



Commentary

Artificial Intelligence and Big Data in Public Health

Kurt Benke ^{1,2,*}  and Geza Benke ³

¹ School of Engineering, University of Melbourne, Parkville, Victoria 3010, Australia

² AgriBio, Centre for AgriBiosciences, State Government of Victoria, Bundoora, Victoria 3083, Australia

³ School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Rd, Melbourne 3004, Australia; geza.benke@monash.edu

* Correspondence: kbenke@unimelb.edu.au

Received: 30 August 2018; Accepted: 5 December 2018; Published: 10 December 2018



Abstract: Artificial intelligence and automation are topics dominating global discussions on the future of professional employment, societal change, and economic performance. In this paper, we describe fundamental concepts underlying AI and Big Data and their significance to public health. We highlight issues involved and describe the potential impacts and challenges to medical professionals and diagnosticians. The possible benefits of advanced data analytics and machine learning are described in the context of recently reported research. Problems are identified and discussed with respect to ethical issues and the future roles of professionals and specialists in the age of artificial intelligence.

Keywords: algorithms; Big Data; machine learning; deep learning; data mining; visualization; epidemiology; predictive analytics; precision medicine; vision; wearable AI

1. Introduction

Big Data is associated with the massive computational resources needed to cope with the increasing volume and complexity of data from many sources, such as the internet and remote sensor networks. Big Data includes information that is structured, semi-structured, or unstructured, and there may be complex interrelationships that are syntactic, semantic, social, cultural, economic, and organizational in nature. There may also be epistemic uncertainties that enter the data processing pipeline and hinder decision making by humans and computers, including (a) data corruption by noise and artefacts, (b) data entry errors, (c) duplicate records, (d) missing records, and (e) incomplete digital records relating to information on date, age, gender, or other variables.

The Big Data culture embraces cyber-physical systems, cloud computing, and the Internet of Things (IoT), also known as Industry 4.0. These massive data processing systems often involve significant levels of process automation. The current global focus on Big Data on the internet has been coupled with the rise of artificial intelligence (AI) in diagnostics and decision making following recent advances in computer technology. An early catalyst was the rapid growth of social media and the rapidly declining cost of microelectronics in connected devices.

In this paper, the fundamentals of AI and Big Data are introduced and described in the context of public health. Examples of potential benefits and critical comments are provided within the framework of recent research published in many prestigious journals such as *JAMA* and *Nature Reviews* [1–11]. The examples provided are illustrative but not exhaustive, and were chosen to cover a variety of interests in public health, such as epidemiology, precision medicine, medical screening, vision augmentation, and psychology. Questions are raised on ethical issues and the implications for the future roles of specialists and expert opinion. Possible answers are discussed and evaluated, and we argue the need to start the conversation on engagement with AI by professionals in the public health sector.

2. Medical Research

Big Data is often associated with large-scale macrosystems with distributed data processing that is often beyond the capability of local desktop computers and traditional software because of constraints imposed by speed and volume of processing. Information processing is diverse and may include the streaming of text messages, images, video, and music files. The following five characteristics of Big Data have been quoted in the past and can occur in different combinations (see also [1,12,13]):

- Variability (lack of structure, consistency, and context);
- Variety (includes audio files, imagery, numerical data, and text data);
- Velocity (real-time processing and very high speed of transmission);
- Veracity (accuracy, noise, and uncertainty in data);
- Volume (extremely large data sets).

The quantity and quality of Big Data are influenced by statistical replications in large-scale scientific experiments and are subject to many epistemic uncertainties. The most significant social impacts of Big Data are likely to occur when it is used in combination with artificial intelligence.

