Machine learning, natural language programming, and electronic health records: The next step in the artificial intelligence journey?

Neil Mehta, MBBS,[a] and Murthy V. Devarakonda, PhD[b] *Cleveland, Ohio, and Scottsdale, Ariz*

***Key words:** Machine learning, artificial intelligence, electronic health records, natural language processing, medical computer science*

The invention of the printing press and the development of the conveyor belt assembly line are good examples of how humans break down complex mechanical tasks into simpler well-defined steps that can then be automated. However, automating cognitive tasks has been a bigger challenge because it is not known precisely how human brains work. But is it necessary for humans to decode cognitive tasks for automation to work? Recent advances in artificial intelligence (AI) suggest otherwise, and the implications for health care are tantalizing.

Alan Turing, considered the father of computational science, defined AI systems as machines that "can act as intelligently as human beings."[1] This concept has been popularized in science fiction by HAL 9000 in *2001: A Space Odyssey* and Ava in *Ex Machina*, but it is not clear whether robots with AI are likely to replace physicians anytime soon. Presently, we are making good progress in developing algorithms and computational systems using machine learning (ML) and natural language processing (NLP) that do well on narrow problems and data sets.

This article reviews current efforts to apply AI in health care and describes how techniques could support intelligent data retrieval from electronic health records (EHR) in the near future.

ML AND NLP

Table I lists definitions of NLP and types of ML along with examples from real life. ML can be considered a form of applied statistics, which was first described in the 1950s. It was somewhat ahead of its time and fell into an "AI winter" (see Appendix E1 in this article's Online Repository at www.jacionline.org), but it is now seeing a huge resurgence. This renewed interest is particularly true of deep learning, which is being fueled by powerful processors, such as graphical processing units, new algorithms, and massive amounts of data. Unlike traditional ML, which requires cognitive engineers to identify the input or predictive variables (called features) for the models that predict the outputs or dependent variables (called labels), deep learning deduces the features that predict the outcomes. The other striking feature of deep learning is that the process used to create the model is often a black box, details of which are unavailable even to the programmer of the system. For example, a facial recognition system breaks down the photograph of a person into 128 numbers that are unique for that person.[2] The system is not specifically taught to use those numbers, and it is not clear to a human what those numbers mean, but the system works very well and keeps improving with more data. Although most of the current health care applications of ML are based on traditional and supervised ML, the hope and hype revolve around applications of deep learning for predictions, precision medicine, and population health.

NLP is concerned with programming computers to "understand" spoken or written language. This notion of understanding is not in the human sense but in the sense of being able to solve a problem. Modern NLP solutions make extensive use of ML. Google search, language translation, next-word suggestion in text-messaging apps, and IBM Watson's Jeopardy are working examples of recent progress in NLP. Sentiment analysis of user reviews of restaurants, movies, and products is a classic form of meaning extraction.

ML IN HEALTH CARE: WHERE ARE WE RIGHT NOW?

With the increasing use of EHRs and digital imaging in medicine, there are vast amounts of data that can be used potentially with ML to help patients and providers. Although quantitative data (vital signs and laboratory results) are easy for computational systems to analyze, there is ongoing research on the use of free-text notes (clinical visit notes and reports of test results) and radiology and pathology imaging data with NLP and ML.

The National Library of Medicine maintains a comprehensive medical metathesaurus (Unified Medical Language System [UMLS]; see the Appendix E2 in this article's Online Repository at www.jacionline.org) and tools to map text to UMLS (eg, Metamap), and this is driving sustained innovation in NLP in health care. Tasks like finding words representing diseases, medications, symptoms, and their modifiers in a clinic note or a medical textbook are becoming straightforward. Separately, the Unstructured Information Management Architecture (UIMA) was developed at IBM for analysis and searching of unstructured data using these techniques. The clinical Text Analysis and Knowledge Extraction System at the Mayo Clinic built on the UIMA to enable extraction of information from the EHR.[3] IBM Watson, which is also built on the UIMA, was able to perform well on questions from the

From [a]Education Informatics and Technology, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, and [b]the Department of Biomedical Informatics, Arizona State University, College of Health Solutions, Scottsdale.

Disclosure of potential conflict of interest: N. Mehta received a grant from IBM. M. V. Devarakonda declares no relevant conflicts of interest.

Received for publication September 12, 2017; revised February 17, 2018; accepted for publication February 23, 2018.

Available online March 5, 2018.

Corresponding author: Neil Mehta, MBBS, Education Informatics and Technology, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, NA2-24, 9500 Euclid Ave, Cleveland, OH 44195. E-mail: mehtan@ccf.org.

J Allergy Clin Immunol 2018;141:2019-21.

