM), and Bernard Widrow and Marcian Hoff at Stanford. Today's computer scientists apply multilayered algorithms using a variety of artificial neural network configurations to solve complex problems. Modern artificial neural networks represent one of the most active areas of artificial intelligence (AI) research.

In 1964, television guru Merv Griffin invented *Jeopardy!*, America's third-longest running game show. In 2011, a supercomputer named for IBM's first chief executive, Thomas J. Watson, used AI to defeat 2 very intelligent humans in an exhibition match culminating with the correct response to this question: "*Which author's most famous novel was inspired by William Wilkinson's 'An Account of the Principalities of Wallachia and Moldavia'?"* (Answer: Bram Stoker's *Dracula*).

ANSWERABLE QUESTIONS

Some questions about AI's role in modern society have been answered:

Why has AI emerged as useful in several diverse sectors (business, science, government)?

Funding: None.

Conflicts of Interest: DDM: None; EWB: Employed by IBM; the employment relationship did not create direct or indirect financial or scientific conflicts in the preparation of this paper.

Authorship: All authors had access to the data and a role in writing this manuscript.

Requests for reprints should be addressed to D. Douglas Miller, MD, CM FACP, Division of Cardiology, Augusta University, 1120 15th Street, BBR-6518, GA, 30912.

E-mail address: ddouglasmiller1@gmail.com

0002-9343/$ - see front matter
https://doi.org/10.1016/j.amjmed.2017.10.035

How does AI differ from standard biostatistics?
What is "*big data*"? How does AI enable big dataset analysis?
How do AI applications differ from smart technologies (medical devices, digital diagnostics, data management systems) already used in medical practice?

INSIDE AI'S BLACK BOX

While AI encompasses a wide range of symbolic and statistical approaches to learning and reasoning (**Figure**), recent advances in algorithms, computational power, and access to large datasets have enabled artificial neural networks to emerge as the leading AI method. Artificial neural networks are flexible mathematical models that use multiple algorithms to identify complex nonlinear relationships within large datasets (analytics). Machines learn when errors encountered in response to minor algorithm modifications are corrected (training), progressively improving predictive model accuracy (confidence).[2]

Deep learning uses ultra-fast computing to rapidly optimize large multilayered datasets organized in a variety of configurations, including filter layers as convolutional neural networks and recursive layers as recurrent neural networks. Deep learning has been applied commercially since the 1990s,[3] and while modern math is similar to that employed in the 1980s, supercomputer speeds and Cloud networking permit deconvolution of massive datasets. In 2006, Hinton et al introduced a novel method to train very deep neural networks by pretraining one hidden algorithm layer at a time using an unsupervised machine learning procedure[4] and Bengio et al validated Hinton's work with test data and used it with other unsupervised techniques such as auto-encoders.[5]

Ten years later, deep learning modeling of big datasets exerts major influences on modern society—from Web searching to social media networking, and from financial technology banking to facial recognition.[3] Advanced algorithms achieve acceptable performance with ~5000 data points per category, and exceed human performance with datasets of >10 million labeled examples.[2] The bigger the dataset, the easier it is for machines to learn (gain confidence) because the burden of standard biostatistical estimation is reduced.[2] Despite this, like human thinking, predictive model confidence never reaches 100%.

CLINICAL SIGNIFICANCE

- Artificial intelligence (AI) medical image analysis achieves diagnostic speed exceeding, and accuracy paralleling, experts.
- AI will impact medical practice by applying natural language processing to "read" the expanding scientific literature and collate diverse electronic medical records.
- Machines learning directly from medical data could avert clinical errors due to human cognitive biases, positively impacting patient care.
- Because AI is neither astute nor intuitive, physicians will remain essential to cognitive medical practice.