In epidemiology, the huge potential of Big Data was illustrated by a pioneering study reported by Deiner [1,13], who demonstrated that the spread of epidemics can be detected early by tracking online queries on disease symptoms using social media such as Google Search and Twitter. Pattern recognition and data analytics were used as tools for the detection, recognition, and classification of patterns of disease relating to the incidence of conjunctivitis. The study results suggested that early warning and detection of biosecurity threats and epidemics of influenza may be possible by the surveillance of online queries. Timeliness of detection and early intervention are obvious benefits subject to further validation by correlation with hard data, including electronic medical records, medication sales, and hospital admissions.

The accuracy of data from social media is affected by limited online access by minority groups and participation by age, gender, or ethnicity. Aside from Google search errors, Twitter queries may be compromised by ambiguities involving Boolean logic using If–Then–Else rules, resulting in incorrect predictions for disease incidence.

Such internet-based approaches have uncertainties that will need to be addressed in future research, such as the accuracy of search algorithms, updating rates, and the authenticity of data [1]. A “hot topic” in the mass media is the revelation that some governments and some data analytics companies have sought to influence public opinion by spreading so-called “fake news” and “fake” Twitter accounts. This problem may be addressed by the application of AI-based screening methods to check online Big Data. For example, a machine learning algorithm such as a neural network could be trained to detect certain recurring words, phrases, and views expressed in social media and then use pattern recognition to correlate the data with news content, candidate names, media outlets, and geographic locations. Suspicious correlations and clusters can be revealed in this manner.

Another example of a Big Data application is precision medicine, where clinical trials are based on patient selection according to DNA profiling, which provides biomarkers for targeted treatment, rather than a standard approach used for the whole population [2]. The selection of individuals sharing the same genetic abnormality for clinical trials leads to more precise drug development and more accurate treatments with existing drugs [2,3]. Additional advantages include reductions in sample size and reduced variability in clinical trials—without the need for a placebo. There is also the potential for genome sequencing to produce greater diagnostic sensitivity and more precise treatment [4]. Large-scale population studies and Big Data analytics can be used for more accurate genotypic and phenotypic data to support investigations of causality.

There are many uncertainties in the precision medicine approach, including whether genetic profiling is more effective than other factors, such as changes in lifestyle and exposure to environmental stressors [14]. Response to drug treatment is affected by diet, exercise, medication, alcohol, tobacco, and the microbiome in the digestive tract [5]. Information on these factors needs to be provided within

a comprehensive Big Data framework. Large-scale population studies are required for information discovery to resolve many of these uncertainties.

The development of expensive genetically-tailored medicines may benefit only a small minority of wealthy citizens. Would limited funding be better deployed on medical research that benefits the majority? The genetic basis of a disease in some cases is secondary in importance to environmental factors, and behavior modification relating to diet or exercise may be just as effective but less expensive. In addition, an expensive new drug developed from a clinical trial may be no more effective than a combination of existing inexpensive drugs. In an era of declining research funding, there is a case for data analytics and the investigation of costs vs. benefits and trade-offs associated with all areas of research.

Important challenges relating to the development of ethics for precision medicine include rights to ownership of genetic information, privacy, control of its dissemination, and potential consumer use or abuse. For example, unauthorized knowledge of genomic information can influence consumer risk perception in personnel selection processes or credit ratings. The DNA of an individual is not a password that can be changed. There is a need for greater debate on genetic privacy, such as finding the right balance between public rights to privacy and evidence required for law enforcement, and potential improvements in regulatory control [11,15].

3. Artificial Intelligence

Artificial intelligence is a generic term for a machine (or process) that responds to environmental stimuli (or new data) and then modifies its operation to maximize a performance index [16,17]. In practice, the learning process is implemented using mathematics, statistics, logic, and computer programming. The learning process enables the AI model to be trained on data in an iterative procedure with parameter adjustment by trial and error using reinforcement rules. The performance index may be the minimization of the difference between model predictions and experimental data.

