REVIEW

THE AMERICAN JOURNAL *of* MEDICINE®

Artificial Intelligence Transforms the Future of Health Care

Nariman Noorbakhsh-Sabet, MD,[a,b] Ramin Zand, MD, MPH,[c,b,d] Yanfei Zhang, PhD,[e] Vida Abedi, PhD[f,d]

[a]St. Jude Children's Research Hospital, Memphis, Tenn; [b]University of Tennessee Health Science Center, Memphis; [c]Geisinger Neuroscience Institute, Geisinger Health System, Danville, Penn; [d]Biocomplexity Institute, Virginia Tech, Blacksburg, Va; [e]Genomic Medicine Institute, Geisinger Health System, Danville,Penn; [f]Biomedical and Translational Informatics Institute, Geisinger Health System, Danville, Penn.

ABSTRACT

Life sciences researchers using artificial intelligence (AI) are under pressure to innovate faster than ever. Large, multilevel, and integrated data sets offer the promise of unlocking novel insights and accelerating breakthroughs. Although more data are available than ever, only a fraction is being curated, integrated, understood, and analyzed. AI focuses on how computers learn from data and mimic human thought processes. AI increases learning capacity and provides decision support system at scales that are transforming the future of health care. This article is a review of applications for machine learning in health care with a focus on clinical, translational, and public health applications with an overview of the important role of privacy, data sharing, and genetic information.

 • *The American Journal of Medicine (2019) 132:795−801*

KEYWORDS: Artificial intelligence (AI); Integrated health care systems; Machine learning; Medical informatics; Precision medicine

INTRODUCTION

Machine learning, a popular subdiscipline of artificial intelligence (AI), uses large data sets and identifies interaction patterns among variables. These techniques can discover previously unknown associations, generate novel hypotheses, and drive researchers and resources toward most fruitful directions.[1] Machine learning can be applied in various fields, such as financial, automatic driving, smart home, etc. In medicine, machine learning is widely used to build automated clinical decision systems.

Most approaches to machine learning fall into two main categories: supervised and unsupervised. Supervised methods are great for classification and regression. Recent examples include detection of a lung nodule from a chest X-ray[2]; risk estimation models of anticoagulation therapy;[3] implantation of automated defibrillators in cardiomyopathy;[4] use in classification of stroke and stroke mimic;[5] modeling of CD4+ T cell heterogeneity;[6] outcome prediction in infectious diseases;[7] detection of arrhythmia in electrocardiogram (ECG);[8] and design and development of *in silico* clinical trial[9] among others.

Unsupervised learning does not require labeled data. It aims to identify hidden patterns present in the data and is often used in data exploration and in the generation of novel hypotheses.[2] In three separate studies in heart failure with preserved ejection fraction among patients who had a heterogeneous condition with no proven therapies without human intervention,[10] researchers used unsupervised learning[2] to revisit failed clinical trials such as treatment with spironolactone,[11] enalapril,[12] and sildenafil[13] compared with placebo to identify a subclass of patients who might benefit from specific therapies.

There are other algorithms, such as reinforcement learning, which can be viewed as a combination of supervised

Funding: This work was in part supported by Geisinger Research, and by funds from the National Institute of Health (NIH) grant No. R56HL116832 to Sutter Health and sub-awarded to VA (Sub-PI, Geisinger) as well as funds from the Defense Threat Reduction Agency (DTRA) grant No. HDTRA1-18-1-0008 to Virginia Tech and sub-awarded to VA (Sub-PI, Geisinger, sub-award No. 450557-19D03). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Conflicts of Interest: None.

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

Requests for reprints should be addressed to Vida Abedi, PhD, Staff Scientist, Biomedical and Translational Informatics Institute at Geisinger Health System, 100 N Academy Ave, Danville, PA 17822.

E-mail addresses: vidaabedi@gmail.com, vabedi@geisinger.edu

0002-9343/
https://doi.org/10.1016/j.amjmed.2019.01.017

and unsupervised learning to maximize the accuracy of using trial and error[14] (Table 1).