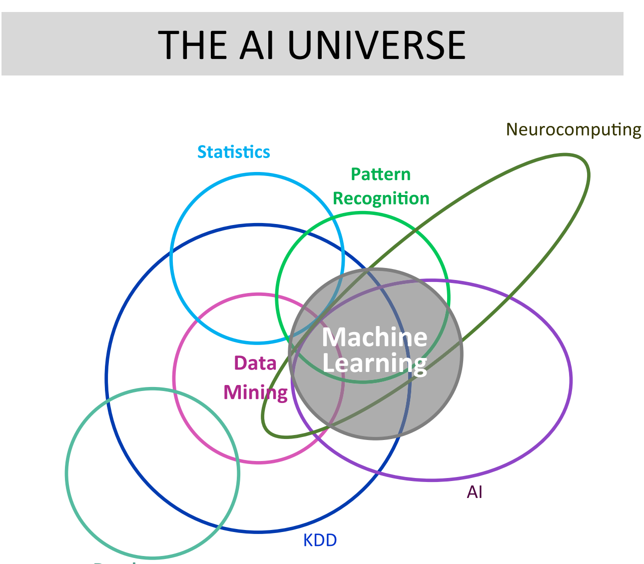

Figure In the computation science universe, artificial intelligence (AI) is distinguished from standard statistics and databases, but overlaps with knowledge discovery and data mining (KDD) methodologies that extract useful insights from large datasets. The mathematics of pattern recognition (kernel machines, cluster analysis) overlaps significantly with machine learning edge-detection algorithms and with neurocomputing based on artificial neural networks. The area of machine learning outside AI and within statistics/pattern recognition is linear regression analysis.

WORKS IN PROGRESS

Questions remain about the applicability, practicality, and value of AI in medical practice:

How is AI use in medical practice distinguished from big data analytics applications for health care delivery and population health?
Can AI address medical practice "*pain points*," providing more efficient and efficacious care while de-escalating physician burnout?
Will AI improve patient outcomes when used at the point of care?
Can Internet-of-Things health care facilities and medical homes become a platform for safer, higher quality, more connected patient care?

USE CASES FOR COGNITIVE MEDICAL PRACTICE

Simple neural networks have been used in medicine since the early 1990s to interpret electrocardiograms,[6] diagnose myocardial infarction,[7] and predict intensive care unit length of stay following cardiac surgery.[8] AI's scientific applications have proliferated, including image analysis (radiographic, histologic), text recognition with natural language processing, drug activity design, and prediction of gene mutation expression.[9,10] Recent AI applications provide proof of concept for AI use in specialty medical practice, while projecting future utility in general medical practice.

Cognitive Diagnostics

Gene chips are widely used to detect cancer cell gene expression. However, despite chips holding diagnostic probes for 20,000-50,000 genetic features, noisy data and experimental limitations reduce their clinical utility. Deep learning addresses this by reducing data diversity (dimensionality) and applying layered auto-encoding analyses to train artificial neural networks to achieve more accurate cancer detection and classification.[11]

Histopathology of 1417 skin images analyzed using deep learning architecture visual pattern analysis to detect basal cell carcinoma and differentiate malignant from benign lesions outperformed prior automated analyses, with diagnostic accuracy of >90% compared with experts.[12] Deep learning histopathology identifies metastatic breast cancer in sentinel lymph node biopsies, with diagnostic accuracy for tumor detection and localization similar to experts.[13] These systems train by comparing the features of millions of tumor-positive and -negative histological patches, postprocessing these data using heat maps to predict tumor probability. Combining pathologists and deep learning optimized performance, reducing the human error rate by 85%.

Convolutional neural networks outperformed 21 dermatologists at keratinocyte carcinoma and melanoma detection by classifying 129,450 images of 2032 malignant and benign skin diseases using multiple layered algorithms trained to identify common deadly skin cancers.[14]

Chronic Disease Management

AI analytics support the practice of precision medicine, especially in the difficult setting of chronic diseases characterized by multiorgan involvement, erratic acute events, and long illness progression latencies.

For >29 million Americans with diabetes, retinopathy is among the most debilitating complications. Using 128,175 retinal photographs from 5871 adults, 2 deep learning systems trained to detect and grade diabetic retinopathy and macular edema achieved high specificities (98%) and sensitivities (87%-90%) for detecting moderately severe retinopathy and macular edema, compared with 54 ophthalmologists and senior residents.[15] The feasibility of this approach in medical practice and its capacity to improve diabetes care and outcomes require validation.

Depression affects 6.8%-8.7% of the adult US population, resulting in 8 million annual ambulatory care visits.[16] Primary care practices are not equipped to manage chronic depressive illnesses. Phenotypic dimensionality and a paucity of objective depression activity markers may be addressable by applying deep learning to magnetic resonance image mapping of white matter neuronal water content.[17] Image heat-map pattern recognition was 74% accurate for predicting major depressive disorder, with certain brain regions contributing more to model confidence.

Congestive heart failure is a clinically and biologically diverse condition affecting 5.8 million Americans, and 23 million worldwide.[18] Heart failure with preserved ejection fraction (HFpEF) is a phenotypically heterogeneous condition influenced by numerous weak genetic factors, without proven therapies. When supervised machine learning was applied to 46 clinical variables from 397 HFpEF patients, phenotypic heat-map clusters predicted patient survival more accurately than commonly employed risk assessments.[19] AI approaches could identify HFpEF subsets or individuals that could benefit from therapies that failed to show survival benefits in clinical trial cohorts.