The training phase is based on the classical scientific method, which has origins in antiquity with Aristotle and was formalized in the writings of Francis Bacon (circa. 1600). Once trained, the AI model can be applied to new data to make decisions. Decision making can involve detection, discrimination, and classification. This approach is known as supervised learning because it is analogous to an adaptive system with a feedback loop and a goal-directed learning process. Examples include the diagnosis of a disease from pathology results, or assigning the correct names to labeled images of different skin cancers. The potential role of AI in processing Big Data in precision medicine has been suggested by experts due to the massive resources needed for predictive analytics [18]. The potential of Big Data is most apparent when viewed as a database for artificial intelligence (Figure 1).

Some AI models carry out unsupervised learning by a clustering process, such as the k -means clustering algorithm, with the aim of discovering important groupings or defining features in the data [17]. A major challenge in the future development of AI is how to exploit Big Data to produce high-level abstractions that can simulate a human subjective response. In a clinical context, there is potential for the development of a digital expert that facilitates automatic conversion of pathology results into a written report or a verbal explanation.

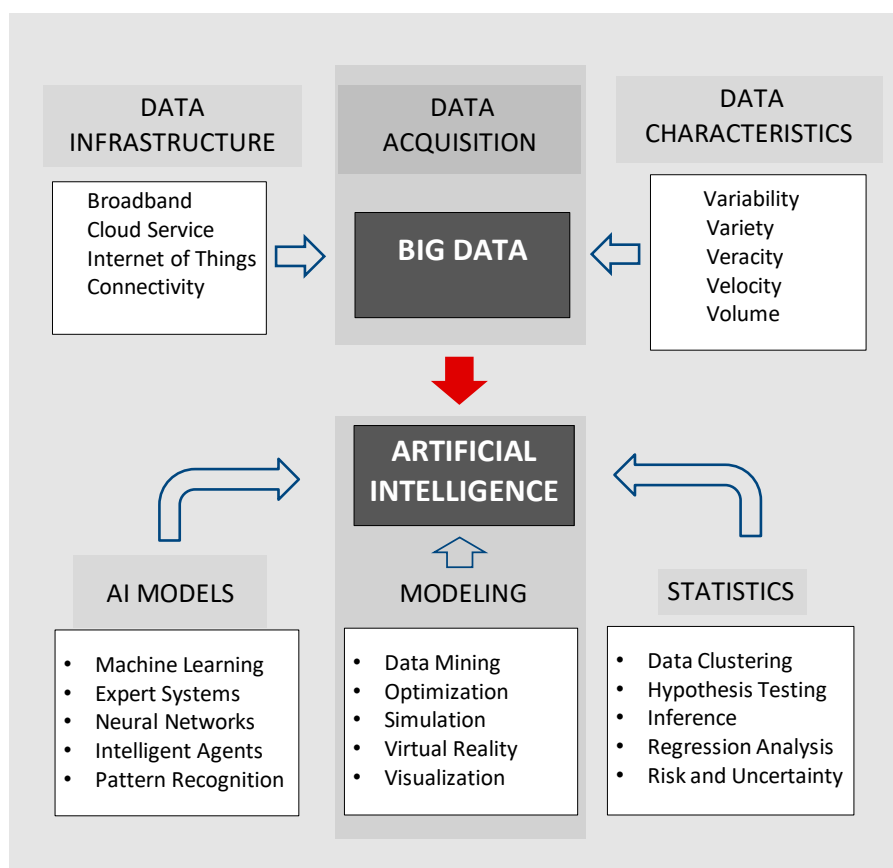


Figure 1. Artificial intelligence (AI) and Big Data.

Examples of AI approaches include machine learning, pattern recognition, expert systems, and fuzzy logic. Many AI models have strong statistical underpinnings and have learning rules based on nonlinear optimization procedures. The typical artificial neural network (ANN) consists of a predefined architecture with a matrix of weights that are adjusted iteratively according to the defined performance index [6,17].