Deep learning is a subset of machine learning that mimics the operation of the human brain using multiple layers of artificial neuronal networks to generate automated predictions from training data sets. Models based on deep learning strategy tend to have multiple parameters and layers; thus, model overfitting could lead to poor predictive performance. Increasing the training sample size, decreasing the number of hidden layers, and ensuring the data is well-balanced can help prevent overfitting. Overall, deep learning is compelling in image recognition[15] and in modeling disease onset[16] using temporal relations among events. A deep neural network was trained on more than 37,000 head computed tomography (CT) scans for intracranial hemorrhage and subsequently evaluated on 9,500 unseen cases, reducing time to diagnosis of new outpatient intracranial hemorrhage by 96% with an accuracy of 84%.[17]

Cognitive computing as a subset of AI involves self-learning systems using pattern recognition and natural language processing for semi- or unstructured data. Cognitive computing mimics the operation of human thought processes, with the goal of creating automated computerized models that can solve problems without human assistance. Examples include research in computer-brain-interface[18,19] and commercial products such as the IBM Watson.[20]

CLINICAL SIGNIFICANCE

- Artificial intelligence (AI) increases learning capacity and provides decision support system at scales that are transforming the future of health care.
- Artificial intelligence has been implemented in disease diagnosis and prognosis, treatment optimization and outcome prediction, drug development, and public health.
- Technological advances require collecting and sharing the massive amount of data and thus generate concerns about privacy.

Although none of these approaches can rapidly and simultaneously consider different disease-related parameters in a user-independent fashion, they are promising venues and are changing the way medicine is practiced. Health care providers should be ready for the upcoming AI age and embrace the added capabilities that would lead to more efficient and effective care. In this article, we review the applications and challenges as well as ethical consideration and perspectives of machine learning in medicine, translational research, and public health (Table 2).

CLINICAL APPLICATION

Disease Prediction and Diagnosis

Despite the increasing application of AI in health care, the research mainly concentrates around cancer, nervous system, and cardiovascular diseases because they are the leading causes of disability and mortality. However, infectious and chronic diseases (eg, type 2 diabetes,[21] inflammatory bowel disease,[22] *Clostridium difficile* infection[9]) have also been getting considerable attention. Early diagnosis can now be achieved for many conditions by improving the extraction of clinical insight and feeding such insight into a well-trained and validated system.[23] For instance, the US Food and Drug Administration (FDA) permitted applying of diagnosis software designed to detect wrist fractures in adult patients.[24] In another study on 1,634 images of

Table 1 Main Machine-Learning Strategies: Their Characteristics, Scope, and Limitations

ML types	Algorithms Description	Characteristics	Limitation
Supervised Learning	Labeled data set System trained with human feedback	Applications include classification, regression, and prediction; ideal for modeling disease prognosis or treatment outcome. Modeling algorithms include Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF)	Requires a large amount of labeled data for training; need validation in an independent cohort.
Unsupervised Learning	Non-labeled data by humans	Applications include mainly pattern recognition; ideal for modeling disease mechanisms, identifying hidden patterns in genotype or phenotype data. Modeling algorithms include various clustering methods	Needs validation in several independent cohorts
Reinforcement Learning	Hybrid approach; the goal is to maximize accuracy by trial and error; especially useful in a complex environment	Applications include chemistry, robotics, games, resource management in computer clusters, personalized recommendations	Memory intensive

ML = machine learning.

Table 2 Selected Areas in Medicine Where Machine Learning Has High Potential and Implications

Field	Application
Clinical	Disease prediction and diagnosis
	Treatment effectiveness and outcome prediction
Translation	Drug discovery and repurposing
	(*In Silico*) Clinical trial
Public health	Epidemic outbreak prediction
	Precision health

cancerous and healthy lung tissue, the algorithm identified healthy cases and distinguished, as accurately as 3 pathologists, between 2 common types of lung cancer.[25] In the United States, more than 6% of adult populations are affected by depression. Predicting major depressive disorder was 74% accurate by image heatmap pattern recognition.[26]

