Artificial intelligence in clinical health care applications

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

Viewpoint

Artificial Intelligence in Clinical Health Care Applications: Viewpoint

Michael van Hartskamp, PhD; Sergio Consoli, PhD; Wim Verhaegh, PhD; Milan Petkovic, PhD; Anja van de Stolpe, MD, PhD

Philips Research, Eindhoven, Netherlands

Corresponding Author:
Anja van de Stolpe, MD, PhD
Philips Research
HTC11, p247
High Tech Campus
Eindhoven, 5656AE
Netherlands
Phone: 31 612784841
Email: anja.van.de.stolpe@philips.com

Abstract

The idea of artificial intelligence (AI) has a long history. It turned out, however, that reaching intelligence at human levels is more complicated than originally anticipated. Currently, we are experiencing a renewed interest in AI, fueled by an enormous increase in computing power and an even larger increase in data, in combination with improved AI technologies like deep learning. Healthcare is considered the next domain to be revolutionized by artificial intelligence. While AI approaches are excellently suited to develop certain algorithms, for biomedical applications there are specific challenges. We propose six recommendations—the 6Rs—to improve AI projects in the biomedical space, especially clinical health care, and to facilitate communication between AI scientists and medical doctors: (1) Relevant and well-defined clinical question first; (2) Right data (ie, representative and of good quality); (3) Ratio between number of patients and their variables should fit the AI method; (4) Relationship between data and *ground truth* should be as direct and causal as possible; (5) Regulatory ready; enabling validation; and (6) Right AI method.

Introduction

The idea of artificial intelligence (AI) has a long history. Since the 1950s there have been several revolutionary promises of AI replacing human work within a few decades. It turned out, however, that reaching intelligence at human levels was more complicated, which led to several "AI winters," where interest in AI disappeared [1]. Currently, we are experiencing a renewed interest in AI, fueled by an enormous increase in computing power and an even larger increase in data generation. In combination with improved algorithms that allow training of deep neural networks, several high-tech companies have reached successes in performing tasks that are close to human or even beyond human performance: playing games like chess and Go, image recognition and computer vision, natural language processing, machine translation, and self-driving cars are just a few examples.

Health care is considered the next domain to be revolutionized by AI [2-5]. In addition to many academic efforts, companies are also getting involved. IBM has developed Watson for several health applications, such as Watson for Oncology and Watson for Genomics, and there is a large number of start-ups addressing all possible aspects of the health continuum [6,7].

Artificial Intelligence

The term *artificial intelligence* is used to indicate development of algorithms that should execute tasks that are typically performed by human beings and are, therefore, associated with intelligent behavior. AI makes use of a variety of techniques, such as deep learning, but also probabilistic methods like Bayesian modeling [8,9]; for definitions, see He et al [2]. Colloquially, the term is applied to a machine that mimics cognitive functions, such as learning and problem solving [2,4,10,11].

Use of Artificial Intelligence Methods in Health Care

It is clear that health care has numerous needs that could benefit from solutions developed with, or by embedding, artificial intelligence [2,4,8]. In this brief article, we focus on the contributions AI can make to clinical health care, a domain that poses new and sometimes unique challenges to

the application of AI. In the next sections, we discuss some important challenges and provide our recommendations on how to deal with them.

While radiology imaging was first in delivering digital data, digital pathology is a more recent revolutionary development [4,12,13]. In addition, for many years, hospitals have been digitizing their medical patient records [14]. Hence, a large and ever-increasing body of reasonably annotated clinical data has been collected: partially structured data in machine-readable formats, such as those from medical imaging, and partially unstructured data in natural language. As in other industrial sectors, it is expected that this *big data* movement can be leveraged to transform health care and drive unprecedented improvements in quality of patient diagnostics, treatment, care, and clinical outcome. Expected results range from identification of individuals at high risk for a disease, to improved diagnosis and matching of effective personalized treatment to the individual patient, as well as out-of-hospital monitoring of therapy response [15]. Although these opportunities and this potential are widely acknowledged, it is important to understand what can be delivered in practice with the current state-of-the-art AI technologies and which applications require further advances in AI to become feasible.

Artificial Intelligence Methods: Supervised or Unsupervised Learning; Knowledge-Based or Data-Driven?