4. Machine Learning

Machine learning is a subfield of AI that is based on training on new data using an adaptive approach, such as a neural network, but without explicit programming of new rules as required by other types of AI algorithms such as expert systems [16,17]. The coefficients (weights) of the model are adjusted iteratively, with repeated presentations of the data, until the performance index is optimized. The recent investigation by Gulshan et al. [6] reported the validation of a “deep learning” algorithm for the detection of diabetic retinopathy and diabetic macular edema (DME). The study produced high sensitivity (87%) and specificity (98%), which exceeded the performance of human specialists. The deep learning approach did not use explicit feature measurements, such as lines or edges in images, as in traditional pattern recognition. It was instead trained on a very large set of labeled images, and discrimination was based on using the entire image as a pattern (i.e., without explicit feature extraction), followed by validation on test images from a database. Very large training sets can help to compensate for image variability without the need for canonical transformations to normalize for various geometrical aberrations.

The deep learning algorithm consisted of a convolutional neural network with a very large number of hidden layers optimized using a multistart gradient descent procedure [19]. Very high processing speeds were achieved in the learning phase using graphics cards and parallel processing with thousands of images. The classification results were very impressive and have been echoed in

similar studies reviewed elsewhere, raising questions about the future roles of human specialists in medical diagnostics [7,20].

The results were deemed to be so significant that *JAMA* published two invited editorials together with a viewpoint article as commentaries on the machine learning procedure as a digital diagnostician [7–9]. Future research is still needed on performance in clinical settings and improvements in the labeling scheme used for training data, which was based on a majority vote from medical experts. The performance index could be improved by the use of more objective metrics for detection of DME, such as optical coherence tomography.

Jha and Topol [9] suggested that adaptation to the challenge of machine learning by medical specialists may be possible if they become generalists and information professionals. Humans could use AI-based automated image classification for screening and detection, but would manage and integrate the information in a clinical context, advise on additional testing, and provide guidance to clinicians. However, this potential symbiosis assumes that AI is restricted to pattern recognition and image classification using machine learning algorithms.

In practice, different AI models can also include systems that can apply logic, inference, strategy, and data integration. This is demonstrated by software such as IBM's "Deep Blue" and "Watson" and Google's "DeepMind", which have employed AI algorithms to defeat the best human experts at strategy games such as chess and GO and win the televised *Jeopardy* general knowledge quiz by accessing the entire knowledge base of Wikipedia. Together with software personal assistants such as Apple's "Siri" and Microsoft's "Cortana" (which provide conversational response and advice), there is a sense that specialized AI algorithms are evolving rapidly in sophistication. Recently, internet giants Amazon (with "Alexa") and Google (with "Assistant") have been reported to be seeking business partners for further development of the next generation of personal assistants, which will feature voice-activated platforms for interactive phone calls that imitate humans.

Human infants learn by a process of trial and error and acquiring a series of specialist skills by interaction and eventual combination (e.g., crawling, walking, and talking). In a similar manner, cooperation between specialist AI algorithms (including machine learning) could one day simulate the human cognitive process to produce generalized AI. For example, a suite of specialized programs could be controlled in a top-down manner by a master AI program with a standard set of rules and axioms (analogous to encoded DNA). Several companies such as Arria NLG, Retresco, and others have already demonstrated that AI and automatic language processing can convert diagnostic data into a final report and verbal counseling to replace a human expert.

The qualitative distinction between AI and human intelligence is arguably a matter of degree only, as measured by the famous Turing test. In the long term, AI may converge with and even surpass human judgment, given current rapid developments in software and hardware coupled with feedback loops incorporating multispectral IoT sensors and complexity theory. Progress in AI applications in recent years has been exponential, not linear, and is now ubiquitous and immersive in our lifestyles.

There is clearly a need for widespread discussion and debate on the future roles of human experts, diagnosticians, and medical specialists. The development of ethical standards is necessary for the application of AI in the medical sector to support regulations for the protection of human rights, user safety, and algorithmic transparency. Algorithms need to be free of bias in coding and data and should be explanatory if possible and not just predictive. Controlling the pace of automation would ensure incremental adoption, if necessary, with constant process evaluation [18].