Several studies are looking at the potential of AI in timely and precise diagnosis of disease. Supervised methods are effective tools at capturing nonlinear relationships for complex and multifactorial disease classification. In a cohort study of 260 patients, Abedi et al[27] found that the model can better diagnose acute cerebral ischemia than trained emergency medical respondents. Although noisy data and experimental limitations reduce the clinical utility of the models, deep learning methods can address these limitations by reducing the dimensionality of the data through layered auto-encoding analyses. Examples include analysis of more than 1,400 images from 308 histopathology region of skin to detect basal cell carcinoma and differentiate malignant from benign lesions, achieving a diagnostic accuracy of >90% compared with experts[28]; or examination of more than 41,000 digital-screening breast mammograms for identifying dense or non-dense breast tissue, where 94% of the 10,763 deep learning assessments were accepted by the interpreting radiologist.[29]

Treatment Effectiveness and Outcome Prediction

Treatment effectiveness and outcome prediction are also important areas with the potential clinical implication in disease-management strategies and personalized care plans. A decade ago, only molecular and clinical information was exploited to predict cancer outcomes. With the development of high-throughput technologies, including genomic, proteomic, and imaging technologies, new types of input parameters have been collected and used for prediction. With a large sample size and integrated multimodal data types, including histological or pathological assessments,[30] these methods could considerably (15%-25%) improve the accuracy of cancer susceptibility, outcome prediction, and prognosis.[31]

Electronic health records (EHRs) are effective tools for documenting and sharing health care information. Integrating machine learning-based modeling designed specifically for administrative data sets can facilitate the detection of potential complications, improve health care resource utilization, and outcomes at a personalized level.[32,33] Utilization of machine learning applied to EHR data has been shown to predict outcome in patients with sepsis.[7] A large-scale mortality study based on machine learning in more than 170,000 patients with 331,317 echocardiographies by Samad et al[34] achieved 96% accuracy to predict patients' survival based on echocardiography combined with EHR data. In terms of algorithm improvement, Smith et al[35] developed a deep neural network model for 12-lead ECG analysis and compared it to the conventional algorithm in emergency department ECGs; their result showed an accuracy of 92% for finding a major abnormality.

AI analytics can be used in chronic disease management characterized by multi-organ involvement, acute variable events, and long illness progression latencies. For instance, retinopathy can be predicted using machine learning. Training 2 validation data sets using deep learning to detect and grade diabetic retinopathy and macular edema achieved a high specificity and sensitivity for detecting moderately severe retinopathy and macular edema after each image was graded by ophthalmologists between 3 and 7 times.[36]

To improve care in congestive heart failure, a study used supervised machine learning on 46 clinical variables from 397 patients with heart failure with preserved ejection fraction. Phenotypic heatmap predicted patient survival more accurately than commonly employed risk assessment tools.[2]

One of the goals of precision medicine in cancer is the accurate prediction of optimal drug therapies from the genomic data of individual patient tumors.[37] In a study, researchers presented an open-access algorithm for the predictive response of cancers to 7 common chemotherapeutic medications.[38] Precision medicine success depends on algorithm ability to translate large compendia of *-omics* data into clinically actionable predictions. For example, Costello et al[39] analyzed 44 drug sensitivity prediction algorithms on 53 breast cancer cell lines with available genomic information to fulfill dose-response values of growth inhibition for each cell line exposed to 28 therapeutic compounds.

TRANSLATION APPLICATION

Drug Discovery and Repurposing

About 25% of all discovered drugs were the result of a chance when different domains were brought together accidentally.[40] Targeted drug discovery is preferred in pharmaceuticals because of the explicit mechanism, higher success rate, and lower cost when compared to traditional blind screening. Machine learning is now used in the process of drug discovery due to high costs of drug development, increasing availability of 3-dimensional structural information that can guide the characterization of drug targets, and extremely low success rates in clinical trials.[41] Machine learning can be used as a bridge to achieve cross-domain linkage. It can identify a newly approved drug by recognizing contextual clues like a discussion of its indication or side effects.[20]

Despite these novel approaches in drug discovery, there are important challenges, including data access and that in general, different data sets are stored in a variety of repositories. Furthermore, raw data from clinical trials and other preclinical studies are typically not available. However, overall, AI has been successful when applied to available sources, including the use of drug information to extract insight about mechanism-of-action by applying techniques such as similarity metrics across all diseases to find shared pathways.[20] Another example includes the use of natural-language processing for identification of hidden or novel associations that might be important in the detection of potential adverse drug effects based on scientific publications.[42]