Multiple AI technologies are available to choose from [2,9,16]. Algorithmic learning-based AI can be performed in a supervised mode; this means that a *ground truth* label is available for every data sample, which guides the AI effort and is based on domain knowledge. It will be obvious that the correctness of *ground truth* labels is a prerequisite for good performance of an AI solution. The alternative, unsupervised mode, is when no *ground truth* is available and only similarities can be found with a yet undefined meaning [2,16].

Machine learning traditionally involves a human to determine features of the data, using domain knowledge. In contrast, deep learning allows finding such features from the data *by itself*. The features are subsequently used in various models. Some of those can be knowledge-based models in which new deep learning-defined features are integrated according to knowledge [17,18]. The current interest in AI from industry comes from the recent breakthroughs in data-driven approaches, such as deep learning, and their applicability in industrial applications such as speech recognition, machine translation, and computer vision. Still, it is expected that combining data-driven and knowledge-based approaches will bring AI to the next level, much closer to human intelligence [17,18].

Artificial Intelligence for Well-Defined Narrow Tasks: Radiology Imaging and Digital Pathology

One of the more studied and successfully executed AI opportunities is in imaging. For example, AI technologies can be applied to problems such as distinguishing cell nuclei or certain cell types present in a tumor sample on a histopathology slide, using slide images obtained with a digital pathology scanner [19-21]. Such images are made with consistent equipment and acquired in a controlled fashion, generating images consisting of uniform data and providing very good representations of the phenomena to be modelled. The problem domain is limited. For instance, in the training process, the AI system gets as input raw images with associated labels for different cell types that are provided by a pathologist. The pathologist is providing a *ground truth* , based on existing expert knowledge (eg, on the different cell types or architecture present in the tissue slide). Deep learning has been applied to this problem and AI technologies already outperform manually crafted tissue analysis technologies [16,20,22,23]. It is expected that they will soon be on par or better than a human pathologist on certain well-defined histology feature recognition and measurement tasks, though not yet for clinical interpretation.

Artificial Intelligence for Wider Purposes: Coupling Patient Clinical Data to Clinical Outcome

On the other hand, research projects are ongoing using multimodal data (ie, a combination of datasets of a different data type), for example, to enable prediction of prognosis of a patient or clinical outcome after a certain treatment. One may, for example, use medical imaging data combined with histopathology and clinical laboratory data and even lifestyle data to try to predict survival, risk of rehospitalization within a certain number of days, etc. Such projects remain challenging and have typically been proven not to be very successful [5,24]. IBM's Watson for Oncology claims to integrate all available cancer patient data and disease information for improved diagnosis and therapy decision making. In 2013, the MD Anderson Cancer Center started using IBM Watson technology to increase effective treatment of cancer patients; however, the project was stopped in 2017 because it "did not meet its goals" [25,26]. In contrast, the concordance with respect to clinical interpretation of single-modality genome sequencing data using Watson for Genomics versus a clinical genomics expert group was reportedly quite good, between 77% and 97%, depending on the type of identified genomic mutations [7].

The most difficult challenge for AI in the coming years will be to move from successful narrow domains into wider-purpose, multimodality, data systems. A promising approach here is not to find one methodology to address every problem but to separate the wider-purpose AI goal into smaller goals. In this approach, subgroups of the data may be processed separately with suitable AI methods to provide meaningful, clinically relevant output. For example, for cardiac ultrasound images, one could zoom in to develop an algorithm with deep learning to measure left ventricular volume; or for pathology slide images, one could zoom in to develop an algorithm for recognition and quantification of a specific cell type (eg, lymphocytes). To increase chances at success, it is important to determine a well-defined, focused, and clinically relevant

question for an AI project that can be adequately answered with the available data.

The conclusion of this section is that a relevant and well-defined clinical question should come first.

The Relationship Between Patient or Sample Numbers and Data Variables

Many AI techniques, especially deep learning, rely on the availability of large datasets or *big data* [15]. Sometimes domain knowledge can help to create additional data derived from the data that are available. It is important, however, to distinguish the type of data that is needed. In games, such as chess and Go, it is easy to artificially synthesize additional data of the right type to increase the size of the dataset. With respect to medical and histopathology imaging, large amounts of data are available since samples are defined on an image pixel basis. Using this type of data, with relatively few images one can create millions of annotated samples with drawing tools. It is relatively easy to further augment every sample with artificially generated variations (eg, mirror copies, rotated versions, modified intensities, and modified colors) without consequences for the annotation.