5. Vision Augmentation

The OrCam portable device for artificial vision employs a miniature television camera mounted on the frame of spectacles for optical character recognition by a microchip [10,21]. Both text and patterns, including faces, can be converted by an algorithm into words heard through an earpiece. This is an example of an AI-based wearable device [22]. A recent evaluation of the OrCam for blind patients produced very encouraging results [10].

Device evaluation was based on a 10-item test that included reading a menu, reading a distant sign, and reading a headline in a newspaper. The score of the evaluation was the number of test items correctly vocalized. Because the items were not weighted by importance, the development of a weighting scheme would further facilitate comparison with similar devices, such as NuEyes [23]. A conventional eye chart could be used as a baseline.

The potential of such a noninvasive AI device for sensory substitution cannot be overemphasized. Even texture and color could be digitized and identified as words through the earpiece. In combination with GPS data and bionic vision [24], a blind person could acquire situational awareness and vision for navigation in outdoor locations. The performance of the OrCam device could be improved by adding text to city pavements with location information such as “road”, “crossing”, or “bus stop”. Text information could be rendered in low contrast to conform with environmental aesthetics. Adding text to walls and floors inside buildings would also enable navigation in new environments while providing continuous information, commentary, and environmental awareness.

6. Data Mining

In data science, passive data sets in stored archives are not very useful unless the information can be processed for decision making, forecasting, and data analytics. Data mining is a process for knowledge discovery that is used to find trends, patterns, correlations, anomalies, and features of interest in a database [25,26]. The results can be used for making decisions, making predictions, and conducting quantitative investigations. Data mining has been described as a combination of computer science, statistics, and database management with rapidly increasing utilization of AI techniques (e.g., machine learning) and visualization (with advanced graphics), especially in the case of Big Data analytics [26].

Common approaches include multivariate regression analysis and neural networks to construct predictive models. Additional approaches are listed in Figure 1, and many other techniques are described in literature reviews [26,27]. An example of a machine learning application would be to train a neural network to detect novel patterns in patient data and correlations between medications and adverse side effects by scanning electronic medical records for risk assessment [28]. Before data mining it may be necessary to “clean” the stored records with appropriate data cleaning software for the detection and correction of records that may be inaccurate, redundant, incomplete, or corrupt—which are typical epistemic uncertainties [29,30]. Software tools are available from various companies to support data quality and data auditing, including data management products designed for Big Data frameworks (see also Hadoop, Cloudera, MongoDB, and Taklend [30,31]).

Data mining may be enhanced by data visualization techniques that use human perception for information discovery in complex data sets showing no obvious patterns. Visualization enables the exploration and identification of trends, clusters, and artifacts and provides a starting point for detailed numerical analysis [32]. Techniques for visualization include 2D and 3D color graphics; bar graphs; maps; polygons with different shapes, sizes, colors, and textures; video data; and stereo imagery. An example in bioinformatics is the Microbial Genome Viewer, which is a web-based tool used for the interactive graphical combination of genomic data including chromosome wheels and linear genome maps from genome data stored in a MySQL database [33].

Basic concepts in data mining and visualization are demonstrated by the following small-scale example in the field of psychology [34]. The correlation between subjective ratings of fear and the appearance and perception of threat were explored in a pilot study involving a mixed-gender group of first-year university students in the age range of 16–43 years ($n = 31$, $F = 24$, $M = 7$). The subjects were requested to rank images of 29 animals and insects on a three-point scale (1 = not; 2 = quite; 3 = very) for both appearance and perception of threat. A standard multiple linear regression (MLR) model was fitted to the data in the original study, with fear as the dependent variable, and revealed that the weighting for perception of threat was 0.85 vs. 0.15 for appearance of threat. In the current paper, a standard feed-forward backpropagation artificial neural network (ANN) was trained on the

data using commercial software, noting that the very limited statistical data would produce indicative results sufficient for illustration [35]. Both MLR and ANN produced correlation coefficients $r > 0.90$ on training data, thus demonstrating the potential of this approach for predictive analytics. For improved confidence, a future study would need more extensive experiments and greater statistical power using Big Data.