Clinical Trial and *in Silico* Clinical Trials

Clinical trial design has its roots in classical experimental design. However, the clinical investigators are not able to control various sources of variability. Ethical issues are paramount in clinical research. Subject enrollment can become lengthy and costly.[43,44]

The machine-learning approach using *in silico* data set was introduced to describe the numerical methods used in drug development in oncology by modeling biological systems in the setting of clinical trial studies and hospital databases, paving the way to predictive, preventive, personalized, and participatory medicine.[45] This approach gives the researchers the ability to partially replace animals or humans in a clinical trial and generates virtual patients with specific characteristics to enhance the outcome of such studies. These methods are especially helpful for pediatric or orphan disease trials and can be applied in pharmacokinetics and pharmacodynamics from the preclinical phase to post-marketing.[45,46] In a study, a large *in silico* randomized, placebo-controlled Phase III clinical trial study was designed in which investigators used virtual treatments on synthetic patients with Crohn disease. Results showed a positive correlation between the initial disease activity score and the drop in the disease activity score but with different medications' efficacy.[47] The model did not highly score the investigational drug GED-0301; this prediction was further validated when the company that was running the clinical trial on GED-0301 stopped the trial after it failed to clear an interim futility review.[48] *In silico* clinical trials can have considerable potentials in design and discovery phases of biomedical product, biomarker identification, dosing optimization, or the duration of the proposed intervention.[49]

PUBLIC HEALTH RELEVANCE

Epidemic Outbreak Prediction

The infectious disease distribution pattern between population groups with known probabilities are based on prior knowledge of ecological and biological features of the environment. Early prediction of the epidemic (such as peak and duration of infection) is possible if model parameters are partially known.[50] Potential outbreak areas for filoviruses were predicted in the west, southwest, and central parts of Uganda, which were related to bat distribution and previous outbreaks areas.[51] In another study, Kesorn et al[52] predicted the morbidity rate of dengue hemorrhagic fever in central Thailand by estimating the infection rate in the female *Aedes aegypti* larvae mosquitoes and achieved a prediction accuracy of >95% and 88%, respectively, in the training and test set.

Precision Health

Genetic and biomedical studies have continued investigation efforts with the goal of revealing connections between genes and human traits or diseases. Regularized logistic regression is an important tool for related applications. Many studies rely on large-scale sensitive genotype or phenotype data, and sharing across institutions is paramount for the success of such studies.[53]

There are many such examples in recent years. For instance, in a recent case-control study with limited sample size, researchers developed an algorithm to integrate personal whole genome sequencing and EHR data and used this algorithm to study abdominal aortic aneurysm. They assessed the effectiveness of modifying personal lifestyles, given personal genome baselines, which demonstrated the model's utility as a personal health management model. Such studies have the potential to shed lights on the biological architecture of other complex diseases.[54] In a recent review, Torkamani et al[55] examine the core disciplines that enable *high-definition medicine*, given our recent technological advances and high-resolution data.

Challenges and Perspectives. Machine learning's ultimate goal is to develop algorithms that are capable of self-improving with experience and continuously learning from new data and insights and to find answers to an array of questions. The compelling opportunities in precision medicine offered by complex algorithms are accompanied by computational challenges. In 2012, the Obama administration announced "Big Data Research and Development Initiative" investment to "help solve some of the Nation's most pressing challenges."[56] The achievement of this potential requires novel approaches to address at least 3 technical challenges:[57] *volume*, which is the scale of data inputs, outputs, and attributes—this challenge can be addressed in part by using clusters of CPUs, data-sharing system or cloud, and deep-learning methods; *variety*, which is the different formats of data (ie, image, video, and text)—

this challenge can be partially addressed by using novel deep-learning methods to integrated data from various sources; and *velocity*, which is the speed of streaming data— to address this challenge, online learning approaches can be developed.