0091-6749/$36.00

https://doi.org/10.1016/j.jaci.2018.02.025

TABLE I. Simplified description of commonly used terms with examples

Terminology	Description	Examples of applications
NLP	Ability of computers to process spoken or written human (natural) language rather than mathematic equations or computer programs	Grammar checking in Microsoft Word or autocorrect and predictive text when messaging on a smartphone
ML	Computational systems that learn from data without being explicitly programmed	Classified into supervised vs unsupervised traditional vs deep learning
Supervised vs unsupervised ML		
Supervised ML	Algorithms are trained using data sets that have inputs and outputs. Once trained, the algorithms can predict an output from a given input.	Face recognition software used by Facebook or Google Photos; once trained, the software can identify subjects from newly uploaded photographs.
Unsupervised ML	There are no defined outputs and thus no training to predict outputs. Algorithms find hidden similarities and patterns to cluster similar data points.	When a site like Netflix provides recommendations for movies based on the pattern of watching movies and finding other customers with a similar pattern, such as "Other people who liked movie X, also like movie Y"
Traditional (usual) vs deep ML		
Traditional ML	The model "features" (ie, decision criteria) are identified by humans, and the machine learns the relationship between input and output through the features. Requires human effort, less data, and less computing power.	Credit card fraud detection algorithms are designed by humans who identify features predictive of fraudulent activity. These are designed to provide an appropriate balance between financial loss to institution and inconvenience to genuine user.
Deep ML	Given a very large data set, the algorithm learns the features that predict or explain the outcomes or labels. This process usually requires tremendous processing power (and training data, if supervised learning) because the models are organized into multiple layers in the form of neural networks.	The facial recognition example above uses deep ML. The machine identifies various characteristics that define a face uniquely and represents these by using vectors. This schema is a black box to humans, even those with the system.

American College of Physicians' Doctor's Dilemma (Medical Jeopardy) competition.[4] In 2015, Google's DeepMind collaborated with the United Kingdom's Royal Free NHS Foundation Trust to develop algorithms to alert health care providers about patients with acute kidney injury. This arrangement ran into problems because of privacy concerns, an important factor to consider as hospitals partner with technology firms to advance ML solutions for health care.

Currently, ML research is targeting higher-order semantics from clinical text, including identifying unknown drug adverse events,[5] problem-list and medication reconciliation, classifying sentences and extracting phenotypes from the EHR for clinical trials or population management.[6] Recently, IBM Watson technology was used to automatically curate a problem list from patients' complex electronic medical records. The system leveraged the UMLS and NLP on free-text notes along with coded orders data to develop an ML algorithm. When used on a validation data set of 15 patients with complex medical histories, practicing internists found this ML-created problem list preferable to the manually created problem lists in the EHR system. In addition, the ML was more sensitive than these internists at identifying clinically significant problems.[7]

There is also prolific ongoing research into the use of ML in radiology, with studies investigating automatic identification of pulmonary embolism in computed tomographic angiograms, microcalcifications in mammograms, polyps in computed tomographic colonography, and changes in patients with Alzheimer disease on brain magnetic resonance imaging. Similarly, clinical photographs and dermoscopic images are being studied for automatic recognition of melanomas and common skin cancers. Although more challenging because of variability in specimen preparation and pathology and density of data in the slides, progress is being made in ML techniques in pathology. In immunology ML techniques have been used to predict MHC-peptide binding for vaccine design, analysis of allergenicity of food components, and prediction of amyloidogenicity of immunoglobulin sequences. References to the clinical applications of ML mentioned in this paragraph are included in Appendix E3 in this article's Online Repository at www.jacionline.org.

WHAT IS NEXT FOR ML IN HEALTH CARE?

Clinicians make deductions and predictions based on available information. With increasing amounts of available data, these deductions and predictions should become more accurate. Unfortunately, reviewing and analyzing this vast amount of data requires time, which is in short supply for busy clinicians. The increased cognitive load and information management needs can lead to errors,[8] missed diagnoses, or ordering of unnecessary tests. As the amount of data continues to increase and the information in the medical literature grows, this process will become increasingly difficult. NLP and ML can create intelligent assistants for automatic information retrieval and semantic searches across multiple data sources for truly smart clinical decision and diagnostic support systems.