For data visualization, the original tabulated data was modeled using a hybrid 2D polynomial function $z = f(x,y)$ by regression analysis, depicted as a response surface in Figure 2, which was produced using commercial software [36]. An alternative function approximation is possible by ANN but was not used on this occasion due to the small sample size. This would have involved a least-squares error regression analysis with model predictions plotted incrementally along each axis. In the present case, data visualization was accomplished by color coding (where the numerical scale was mapped linearly onto a color scale) followed by a 90 Deg clockwise rotation, revealing two hidden areas of zero-fear response under certain combinations of appearance and perception of threat (e.g., animals that are perceived to be dangerous but are dead or disabled).

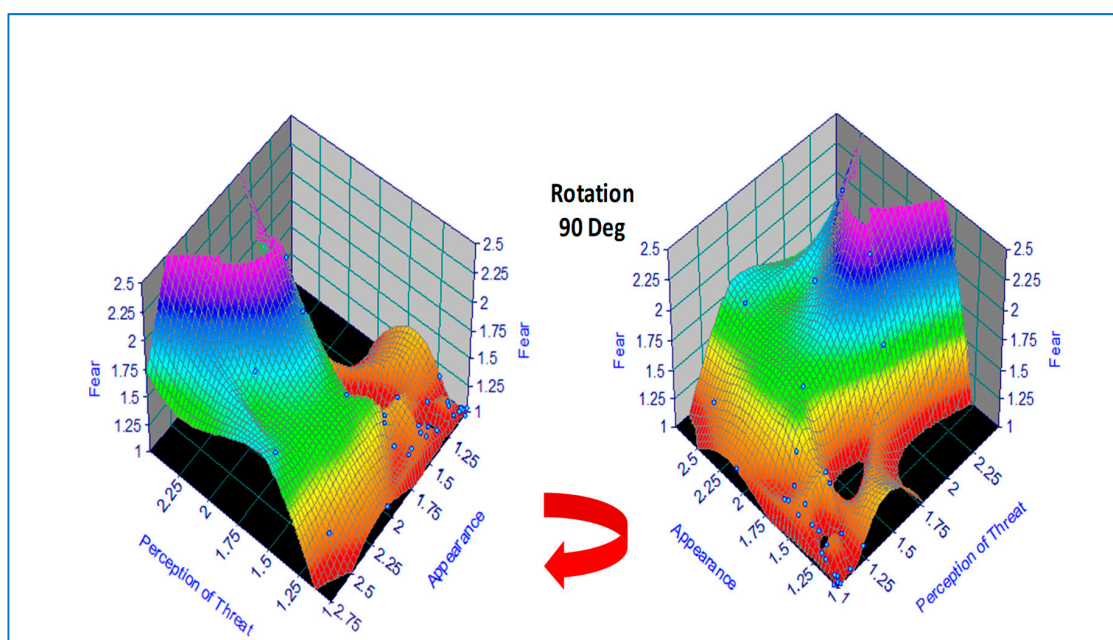


Figure 2. Illustrative example of a response surface for fear as a function of perception of threat and appearance of threat. Visualization reveals discontinuities (two black holes in the right image) exposing regions of zero-fear response under certain combinations of appearance and perception of threat.

Visualization in color and 3D geometry can therefore increase the probability of information discovery in a database or spreadsheet. The huge volume of information that is characteristic of Big Data suggests that there is potential for greater use in data mining using AI for automation and visualization for initial trend detection. Percepts and gestalts are characteristic of human vision and provide additional tools to augment the operation of data mining.