The ethical challenges presented by data science have also been an area of debate. These challenges can be mapped within the conceptual space and described by 3 branches of research: the ethics of data and privacy, the ethics and morality of algorithms, and the ethics and values of practices.[58] Among those, privacy has been the center of attention. Privacy is defined as a fundamental human right in the Universal Declaration of Human Rights at the 1948 United Nations General Assembly. Machine learning plays a key role in the development of precision medicine, whereby treatment is customized to the clinical or genetic risk factors of the patient. These advances require collecting and sharing a massive amount of data and, thus, generate concern about privacy.[59]

At the same time, health care institutions need to communicate with the public and collaborate with scientific communities and government agencies.[60] In this situation, a privacy-preserving framework is necessary and should be applied to a large range of domains in which the privacy and confidentiality of study participants and institutions is of concern.[61] As a standard practice, many institutions collaborate and use the deidentification process to share clinical data or perform a meta-analysis; each contributing site performs an analysis in house. These processes reduce the scope of clinical data sharing. For example, the DNAnexus clinical trial solution service powers the US Food and Drug Administration's platform for advancing regulatory standards.[62] St. Jude Cloud is a data-sharing resource for the global research community.[63] eMERGE is a national network organized and funded by the National Human Genome Research Institute (NHGRI) that combines DNA biorepositories with electronic medical record (EMR) systems for large-scale, high-throughput genetic research in support of implementing genomic medicine.[64] In Europe, the UK Biobank is a national and international health resource with unparalleled research opportunities and is open to all bona fide health researchers.[65]

The most important issue when developing machine learning in a clinical setting is the issue of trust when both clinicians and patients accept the recommendations provided by the system.[66] The data is noisy, complex, high-dimensional with thousands of variables, and biased for the catchment area of the originating hospital systems where the model was trained. Furthermore, missing data is not at random. Missingness can be to the result of incompleteness, inconsistency, or inaccuracy.[67,68] Imputation, or predicting missing values, also has its unique challenges. Standardized techniques such as the MICE algorithm[69] or novel imputation methods[70] have been proposed. Other challenges in mining the EHR data include different protocols and changes that are introduced at various time periods, without documentation for the research team, and policy changes and reimbursement rules that are introduced that may affect how patients seek care and how the treatment is redesigned based on their needs and their insurance coverage. Therefore, to develop models using EHR, the researchers must work closely with care providers and others within the health care system to increase the predictive power of the modeling-enabled discoveries.

Other limitations are lack of interoperability across technology platforms over time, and massive expansion of structured and unstructured data elements. Natural-language processing can be used to process and contextualize different medical words and expressions.[71] However, robust infrastructures have to be in place to be able to handle a large number of clinical notes. For instance, it is possible to use robust infrastructure to process millions of notes and identify patients who need a follow-up appointment for preventive care in hospital settings.[72]

Today's approaches to machine learning are near to real-world conditions. Due to the rapid technological advancements, tasks previously limited to humans will be taken on by algorithms.[73] The ability of machine learning to transform data into insight will affect the field of medicine, displacing much of the work of radiologists and anatomical pathologists. However, clinical medicine has always required doctors to handle huge amounts of data, from history and physical examinations to laboratory and imaging studies and, now, genetic data. The ability to manage this complexity has always set good doctors apart.[74]

References

1. Koprowski R, Foster KR. Machine learning and medicine: book review and commentary. *Biomed Eng Online.* 2018;17:17.
2. Deo RC. Machine learning in medicine. *Circulation.* 2015;132:1920–1930.
3. Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest.* 2010;137:263–272.
4. O'Mahony C, Jichi F, Pavlou M, et al. A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM risk-SCD). *Eur Heart J.* 2014;35:2010–2020.
5. Abedi V, Goyal N, Tsivgoulis G, et al. Novel screening tool for stroke using artificial neural network. *Stroke.* 2017;48:1678–1681.
6. Lu P, Abedi V, Mei Y, et al. Supervised learning methods in modeling of CD4+ T cell heterogeneity. *BioData Min.* 2015;4:27.
7. Bogle B, Balduino R, Wolk DM, et al. Predicting mortality of sepsis patients in a multi-site healthcare system using supervised machine learning. In: *Int'l Conf. of Health Informatics and Medical Systems*; 2018.; . Available at: https://csce.ucmss.com/cr/books/2018/LFS/CSREA2018/HIM3645.pdf [Accessed February 5, 2019 .
8. Chen Y, Wang X, Jung Y, et al. Classification of short single lead electrocardiograms (ECGs) for atrial fibrillation detection using piecewise linear spline and XGBoost. *Physiol Meas.* 2018;39:104006. https://doi.org/10.1088/1361-6579/aadf0f.
9. Leber A, Hontecillas R, Abedi V, et al. Modeling new immunoregulatory therapeutics as antimicrobial alternatives for treating *Clostridium difficile* infection. *Artif Intell Med.* 2017;78:1–13.
10. Udelson JE. Heart failure with preserved ejection fraction. *Circulation.* 2011;124:540–543.
11. Pitt B, Pfeffer MA, Assmann SF, et al. Spironolactone for heart failure with preserved ejection fraction. *N Engl J Med.* 2014;370:1383–1392.