In contrast, in clinical health care the type of data typically is a pathology or radiology report from a patient, associated with a clinical annotation such as diagnosis or response to therapy. In this case, the number of samples is generally equal to the number of patients. The annotation is often more difficult, as it requires an expert physician to provide the *ground truth*. When using multimodal data to find parameters that, for example, predict clinical outcome, despite all digital records and digital health devices, there are not enough data. The number of patients for which the necessary multimodal data are available is, in general, the limiting factor for using AI methods on such combined data sources to create a valid algorithm for risk prediction, diagnosis, or a therapeutic decision. When the number of patients of a specific defined disease (sub)type is low, the often-heard strategy is to extend a study to include more patients, even all patients worldwide, which requires addressing various legal and technical barriers. However, this is still likely to fail in reaching the required patient number; with efforts to increase the number of patients for inclusion in data analysis, the amount of variation per patient, including many unknown features and variables, tends to grow as well, leading to uncontrolled data variation. This is caused by the large variation in human individuals: their DNA (ie, just think of the 3 billion base pairs and the near-infinite combinations of genomic variations), their lifestyle, family medical history, use of medication, etc. Moreover, patients are never treated in exactly the same manner in the various hospitals, bringing in many additional variables. It is a well-recognized issue in clinical trials run by pharma companies [24]. The challenge is to minimize such unwanted variables in the patient or sample set to analyze. Much of this uncontrolled variation is not recorded or, at best, only in a very noisy way. The number of unknown parameters that may have influenced the outcome, especially if its measurement lies many years after the diagnosis and treatment, is typically underestimated. Examples of failure of AI methods caused by these issues

include many genome-wide association studies aimed at identification of clinically useful genomic risk factors for complex diseases and genomic studies aimed at identification of biomarkers for cancer diagnostics and treatment decisions [27].

Similar challenges are present in other domains, but solutions in those areas can be invoked that are not possible in the health care domain. In natural language processing, Google Translate is a well-known example. When it started, translations were of very poor quality and heavily criticized, but Google decided to keep the service up and running; online feedback was used to collect a large amount of translation data, enabling continuous improvement of the performance of the translation algorithm [28].

In summary, for applying AI to multimodal patient data, the number of patients from whom the complete set of multimodal data is available is frequently too limited to address the *curse of dimensionality*. In the scientific community, dimensionality reduction remains an active research area [29-31]. In clinical application areas for AI, it remains the main challenge to address. The first solution lies in reducing data modality and bringing the number of variables (P) on the right level in relation to the number of patients or samples (N) for which a *ground truth* is available. The desired solution will reduce high-dimensional data to biologically sound knowledge-based features. Introducing knowledge-based computational approaches is expected to provide a way forward to reduce model freedom and handle high-dimensional data [32].

The conclusion of this section is that the ratio between the number of patients and their variables should fit the AI method.

Insufficient Data Quality and Many Subjective Parameters

For the patient data that is available, it turns out that this data is usually neither 100% complete nor 100% correct. For example, diagnoses are not always complete or correct, or they were not correctly entered into the digital domain. The main diagnosis is, in general, reasonably well-documented; however, side diagnoses and complications that arise, for example, during hospital admission or in the home setting, as well as treatment details are less-accurately or not documented [5]. For many clinical variables, such as diagnoses, the *ground truth* comes from a physician's judgement and cannot be objectively measured or quantified. For example, it is documented that histopathology diagnoses differ to a varying extent among pathologists that diagnose the same slide [33-35]. As a consequence, datasets may be incomplete and noisy, and presumed *ground truths* may not always be correct.

The conclusion of this section is that the right data (ie, representative and of good quality) needs to be obtained.