7. Conclusions

The fundamentals of AI and Big Data have been introduced and described in the context of potential applications to data analytics in public health and the medical sciences. The advent of AI and Big Data will change the nature of routine screening in specialist diagnostics and perhaps even in general practice and routine pathology [8–10,19,37]. A deeper understanding of these new technologies is required for public health and policy development by bureaucrats, politicians, and business leaders. The important issues of ethics and the need for an overarching regulatory framework for AI and precision medicine have not been sufficiently addressed in public health, and need further attention.

These same issues will also have increasing prominence in the public discourse on autonomous cars and battlefield robots.

The combination of AI and Big Data has the potential to have profound effects on our future. The role of the medical specialist will be challenged as these technologies become more widespread and integrated. There is not yet a vigorous debate about the future roles of specialists and diagnosticians if specialist AI programs can perform with greater sensitivity and specificity. Will AI eventually take over for humans in medical diagnostics? Will the future role of the human specialist move to case management, thus leaving screening, detection, and diagnostics to become the domain of intelligent machines?

It has been argued that these questions may be answered within a framework that allows human experts to occupy new roles as generalists and information specialists [9,10]. For example, AI and automation could be used as a front-end for diagnostics involving the interpretation of pathology results and image classification. This would leave humans free to manage and integrate the information in a clinical context, advise on additional testing if necessary, and provide continuing guidance to clinicians and patients. This approach is feasible in the short term and medium term with specialist AI, but will this approach suffice in the longer term with the rise of generalized AI?

Author Contributions: K.B. provided concepts and information and wrote the manuscript. G.B. contributed to concepts, writing, and editing. All authors read and approved the final manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare that they have no competing interests.

References

1. Benke, K.K. Uncertainties in Big Data when using Internet Surveillance Tools and Social Media for determination of Patterns in Disease Incidence. *JAMA Ophthalmol.* **2017**, *135*, 402. [CrossRef]
2. Rubin, R.A. Precision Medicine Approach to Clinical Trials. *JAMA* **2016**, *316*, 1953–1955. [CrossRef] [PubMed]
3. Rubin, R. Precision Medicine: The Future or Simply Politics? *JAMA* **2015**, *313*, 1089–1091. [CrossRef] [PubMed]
4. Ashley, E.A. Towards Precision Medicine. *Nat. Rev. Genet.* **2016**, *17*, 507–522. [CrossRef] [PubMed]
5. Kyriacos, D.N.; Lewis, R.J. Confounding by Indication in Clinical Research. *JAMA* **2016**, *316*, 1818–1819. [CrossRef] [PubMed]
6. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.C.; Wu, D.; Narayanaswamy, A.; Venugopalan, S.; Widner, K.; Madams, T.; Cuadros, J.; et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **2016**, *316*, 2402–2410. [CrossRef]
7. Wong, T.Y.; Bressler, N.M. Artificial Intelligence with Deep Learning Technology Looks into Diabetic Retinopathy Screening. *JAMA* **2016**, *316*, 2366–2367. [CrossRef]
8. Beam, A.L.; Kohane, I.S. Translating Artificial Intelligence into Clinical Care. *JAMA* **2016**, *316*, 2368–2369. [CrossRef]
9. Jha, S.; Topol, E.J. Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists. *JAMA* **2016**, *316*, 2353–2354. [CrossRef]
10. Moisseiev, E.; Mannis, M.J. Evaluation of a Portable Artificial Vision Device among Patients with Low Vision. *JAMA Ophthalmol.* **2016**, *134*, 748–752. [CrossRef]
11. Lunshof, J.E.; Chadwick, R.; Vorhaus, D.B.; Church, G.M. From genetic privacy to open consent. *Nat. Rev. Genet.* **2008**, *9*, 406–411. [CrossRef]
12. SAS. Big Data—What It Is and Why It Matters. Available online: http://www.sas.com/en_us/insights/big-data/what-is-big-data.html (accessed on 6 September 2017).
13. Deiner, M.S.; Lietman, T.M.; McLeod, D.; Chodosh, J.; Porco, T.C. Surveillance tools emerging from search engines and social media data for determining eye disease patterns. *JAMA Ophthalmol.* **2016**, *134*, 1024–1029. [CrossRef] [PubMed]