12. Kitzman DW, Hundley WG, Brubaker PH, et al. A randomized double-blind trial of enalapril in older patients with heart failure and preserved ejection fraction: effects on exercise tolerance and arterial distensibility. *Circ Heart Fail*. 2010;3:477–485.
13. Guazzi M, Vicenzi M, Arena R, Guazzi MD. Pulmonary hypertension in heart failure with preserved ejection fraction: a target of phosphodiesterase-5 inhibition in a 1-year study. *Circulation*. 2011;124:164–174.
14. Krittanawong C, Zhang H, Wang Z, Aydar M, Kitai T. Artificial intelligence in precision cardiovascular medicine. *J Am Coll Cardiol*. 2017;69:2657–2664.
15. Lee EJ, Kim YH, Kim N, Kang DW. Deep into the brain: artificial intelligence in stroke imaging. *J Stroke*. 2017;19:277–285.
16. Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. *J Am Med Inform Assoc*. 2017;24:361–370.
17. Arbabshirani MR, Fornwalt BK, Mongelluzzo GJ, et al. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical work flow integration. *npj Digit Med.* 2018;1:9. https://doi.org/10.1038/s41746-017-0015-z.
18. Bashivan P, Yeasin M, Bidelman GM. Temporal progression in functional connectivity determines individual differences in working memory capacity. In: *2017 International Joint Conference on Neural Networks (IJNN)*, IEEE; 2017:2943–2949. https://doi.org/10.1109/IJCNN.2017.7966220.
19. Elahian B, Yeasin M, Mudigoudar B, Wheless JW, Babajani-Feremi A. Identifying seizure onset zone from electrocorticographic recordings: A machine learning approach based on phase locking value. *Seizure*. 2017;51:35–42.
20. Chen Y, Elenee Argentinis JD, Weber G. IBM Watson: how cognitive computing can be applied to big data challenges in life sciences research. Clin Ther. 2016;38:688–701.
21. Kagawa R, Kawazoe Y, Ida Y, et al. Development of type 2 diabetes mellitus phenotyping framework using expert knowledge and machine learning approach. *J Diabetes Sci Technol*. 2017;11:791–799.
22. Bassaganya-Riera J. Introduction to accelerated path to cures and precision medicine in inflammatory bowel disease. In: Bassaganya-Riera J, ed. *Accelerated Path to Cures*. New York: Springer International; 2018:1–6.
23. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017;2:230–243.
24. US Food and Drug Administration. FDA permits marketing of artificial intelligence algorithm for aiding providers in detecting wrist fractures [press release]. May 24, 2018. Available at: https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm608833.htm. Accessed February 5, 2019.
25. Razavian N, Tsirigos A. Pathologists meet their match in tumour-spotting algorithm. *Nature*. 2018;561:436–437.
26. 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 Neuroimaging*. 2017;264:1–9.
27. Abedi V, Goyal N, Tsivgoulis G, et al. Novel screening tool for stroke using artificial neural network. *J Stroke*. 2017;48:1678–1681.
28. Cruz-Roa AA, Arevalo Ovalle JE, Madabhushi A, González Osorio FA. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. *Med Image Comput Comput Interv*. 2013;16:403–410.
29. Lehman CD, Yala A, Schuster T, et al. Mammographic breast density assessment using deep learning: clinical implementation. *Radiology*. 2019;290:52–58.
30. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotodias DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J*. 2014;15:8–17.
31. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform*. 2007;11:59–77.
32. Rivers EP, McIntyre L, Morro DC, Rivers KK. Early and innovative interventions for severe sepsis and septic shock: taking advantage of a window of opportunity. *CMAJ*. 2005;173:1054–1065.
33. Hillman K, Chen J, Cretikos M, et al. Introduction of the medical emergency team (MET) system: a cluster-randomised controlled trial. *Lancet*. 2005;365:2091–2097.