Causal Relations Versus Correlations: A Role for Bayesian Reasoning

Any data-driven approach on data for which the ratio of number of patients (ie, samples) to variables is too low can

lead to multiple spurious correlations [36]. This means that the data suggest a correlation between two factors, but this is purely due to chance and there is no underlying explanation or causal relationship. In machine learning, this can easily lead to overfitting and finding of irrelevant correlations [16,37]. Also, for clinical implementation, any interesting correlation (eg, a feature or combination of features associated with increased disease risk) needs to be clinically validated at high cost, where lack of causality generally results in very low success rates. Therefore, turning an algorithm, based on correlations, into a successful proposition will, in general, be easier if causal relations underlie the found correlations. Knowledge-based reasoning techniques, such as Bayesian network models, can reduce the number of spurious relations and overfitting problems by using existing knowledge on causal data relations to eliminate noisy data [11,14]. As an additional advantage, Bayesian models can deal very well with uncertainty and missing variables, which is the rule rather than the exception in clinical data [11]. Not a coincidence, with respect to patient data interpretation, Bayesian models reason the way a medical doctor does [38].

The conclusion of the section is that the relationship between data and *ground truth* should be as direct and causal as possible.

Validation of Artificial Intelligence-Based Solutions

For many of the success stories of AI, a robust and reliable result is usually not necessary. For a free translation service, the consequence of a wrong decision is at most a dissatisfied customer. Improvements of those services could happen relatively quickly because many of those AI applications are deployed in the field and iteratively improve their performance on the basis of new data, thus learning from their mistakes. In sharp contrast to most of these consumer or lifestyle solutions based on AI, every clinical application, be it hardware or software, requires a thorough clinical validation in order to be adopted by the professional clinical community for use in patient care, such as diagnostics or treatment decisions, and must be approved by regulatory authorities [24]. The requirements for clinical validation will be more stringent when errors or mistakes can have greater consequences. In a clinical trial, it needs to be demonstrated how accurately the developed AI solution performs compared to the clinical standard (eg, sensitivity and specificity of a diagnostic test). Still, it is not completely clear whether good performance of an algorithm is acceptable if the solution is a "black box" and not transparent and rationally explainable [2]. On top of that, it is not obvious what proper validation of a continuous learning-based solution implies. An important issue is that because of lack of transparency, deep learning-based "black box" algorithms cannot be easily improved, in contrast to, for example, Bayesian models that are based on a transparent structure. Initial attempts to tackle this challenge are on the way [39].

Have AI-based solutions already been approved for clinical use? The earlier mentioned Watson for Oncology system operates as a "black box" and its advice could not be

clinically validated [26]. On the other hand, in 2017 it was claimed that the first deep learning-based algorithm, which identifies contours of cardiac ventricles from a magnetic resonance imaging (MRI) image to calculate ventricular volume, was validated and approved by the US Food and Drug Administration (FDA) for performing the calculation faster than a clinician [40]. Obviously, this system's scope is far more restricted than Watson's; the unimodal imaging data that were used were directly and causally related to the *ground truth* provided during every image analysis by the clinician. Also, it can be considered a measurement algorithm and does not include a clinical interpretation claim. Clinical validation and obtaining regulatory approval are much more difficult for those algorithms for which such an interpretation claim is added [4].

Several new solutions are ready or able to perform continuous (ie, incremental) learning [41]. However, within current regulations, an AI system for clinical applications should be "frozen" and can, therefore, not learn online and immediately apply its new knowledge. Rather, it needs to have an offline validation of the obtained "frozen" model on an independent series of patient or sample data. Following a next continuous-learning cycle, the validation process needs to be repeated again prior to renewed implementation of the model. Ideally, new clinically acceptable ways to shorten validation tracks for digital applications in a patient-safe manner should be found; it is expected that special procedures will be put in place to facilitate regulatory approval of updated algorithms. In line with this, the FDA is actively developing a strategy to deal with AI-based software solutions [42]. Maximal use of existing knowledge in transparent and causal model algorithms, as in Bayesian modeling, is expected to facilitate both clinical validation and obtaining regulatory approval, both for unimodal as well as for multimodal data.

The conclusion of this section is that procedures must be put in place to facilitate algorithms to be regulatory ready and to enable validation.

Methods to Use

Technology-wise, numerous methods from the domain of AI have been explored for the development of clinical applications [8,11,16,43]. Some have been more successful than others, mostly depending on application type. For automating pathology diagnosis using tissue slide images, deep learning has proven to be an appropriate technology. When dealing with more general multimodal problems, such as predicting clinical outcomes, patient assessments, and risk predictions, other methods that often include domain knowledge are likely to be more appropriate choices. Probabilistic methods using knowledge representation are increasingly used and enable reduction of the number of influencing variables, determining *sensible* features or latent variables. Probabilistic Bayesian modeling is well-suited to deal with complex biological (eg, "omics" data, such as genomics and transcriptomics data) as well as medical and clinical data; it is finding its way into diagnostic applications as well as drug development [9,44-49]. However, where knowledge is lacking, knowledge-agnostic AI approaches

become valuable; Bayesian reasoning networks are thought to have high potential for use in combination with deep learning, combining the best of two worlds in Bayesian deep learning [17,18,50].