Let us consider an example of a clinician who has to spend a long time reviewing the records of a new patient with an increased creatinine level. If the patient's EHR had been analyzed by using NLP and ML, all the quantitative data from laboratory tests and vital signs and all the text from reports and clinician notes would be available in a form of a visual synopsis. Fig 1 displays an example of how this might work. The timeline tracks the level of creatinine superimposed with events that are known to affect renal function. The system is able to mine the medical literature to stay up to date with causes of kidney injury. It uses the National Library of Medicine taxonomy to "translate" the episode of

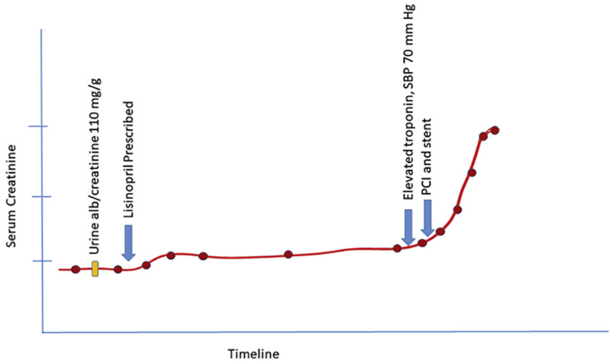

FIG 1. Visual summary of trends and possible causes of increased serum creatinine levels by using NLP and ML. *PCI,* Percutaneous coronary intervention.

hypotension and increased troponin level to acute myocardial infarction and shock and the percutaneous coronary intervention and stent with exposure to contrast material. In addition, it would be able to present clinical notes that specifically address the renal dysfunction along the timeline, and therefore the clinician could quickly scan what a nephrologist might have written regarding this topic at any specific point in the timeline. The system could automatically generate a customized prioritized problem list for each patient and render a similar smart visual summary for each.

In addition, it is also possible that the interface would capture this review and allow the clinician to document her assessment without having to write a separate note. Such assistants would free up clinicians from mundane information retrieval and documentation tasks, reduce cognitive load and burnout, and thus help clinicians provide more efficient and effective care.

CONCLUSION

The scenario described above is feasible in the near term. In addition to EHRs and digital imaging, increasing amounts of data from proteomics, genomics, and individual sensors, such as activity monitors, are becoming available. Combined with socioeconomic data, this network of real-time data streams could be made available for ML to potentially solve complex health care needs, such as predictive modeling, precision medicine, and population health.

There are many challenges before this future can come to pass. The first issue is one of trust. When a computer provides a recommendation, clinicians and patients would like to understand the logic behind it. ML algorithms can be "black boxes" that do not provide this information. It will take long years of good outcomes with ML to earn the trust of patients and providers. In addition, integrating these multiple data streams will need regulatory changes and a tremendous amount of effort to analyze and map the data.

Lastly, there is a fear expressed by various experts, such as Stephen Hawking and Elon Musk, that this future could become a dystopian one, with concerns of AI taking over control. Because health care lags other industries in adoption of ML, such a scenario, unlikely and futuristic as it is, should hardly be our primary concern. Instead, we should learn from other sectors and apply the lessons learned to solve the many problems facing health care. While these challenges are being addressed, appropriately designed and managed ML can help providers manage cognitive and information overload while providing better care to their patients.

REFERENCES

1. Turing AM. Computing machinery and intelligence. Mind 1950;59:433-60.
2. Schroff F, Kalenichenko D, Philbin J. FaceNet: a unified embedding for face recognition and clustering. Available at: http://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Schroff_FaceNet_A_Unified_2015_CVPR_paper.html. Accessed June 24, 2017.
3. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010;17:507-13.
4. Ferrucci D, Levas A, Bagchi S, Gondek D, Mueller ET. Watson: beyond Jeopardy! Artif Intell 2013;199:93-105.
5. Page D, Costa VS, Natarajan S, Barnard A, Peissig P, Caldwell M. Identifying adverse drug events by relational learning. Proc AAAI Conf Artif Intell 2012;2012:790-3.
6. Li D, Simon G, Chute CG, Pathak J. Using association rule mining for phenotype extraction from electronic health records. AMIA Summits Transl Sci Proc 2013;2013:142-6.
7. Devarakonda MV, Mehta N, Tsou C-H, Liang JJ, Nowacki AS, Jelovsek JE. Automated problem list generation and physicians perspective from a pilot study. Available at: http://www.ijmijournal.com/article/S1386-5056(17)30164-8/abstract. Accessed June 24, 2017.
8. Shachak A, Hadas-Dayagi M, Ziv A, Reis S. Primary care physicians' use of an electronic medical record system: a cognitive task analysis. J Gen Intern Med 2009;24:341-8.

APPENDIX E1. TIMELINE OF AI

1956: First AI conference at Dartmouth College[E1]

John McCarthy coined the term "Artificial Intelligence"

1956-1974: The golden years

Huge amount of optimism that machines would, by the 1980s, have the intelligence of an average human being and, for example, would be able to beat any human at chess. AI received wide publicity, including in the popular press, and large amounts of funding from the government. The latter was prompted partly because of hope that a machine would be able to translate Russian into English in real time, which was of great value in the Cold War era.[E2]

1974-1980: First AI winter

Disillusionment in the setting of very high expectations. Two reports, the ALPAC report[E3] and the Lighthill report,[E4] led to loss of funding for AI in the United States and United Kingdom.