34. Samad MD, Ulloa A, Wehner GJ, et al. Predicting survival from large echocardiography and electronic health record datasets: optimization with machine learning [e-pub ahead of print]. *JACC Cardiovasc Imaging*. doi:10.1016/j.jcmg.2018.04.026, accessed February 5, 2019.
35. Smith SW, Walsh B, Grauer K, et al. A deep neural network learning algorithm outperforms a conventional algorithm for emergency department electrocardiogram interpretation. *J Electrocardiol.* 2019;52:88–95. https://doi.org/10.1016/j.jelectrocard.2018.11.013.
36. 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.
37. Haverty PM, Lin E, Tan J, et al. Reproducible pharmacogenomic profiling of cancer cell line panels. *Nature*. 2016;533:333–337.
38. Huang C, Mezencev R, Mcdonald JF, Vannberg F. Open source machine-learning algorithms for the prediction of optimal cancer drug therapies. *PIoS One*. 2017;12:1–14.
39. Costello JC, Heiser LM, Georgii E, et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nat Biotechnol*. 2014;32:1202–1212.
40. Hargrave-Thomas E, Yu B, Reynisson J. Serendipity in anticancer drug discovery. *World J Clin Oncol*. 2012;3:1–6.
41. Pinyi L, Bevan DR, Leber A, Hontecillas R, Tubau-Juni N, Bassaganya-Riera J. Computer-aided drug discovery. In: Bassaganya-Riera J, ed. *Accelerated Path to Cures*. New York: Springer International; 2018:7–24.
42. Abedi V, Yeasin M, Zand R. Empirical study using network of semantically related associations in bridging the knowledge gap. *J Transl Med*. 2014;12:324.
43. Design and Analysis of Clinical Trials. *PennState Eberly College of Science* https://link.springer.com/chapter/10.1007%2F978-3-319-73238-1_5
44. Zand R, Abedi V, Hontecillas R, et al. Development of synthetic patient population and in silico clinical trials. In: Bassaganya-Riera J, ed. *Accelerated Path to Cures*, New York: Springer International; 2018:57-77. https://doi.org/10.1007/978-3-319-73238-1_5.
45. Gal J, Milano G, Ferrero JM, et al. Optimizing drug development in oncology by clinical trial simulation: why and how? *Brief Bioinform.* 2018;19:1203–1217. https://doi.org/10.1093/bib/bbx055.
46. Harnisch L, Shepard T, Pons G, Della Pasqua O. Modeling and simulation as a tool to bridge efficacy and safety data in special populations. *CPT Pharmacometrics Syst Pharmacol*. 2013;27:e28.
47. Abedi V, Lu P, Hontecillas R, et al. Chapter 28 phase iii placebo-controlled, randomized clinical trial with synthetic crohn's disease patients to evaluate treatment response. *Emerg Trends Appl Infrastructures Comput Biol Bioinformatics Syst Biol*. 2016;27:411–427.
48. Taylor NP. Celgene cans phase 3 trial of $710M Crohn's drug GED-0301. *Fierce Biotech*. October 20, 2017. Available at: https://www.fiercebiotech.com/biotech/celgene-cans-phase-3-trial-710m-crohn-s-drug-ged-0301 [Accessed February 5, 2019 .
49. Carlier A, Vasilevich A, Marechal M, de Boer J, Geris L. In silico clinical trials for pediatric orphan diseases. *Sci Rep*. 2018;8:2465.
50. Zamiri A, Yazdi HS, Goli SA. Temporal and spatial monitoring and prediction of epidemic outbreaks. *IEEE J Biomed Heal Informatics*. 2015;19:735–744.
51. Nyakarahuka L, Ayebare S, Mosomtai G, et al. Ecological niche modeling for filoviruses: a risk map for ebola and marburg virus disease outbreaks in Uganda. *PLoS Curr*. 2017;5. https://doi.org/10.1371/currents.outbreaks.07992a87522e1f229c7cb023270a2af1.
52. Kesorn K, Ongruk P, Chompoosri J, et al. Morbidity rate prediction of dengue hemorrhagic fever (dhf) using the support vector machine and the *Aedes aegypti* infection rate in similar climates and geographical areas. *PLoS One*. 2015;10:e0125049.