The conclusion of this section is that the right AI method must be used for the problem.

Recommendations for Use of Artificial Intelligence Methods to Develop Clinical Applications: The Six Rs

In view of the challenges related to the use of AI for health care and biomedical applications, we believe it will be of value to have some guidelines when designing a study. They may also serve to facilitate communication between scientists involved in AI and medical doctors. From the discussion above, we have extracted six basic recommendations.

1. Relevant and well-defined clinical question first. Data analytics without domain knowledge can be applied in the health care domain, but at high risk of getting clinically irrelevant outcomes. For every new AI project, the clinical questions should be well-defined and reviewed with clinical experts. The outcome of the analysis should also be reviewed for clinical and/or biological sense.

2. Right data (ie, representative and of good quality). Carefully define the dataset that is needed to answer the clinical question. A clinical dataset with *ground truth* should be sufficiently clean and reliable. Be aware of hidden variation between samples that is not visible in the dataset. The dataset should be appropriate for the question at hand as well as representative for the population under study.

3. Ratio between number of patients and their variables should fit the AI method. To obtain useful results, ensure working with adequately large datasets (ie, numbers of patients or samples) for the AI method to be used, and reduce patient variables where possible. Use domain knowledge to limit spurious correlations.

4. Relationship between input variables and predicted output variable, as the dependent value, should be as direct and causal as possible. The clinical question should as closely as possible relate the *ground truth* to the data. Hence, finding new pathology features that best distinguish between two different pathology diagnoses can be successful; using lifestyle information to predict 10-year survival might not.

5. Regulatory ready; enabling validation. Upfront, consider how a certain solution can be validated and pass regulatory

requirements. Consider how using domain knowledge could speed up the validation process, for instance, by breaking up the AI system into smaller AI systems. This effectively excludes systems that iteratively change by continuous learning.

6. Right AI method. Use the right method for the question at hand. Data-driven methods can be used if the data available allows it, and knowledge-based methods can be applied if there is knowledge available but not enough data; a mixture of the two, combined in a wise manner, may be highly productive for development of clinically applicable health care solutions.

Privacy Issues

Driven by the *big data* analysis developments in health care, new privacy regulations were recently implemented in Europe —General Data Protection and Regulation (GDPR) [51]. To protect privacy, individuals control their own personal data, and explicit informed consent is required for access to the data and use in AI. This regulation is expected to make it more difficult to share patient data between multiple medical centers and with companies involved in development of AI solutions.

Key Takeaways

While AI approaches are excellently suited to develop algorithms for analysis of unimodal imaging data (eg, radiological or digital pathology images), for clinical (ie, patient-related) applications, major challenges lie in the usually limited patient or sample numbers (N). This is in comparison to the number of multimodal variables (P) due to patient variation, inadequate *ground truth* information, and a requirement for robust clinical validation prior to clinical implementation. Artificial Intelligence solutions that combine domain knowledge with data-driven approaches are, therefore, preferable over solutions that use only domain knowledge or are fully data driven. We introduce the following 6R model to keep in mind for AI projects in the biomedical and clinical health care domain:

1. Relevant and well-defined clinical question first.
2. Right data (ie, representative and of good quality).
3. Ratio between number of patients and their variables should fit the AI method.
4. Relationship between data and *ground truth* should be as direct and causal as possible.
5. Regulatory ready; enabling validation.
6. Right AI method.

Conflicts of Interest

All authors are regular employees of Royal Philips, Eindhoven, The Netherlands.