Electronic Medical Record Applications

Electronic medical records (EMRs) are purported tools for documenting and sharing medical care information. EMR challenges include lack of interoperability across technology platforms over time, and massive expansion of structured and unstructured data elements. Natural language processing is an AI tool that "reads" and contextualizes different medical words and expressions in EMRs. Available products can accurately compile and connect decades of accumulated diverse EMR data—history, physical, laboratory, imaging, medications—in a user-friendly manner. IBM Watson generates accurate universal problem lists from diverse EMRs in seconds, while also compiling relevant medical literature in response to clinical queries.[20] Deep learning modeling of EMR data memory can predict future illness trajectories and medical outcomes, confidently predicting interventions and readmissions in 2 patient cohorts that exert heavy economic and societal burdens—diabetes and mental health.[21]

POTENTIAL JEOPARDIES

Concerns about cognitive medical practice are largely the result of existing information deficits:

Will providers perceive AI as another technology barrier to direct patient care?
Does AI enhance or dis-intermediate patient–physician engagement?
What nonmedical barriers exist to the use of AI in direct patient care (eg, reimbursement, regulatory)?
Will AI put some physicians out of work (obsolescence) and/or reduce physician compensation (relative value)?
Are physicians using AI at risk for skill erosion in diagnostic expertise, clinical acumen, or critical thinking?
Will younger tech-savvy learners and clinicians become early technology adopters, driving the development of AI-infused cognitive practice?

TECHNOLOGY INSERTION

Tracey Kidder's 1981 Pulitzer Prize-winning book *The Soul of a New Machine*[22] underscored how imperfect humans remained critical to intelligent computer design. The current AI medical literature reproducibly supports a widely held tautology—that collaborative human-machine tasking improves performance over either alone. While AI's technology displacement curve is paralleled by an opportunity curve, concerns abide that AI will dislocate highly skilled health professionals from their jobs.

A tool is a device or implement used by humans for a particular function; tools are combined into machines for industrial production. At the turn of the 20th century, combustion engines combined tools to autonomously power vehicles over land. In 2016, global automobile sales increased to 88 million units, with China leading all nations. Public health evidence indicates that fossil-fueled vehicles emit multiple air pollutants, contributing to 2.1 million excess deaths in Asia alone between 1990 and 2010.[23] At the turn of the 21st century, mobile devices placed the data capture and analytics power of computers into human hands. By 2018, average daily mobile device use for Internet access alone will increase to 113 minutes per human. Research associates mobile device use to higher risks of cancer, accidents, and medical device interference.[24] Just as machines have created unanticipated risks for humans, there may be risks to AI use in medical practice.

The defense and aerospace industries often insert new or improved technology into an existing product or system.[25] Associated process management challenges include, "*platform modernization and achieving the rapid fielding of the new technology*." But the primary impediment to successful technology insertion is a lack of common understanding of the technology among key users.

Industrial technology insertion differs from new medical device or software regulation, under the aegis of the Food and Drug Administration. Although the Food and Drug Administration is establishing a digital health unit, US and European regulatory platforms are not yet equipped to oversee AI's insertion into medical practice. It is unclear whether the cost of using AI technologies in medical practice will be reimbursed by value-conscious insurers.

"IS THAT YOUR FINAL ANSWER?"

Final *Jeopardy!* answer: An 1816 medical instrument invented by Dr. René Laennec to avoid patient contact. Correct question: *"What is the stethoscope?"*

There are 2 reasons why medical schools still teach students to use a centuries-old tool. The first is that the stethoscope reveals diagnostic information helpful to patient care. The second is that the hand-held device requires learners to physically contact the precordium, a connection that is both humanistic of doctors and reassuring to patients. While experienced auscultators glean 75%-80% of the information generated by a Doppler echocardiogram, best medical practices and third-party reimbursement require that humans use this simple tool *prior to* employing more modern machines.

Today's cognitive machines have sophisticated sensors that capture big and little data, and generate corrective computer models simulating a rudimentary human nervous system. In response to driver-reported battery fires from road debris impact, Tesla Motors downloaded chassis height adjustments to *all* of its smart vehicles to mitigate further risk.

The daily practice of medicine is a game, of sorts, requiring repeated situational assessments, pattern recognition based on case experience, and evidence-based risk–benefit adjustments. Mounting performance pressures can prompt reliance on information-processing shortcuts—heuristic thinking or gaming cheats—to improve decision-making efficiency and workflow. Unfortunately, resulting cognitive biases may foster clinical errors. Machines that learn directly from medical data could avert such human cognitive biases, thereby contributing positively to patient care.