1980-1987: The boom

Development of expert systems and special computers for Lisp programming.

1987-1993: The second AI winter

There were many reasons for the "failure" of AI dependency on rule-based expert systems that were too brittle to maintain, use of Lisp as a programming language, and use of special-purpose machines for it.[E5] Desktop and desk-side general-purpose computers became more powerful and were cheaper than the LISP machines. The special-purpose computer industry faltered commercially.

1993: The present

This period was studded with milestone events, such as IBM's Deep Blue beating Gary Kasparov at chess, teams from Stanford and Carnegie Mellon developing driverless cars, and IBM Watson winning Jeopardy. The availability of faster processors and vast amounts of data led to deep-learning algorithms that allowed systems to be trained from data instead of human-created rules. Integration of such tools into daily life began to occur, such as Google search, banking software, and speech recognition.

REFERENCES

E1. McCarthy J, Minsky M, Rocherster N, Shannon C. A proposal for the Dartmouth Summer Research Project on artificial intelligence Available at: http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html. Accessed September 9, 2017.

E2. Roland A, Shiman P. Strategic Computing: DARPA and the Quest for Machine Intelligence, 1983-1993. Boston: MIT Press; 2002:478.

E3. Pierce J, Carroll J, Hamp E, Hays D, Hockett C, Oettinger A, et al. Language and Machines: Computers in Translation and Linguistics. Washington (DC): Automatic Language Processing Advisory Committee; 196634, Report no. 1416. Available at: https://www.nap.edu/resource/alpac_lm/ARC000005.pdf. Accessed September 9, 2017.

E4. Lighthill J. Lighthill report: artificial intelligence: a paper symposium. Available at: https://pdfs.semanticscholar.org/b586/d050caa00a827fd2b318742dc80a304a3675.pdf. Accessed September 9, 2017.

E5. Nof SY. Springer Handbook of Automation. New York: Springer Science & Business Media; 2009:1841.

APPENDIX E2. THE UNIFIED MEDICAL LANGUAGE SYSTEM

The UMLS[E1] is a manually curated collection of biomedical terms drawn from many controlled vocabularies, such as all versions of the International Classification of Diseases and MeSH (medical subheadings for classification of journal articles and books) and are mapped to a single terminology.[E2] As a result, it allows mapping synonymous words and phrases in an article or in a clinical note to unique medical terms with specific meaning. It also allows for translation from one vocabulary to another.

UMLS also provides a taxonomy of medical terms, and therefore it can be used to find generalizations of a medical term (eg, hypertension is a vascular disease) and specific instances of a general term. Although not comprehensive, UMLS also maintains several useful relationships, such as metformin treats diabetes mellitus. Several knowledge sources (databases) derived from terminology and relationships, as well as a set of software tools,[E3] are available as a part of UMLS. It is common for most medical and clinical NLP software to use terminology mapping and/or relationships as a foundational step.[E4]

The UMLS was designed and is maintained by the US National Library of Medicine, is updated quarterly, and can be used for free. The project was initiated in 1986 by Donald A. B. Lindberg, MD, then Director of the Library of Medicine (source: Wikipedia).

REFERENCES

E1. Available at: https://www.nlm.nih.gov/research/umls.

E2. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004;32:D267-70.

E3. Demner-Fushman D, Rogers WJ, Aronson AR. MetaMap Lite: an evaluation of a new Java implementation of MetaMap. J Am Med Inform Assoc 2017;24:841-4.

E4. Cohen AM, Hersh WR. A survey of current work in biomedical text mining. Brief Bioinform 2005;6:57-71.

APPENDIX E3. CLINICAL APPLICATIONS OF ML

E1. Wang S, Summers RM. Machine learning and radiology. Med Image Anal 2012;16:933-51.

E2. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542:115-8.

E3. Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: comprehensive tutorial with selected use cases. J Pathol Inform 2016;7:29.

E4. Lafuente EM, Reche PA. Prediction of MHC-peptide binding: a systematic and comprehensive overview. Curr Pharmacol Des 2009;15:3209-20.

E5. Muh HC, Tong JC, Tammi MT. AllerHunter: a SVM-pairwise system for assessment of allergenicity and allergic cross-reactivity in proteins. PLoS One 2009;4:e5861.

E6. David MPC, Concepcion GP, Padlan EA. Using simple artificial intelligence methods for predicting amyloidogenesis in antibodies. BMC Bioinformatics 2010;11:79.