References

1. Chouard T, Venema L. Machine intelligence. Nature 2015 May 28;521(7553):435. [doi: 10.1038/521435a] [Medline: 26017441]

2. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan;25(1):30-36. [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]

3. Murdoch TB, Detsky AS. The inevitable application of big data to health care. JAMA 2013 Apr 03;309(13):1351-1352. [doi: 10.1001/jama.2013.393] [Medline: 23549579]

4. Dreyer KJ, Geis JR. When machines think: Radiology's next frontier. Radiology 2017 Dec;285(3):713-718. [doi: 10.1148/radiol.2017171183] [Medline: 29155639]

5. Kruse CS, Goswamy R, Raval Y, Marawi S. Challenges and opportunities of big data in health care: A systematic review. JMIR Med Inform 2016 Nov 21;4(4):e38 [FREE Full text] [doi: 10.2196/medinform.5359] [Medline: 27872036]

6. CB Insights. 2017 Feb 03. From virtual nurses to drug discovery: 106 artificial intelligence startups in healthcare URL:https://www.cbinsights.com/research/artificial-intelligence-startups-healthcare/ [accessed 2019-03-04] [WebCite Cache ID 76dBlzMnJ]

7. Itahashi K, Kondo S, Kubo T, Fujiwara Y, Kato M, Ichikawa H, et al. Evaluating clinical genome sequence analysis by Watson for Genomics. Front Med (Lausanne) 2018;5:305 [FREE Full text] [doi: 10.3389/fmed.2018.00305] [Medline: 30474028]

8. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]

9. Zarringhalam K, Enayetallah A, Reddy P, Ziemek D. Robust clinical outcome prediction based on Bayesian analysis of transcriptional profiles and prior causal networks. Bioinformatics 2014 Jun 15;30(12):i69-i77 [FREE Full text] [doi: 10.1093/bioinformatics/btu272] [Medline: 24932007]

10. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge, MA: MIT Press; 2016.

11. Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature 2015 May 28;521(7553):452-459. [doi: 10.1038/nature14541] [Medline: 26017444]

12. Williams BJ, Bottoms D, Treanor D. Future-proofing pathology: The case for clinical adoption of digital pathology. J Clin Pathol 2017 Dec;70(12):1010-1018. [doi: 10.1136/jclinpath-2017-204644] [Medline: 28780514]

13. Griffin J, Treanor D. Digital pathology in clinical use: Where are we now and what is holding us back? Histopathology 2017 Jan;70(1):134-145. [doi: 10.1111/his.12993] [Medline: 27960232]

14. Atasoy H, Greenwood BN, McCullough JS. The digitization of patient care: A review of the effects of electronic health records on health care quality and utilization. Annu Rev Public Health 2018 Dec 19;40:1. [doi: 10.1146/annurev-publhealth-040218-044206] [Medline: 30566385]

15. Raghupathi W, Raghupathi V. Big data analytics in healthcare: Promise and potential. Health Inf Sci Syst 2014;2:3 [FREE Full text] [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]

16. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]

17. Wang H, Yeung DY. Towards Bayesian deep learning: A framework and some existing methods. IEEE Trans Knowl Data Eng 2016 Dec 01;28(12):3395-3408 [FREE Full text] [doi: 10.1109/TKDE.2016.2606428]

18. Wang H, Yeung DY. arXiv. Ithaca, NY: arXiv; 2016 Apr 07. Towards Bayesian deep learning: A survey URL:https://arxiv.org/pdf/1604.01662 [accessed 2019-03-06] [WebCite Cache ID 76fuYVJOg]

19. Tizhoosh HR, Pantanowitz L. Artificial intelligence and digital pathology: Challenges and opportunities. J Pathol Inform 2018;9:38 [FREE Full text] [doi: 10.4103/jpi.jpi_53_18] [Medline: 30607305]

20. Robertson S, Azizpour H, Smith K, Hartman J. Digital image analysis in breast pathology: From image processing techniques to artificial intelligence. Transl Res 2018 Dec;194:19-35. [doi: 10.1016/j.trsl.2017.10.010] [Medline: 29175265]

21. Vink JP, Van Leeuwen MB, Van Deurzen CHM, De Haan G. Efficient nucleus detector in histopathology images. J Microsc 2013 Feb;249(2):124-135 [FREE Full text] [doi: 10.1111/jmi.12001] [Medline: 23252774]

22. Liu Y, Gadepalli K, Norouzi M, Dahl GE, Kohlberger T, Boyko A, et al. arXiv. Ithaca, NY: arXiv; 2017 Mar. Detecting cancer metastases on gigapixel pathology images URL:https://arxiv.org/pdf/1703.02442 [accessed 2019-03-06] [WebCite Cache ID 76fv5bego]