AI was not specifically developed as a tool for health care. And while AI is poised to address indurate medical practice "*pain points*," it is neither astute nor intuitive. So it is that humans will remain essential to the intelligent use of AI in medical practice.

References

1. Turing AM. On computable numbers, with an application to the entscheidungsproblem. *Proc Lond Math Soc*. 1936;Series 2(42):230-265.
2. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge MA, London, UK: The MIT Press; 2016; 96-161. Available at: www.deeplearningbook.org. Accessed May 30, 2017.
3. Le Cun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436-444.
4. Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. *Neural Comput*. 2006;18:1527-1554.
5. Bengio Y, Lamblin P, Popovivi D, Larochelle H. Greedy layer-wise training of deep networks. proc. *Adv Neural Inf Process Syst*. 2006;19:153-160.
6. Willems JL, Abreu-Lima C, Arnaud P, van Bimmel JH. The diagnostic performance of computer programs for the interpretation of electrocardiograms. *N Engl J Med*. 1991;325:1767-1773.
7. Baxt WG. Use of an artificial neural network for the diagnosis of myocardial infarction. *Ann Intern Med*. 1991;115(11):843-848.
8. Tu JV, Guerriere MR. Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery. *Proc Annu Symp Comput Appl Med Care*. 1992;666-672.
9. Leung MK, Xiong HY, Lee LJ, Frey B. Deep learning of the tissue-regulated splicing code. *Bioinformatics*. 2014;30:i121-i129.
10. Angermueller C, Parnamaa T, Parts L. Stegle. Deep learning for computational biology. *Mol Syst Biol*. 2016;12:878. Available at: https://dx.doi.org/10.15252%2Fmsb.20156651.
11. Fakoor R, Ladhak F, Nazi A, Huber M. Using deep learning to enhance cancer diagnosis and classification. In: *Proceedings of the 30th International Conference on Machine Learning*, Vol. 28. Atlanta, GA: JMLR; 2013.
12. Cruz-Roa AA, Arevalo Ovalle JE, Madabhushi A, Gonzales Osorio FA. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. In: Mori K, Sakuma I, Sato Y, Barillot C, Navab N, eds. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Vol. 8150. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2013; 403-410.

13. Wang D, Khosla A, Gargeya R, Irshad H, Beck AH. Deep learning for identifying breast cancer. In: *Proceedings of the International Society on Biomedical Imaging (ISBI), Quantitative Methods*. 2016.
14. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115-119.
15. 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.
16. Bishop TF, Ramsay PP, Casalino LP, Bao Y, Pincus HA, Shortell SM. Care management processes used less often for depression than for other chronic conditions in us primary care practices. *Health Aff (Millwood)*. 2016;35(3):394-400.
17. Schnyer DM, Clasen PC, Gonzalez C, Beevers CG. Evaluating the diagnostic utility of applying a machine learning algorithm to diffusion tensor mri measures in individuals with major depressive disorder. *Psychiatry Res*. 2017;264:1-9. Available at: http://dx.doi.org/10.1016/j.pscychresns.2017.03.003.
18. Roger VL. Epidemiology of heart failure. *Circ Res*. 2013;113:646-659.
19. Deo RC. Machine learning in medicine. *Circulation*. 2015;132:1920-1930.
20. Tsou C-H, Devarakonda M, Liang JJ. Toward generating domain-specific/personalized problem lists from electronic medical records. AAAI Fall Symposium, November 2015. Available at: https://www.aaai.org/ocs/index.php/FSS/FSS15/paper/viewFile/11733/11479. Accessed June 25, 2017.
21. Pham T, Tran T, Phung D, Venkatesh S. Predicting healthcare trajectories from medical records: a deep learning approach. *J Biomed Inform*. 2017;69:218-229. Available at: http://doi.org/10.1016/j.jbi.2017.04.001.
22. Kidder T. The Soul of a New Machine. Boston, MA: Little, Brown and Company; 1981.
23. Lim SS, Vos T, Flaxman AD, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systemic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(9859):2224-2260. Available at: http://dx.doi.org?10.1016/S0140-6736(12)61766-8.
24. NIH National Cancer Institute. Cell phones and cancer risk. May 2016. Available at: https://www.cancer.gov/about-cancer/causes-prevention/risk/radiation/cell-phones-fact-sheet. Accessed May 10, 2017.
25. Kerr CIV, Phaal R, Probert DR. Technology insertion in the defense industry: a primer. *Proc Inst Mech Eng B J Eng Manuf*. 2008;222:1009-1023. Available at: http://journals.sagepub.com. Accessed May 15, 2017.