53. Xie W, Kantarcioglu M, Bush WS, et al. SecureMA: protecting participant privacy in genetic association meta-analysis. *Bioinformatics.* 2014;30:3334–3341.
54. Li J, Pan C, Zhang S, et al. Decoding the genomics of abdominal aortic article decoding the genomics of abdominal aortic aneurysm. *Cell.* 2018;174:1361–1372 [e10].
55. Torkamani A, Andersen KG, Steinhubl SR, Topol EJ. High-definition medicine. *Cell.* 2017;170:828–843.
56. The White House. Obama administration unveils 'big data' initiative: announces $200 million in new R&D investments [press release]. March 29, 2012. Available at: https://obamawhitehouse.archives.gov/the-press-office/2015/11/19/release-obama-administration-unveils-big-data-initiative-announces-200. Accessed February 5, 2019.
57. Chen X, Member S, Lin X. Big data deep learning: challenges and perspectives. 2014;2. doi: 10.1109/ACCESS.201.2325029.
58. Floridi L, Taddeo M. What is data ethics? *Philos Trans A Math Physi Eng Sci.* 2016;374 [pii: 20160360].
59. Azencott C. Machine learning and genomics: precision medicine versus patient privacy. *Philos Trans Ser A Math Phys Eng Sci.* 2018;376 [pii: 20170350].
60. Kayaalp M. Patient privacy in the era of big data. *Balkan Med J.* 2018;35:8–17.
61. Li W, Liu H, Yang P, Xie W. Supporting regularized logistic regression privately and efficiently. *PLoS One.* 2016;11:e0156479.
62. DNANEXUS. OPTIMIZE & DE-RISK YOUR CLINICAL TRIALS. (2018). https://www.dnanexus.com/clinicaltrials
63. St Jude Children's Research Hospital. St. Jude Cloud. (2018). https://www.stjude.cloud
64. eMERGE. Electronic Medical Records and Genomics network. (2018). https://emerge.mc.vanderbilt.edu
65. UK Biobank. (2018). https://www.ukbiobank.ac.uk
66. Mehta N, Devarakonda M. Machine learning, natural language programming, and electronic health records: The next step in the artificial intelligence journey? *J Allergy Clin Immunol.* 2018;141:2019–2021.
67. Botsis T, Hartvigsen G, Chen F, Weng C. Secondary use of ehr: data quality issues and informatics opportunities. *AMIA Jt Summits Transl Sci Proc.* 2010;1:1–5.
68. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009;338:b2393. https://doi.org/10.1136/bmj.b2393.
69. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med.* 2011;30:377–399.
70. Abedi V, Shivakumar MK, Lu P, et al. Latent-based imputation of laboratory measures from electronic health records: case for complex diseases. *bioRxiv.* 2018. https://doi.org/10.1101/275743.
71. Miller DD, Brown EW. Artificial intelligence in medical practice: the question to the answer? *Am J Med.* 2018;131:129–133.
72. Karunakaran B, Misra D, Marshall K, Mathrawala D, Kethireddy S. Closing the loop—Finding lung cancer patients using NLP. In: *2017 IEEE International Conference on Big Data (Big Data)*, IEEE; 2017:2452–2461. https://doi.org/10.1109/BigData.2017.8258203.
73. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. *Radiographics.* 2017;37:505–515.
74. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *N Engl J Med.* 2016;375:1216–1219.