23. 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 Dec 02;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]

24. McShane LM, Cavenagh MM, Lively TG, Eberhard DA, Bigbee WL, Williams PM, et al. Criteria for the use of omics-based predictors in clinical trials. Nature 2013 Oct 17;502(7471):317-320 [FREE Full text] [doi: 10.1038/nature12564] [Medline: 24132288]

25. IBM Watson Health. URL:https://www.ibm.com/watson/health/ [accessed 2019-03-04] [WebCite Cache ID 76dDBgH6M]

26. Herper M. Forbes. 2017 Feb 19. MD Anderson benches IBM Watson in setback for artificial intelligence in medicine URL:https://www.forbes.com/sites/matthewherper/2017/02/19/md-anderson-benches-ibm-watson-in-setback-for-artificial-intelligence-in-medicine/ [accessed 2019-03-04] [WebCite Cache ID 76dDZaVQF]

27. Yaffe MB. The scientific drunk and the lamppost: Massive sequencing efforts in cancer discovery and treatment. Sci Signal 2013 Apr 02;6(269):pe13. [doi: 10.1126/scisignal.2003684] [Medline: 23550209]

28. Li H, Graesser AC, Cai Z. Comparison of Google translation with human translation. In: Proceedings of the Twenty-Seventh International Florida Artificial Intelligence Research Society Conference. Palo Alto, CA: AAAI Press; 2014 May 03 Presented at: Twenty-Seventh International Florida Artificial Intelligence Research Society Conference; May 21-23, 2014; Pensacola Beach, FL p. 190-195 URL:https://www.aaai.org/ocs/index.php/FLAIRS/FLAIRS14/paper/view/7864/7823

29. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. Science 2000 Dec 22;290(5500):2323-2326 [FREE Full text] [doi: 10.1126/science.290.5500.2323] [Medline: 11125150]

30. Fodor IK. A Survey of Dimension Reduction Techniques. Oak Ridge, TN: US Department of Energy; 2002 May 09. URL:https://www.osti.gov/servlets/purl/15002155 [accessed 2019-03-06] [WebCite Cache ID 76fyAJRY1]

31. Lakshmi Padmaja D, Vishnuvardhan B. Comparative study of feature subset selection methods for dimensionality reduction on scientific data. In: Proceedings of the 2016 IEEE 6th International Conference on Advanced Computing (IACC). New York, NY: IEEE; 2016 Presented at: 2016 IEEE 6th International Conference on Advanced Computing (IACC); February 27-28, 2016; Bhimavaram, India p. 31-34. [doi: 10.1109/IACC.2016.16]

32. Cooper GF, Bahar I, Becich MJ, Benos PV, Berg J, Espino JU, Center for Causal Discovery team. The center for causal discovery of biomedical knowledge from big data. J Am Med Inform Assoc 2015 Nov;22(6):1132-1136 [FREE Full text] [doi: 10.1093/jamia/ocv059] [Medline: 26138794]

33. Mills AM, Gradecki SE, Horton BJ, Blackwell R, Moskaluk CA, Mandell JW, et al. Diagnostic efficiency in digital pathology: A comparison of optical versus digital assessment in 510 surgical pathology cases. Am J Surg Pathol 2018 Jan;42(1):53-59. [doi: 10.1097/PAS.0000000000000930] [Medline: 28877052]

34. Goacher E, Randell R, Williams B, Treanor D. The diagnostic concordance of whole slide imaging and light microscopy: A systematic review. Arch Pathol Lab Med 2017 Jan;141(1):151-161. [doi: 10.5858/arpa.2016-0025-RA] [Medline: 27399211]

35. Elmore JG, Barnhill RL, Elder DE, Longton GM, Pepe MS, Reisch LM, et al. Pathologists' diagnosis of invasive melanoma and melanocytic proliferations: Observer accuracy and reproducibility study. BMJ 2017 Jun 28;357:j2813 [FREE Full text] [doi: 10.1136/bmj.j2813] [Medline: 28659278]

36. Calude CS, Longo G. The deluge of spurious correlations in big data. Found Sci 2017 Sep;22(3):595-612. [doi: 10.1007/s10699-016-9489-4]

37. Venet D, Dumont JE, Detours V. Most random gene expression signatures are significantly associated with breast cancer outcome. PLoS Comput Biol 2011 Oct;7(10):e1002240 [FREE Full text] [doi: 10.1371/journal.pcbi.1002240] [Medline: 22028643]

38. Gill CJ, Sabin L, Schmid CH. Why clinicians are natural Bayesians. BMJ 2005 May 07;330(7499):1080-1083 [FREE Full text] [doi: 10.1136/bmj.330.7499.1080] [Medline: 15879401]

39. Sussillo D, Barak O. Opening the black box: Low-dimensional dynamics in high-dimensional recurrent neural networks. Neural Comput 2013 Mar;25(3):626-649. [doi: 10.1162/NECO_a_00409] [Medline: 23272922]

40. Marr B. Forbes. 2017 Sep 20. First FDA approval for clinical cloud-based deep learning in healthcare URL:https://www.forbes.com/sites/bernardmarr/2017/01/20/first-fda-approval-for-clinical-cloud-based-deep-learning-in-healthcare/ [accessed 2019-03-04] [WebCite Cache ID 76dF5Q2HY]

41. Zhu L, Ikeda K, Pang S, Ban T, Sarrafzadeh A. Merging weighted SVMs for parallel incremental learning. Neural Netw 2018 Apr;100:25-38. [doi: 10.1016/j.neunet.2018.01.001] [Medline: 29432992]

42. US Food and Drug Administration. Digital Health Software Precertification (Pre-Cert) Program URL:https://www.fda.gov/MedicalDevices/DigitalHealth/UCM567265 [accessed 2019-03-04] [WebCite Cache ID 76dFDAef9]

43. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: Past, present and future. Stroke Vasc Neurol 2017 Dec;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]

44. Verhaegh W, Van de Stolpe A. Knowledge-based computational models. Oncotarget 2014 Jul 30;5(14):5196-5197 [FREE Full text] [doi: 10.18632/oncotarget.2276] [Medline: 25114037]

45. Verhaegh W, van Ooijen H, Inda MA, Hatzis P, Versteeg R, Smid M, et al. Selection of personalized patient therapy through the use of knowledge-based computational models that identify tumor-driving signal transduction pathways. Cancer Res 2014 Jun 01;74(11):2936-2945 [FREE Full text] [doi: 10.1158/0008-5472.CAN-13-2515] [Medline: 24695361]

46. van Ooijen H, Hornsveld M, Dam-de Veen C, Velter R, Dou M, Verhaegh W, et al. Assessment of functional phosphatidylinositol 3-kinase pathway activity in cancer tissue using forkhead box-O target gene expression in a knowledge-based computational model. Am J Pathol 2018 Sep;188(9):1956-1972 [FREE Full text] [doi: 10.1016/j.ajpath.2018.05.020] [Medline: 30030980]

47. van de Stolpe A, Holtzer L, van Ooijen H, de Inda MA, Verhaegh W. Enabling precision medicine by unravelling disease pathophysiology: Quantifying signal transduction pathway activity across cell and tissue types. Sci Rep 2019 Feb 07;9(1):1603 [FREE Full text] [doi: 10.1038/s41598-018-38179-x] [Medline: 30733525]

48. Zarringhalam K, Enayetallah A, Gutteridge A, Sidders B, Ziemek D. Molecular causes of transcriptional response: A Bayesian prior knowledge approach. Bioinformatics 2013 Dec 15;29(24):3167-3173 [FREE Full text] [doi: 10.1093/bioinformatics/btt557] [Medline: 24078682]

49.   Gupta SK. Use of Bayesian statistics in drug development: Advantages and challenges. Int J Appl Basic Med Res 2012 Jan;2(1):3-6 [FREE Full text] [doi: 10.4103/2229-516X.96789] [Medline: 23776799]

50.   van de Stolpe A, Kauffmann RH. Innovative human-specific investigational approaches to autoimmune disease. RSC Adv 2015;5(24):18451-18463. [doi: 10.1039/C4RA15794J]

51.   Official Journal of the European Union. Luxembourg: Publications Office of the European Union; 2016 May 04. Legislation URL:https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2016:119:FULL&from=EN [accessed 2019-03-06] [WebCite Cache ID 76g0PXfqh]

Abbreviations

AI: artificial intelligence
FDA: US Food and Drug Administration
GDPR: General Data Protection and Regulation
MRI: magnetic resonance imaging