14. Powledge, T.M. That ‘Precision Medicine’ Initiative? A Reality Check. Genetic Literacy Project. Available online: <https://www.geneticliteracyproject.org/2015/02/03/that-precision-medicine-initiative-a-reality-check/> (accessed on 7 November 2016).
15. Chadwick, R.; Levitt, M.; Shickle, D. (Eds.) *The Right to Know and the Right Not to Know: Genetic Privacy and Responsibility*; Cambridge University Press: Cambridge, UK, 2014.
16. Russell, S.J.; Norvig, P. *Artificial Intelligence: A Modern Approach*; Prentice Hall: Englewood Cliffs, NJ, USA, 1995.
17. Nilsson, N.J. *Principles of Artificial Intelligence*; Morgan Kaufmann: San Francisco, CA, USA, 2014.
18. Bertalan, M. The role of artificial intelligence in precision medicine. *Expert Rev. Precis. Med. Drug Dev.* **2017**, *2*, 239–241.
19. Dean, J.; Corrado, G.; Monga, R.; Chen, K.; Devin, M.; Mao, M.; Senior, A.; Tucker, P.; Yang, K.; Le, Q.V.; et al. Large Scale Distributed Deep Networks. *Adv. Neural Inf. Process. Syst.* **2012**, 1223–1231.
20. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)] [[PubMed](#)]
21. OrCam. See for Yourself. Available online: <http://www.orcam.com> (accessed on 29 August 2016).
22. Moshtael, H.; Aslam, T.; Underwood, I.; Dhillon, B. High tech aids low vision: A review of image processing for the visually impaired. *Trans. Vis. Sci. Technol.* **2015**, *4*, 6. [[CrossRef](#)] [[PubMed](#)]
23. Greget, M.; Moore, J. NuEyes. Available online: <https://nueyes.com/product/> (accessed on 2 September 2016).
24. Ong, J.M.; da Cruz, L. The Bionic Eye: A Review. *Clin. Exp. Ophthalmol.* **2012**, *40*, 6–17. [[CrossRef](#)]
25. Han, J.; Pei, J.; Kamber, M. *Data Mining: Concepts and Techniques*; Elsevier: New York, NY, USA, 2011.
26. Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. *Data Mining: Practical Machine Learning Tools and Techniques*; Morgan Kaufmann: San Francisco, CA, USA, 2016.
27. Banks, D.L. Statistical Data Mining. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 9–25. [[CrossRef](#)]
28. Jensen, P.B.; Jensen, L.J.; Brunak, S. Mining electronic health records: Towards better research applications and clinical care. *Nat. Rev. Genet.* **2012**, *13*, 395–405. [[CrossRef](#)]
29. Brzozek, C.; Benke, K.; Zeleke, B.; Abramson, M.; Benke, G. Radiofrequency Electromagnetic Radiation and Memory Performance: Sources of Uncertainty in Epidemiological Cohort Studies. *Int. J. Environ. Res. Public Health* **2018**, *15*, 592. [[CrossRef](#)]
30. Rahm, E.; Do, H. Data Cleaning: Problems and Current Approaches. *IEEE Data Eng. Bull.* **2000**, *23*, 3–13.
31. Hernández, M.A.; Stolfo, S.J. Real-world data is dirty: Data cleansing and the merge/purge problem. *Data Min. Knowl. Discov.* **1998**, *2*, 9–37.
32. Fayyad, U.M.; Wierse, A.; Grinstein, G.G. (Eds.) *Information Visualization in Data Mining and Knowledge Discovery*; Morgan Kaufmann: San Diego, CA, USA, 2002.
33. Kerkhoven, R.; Van Enckevort, F.H.; Boekhorst, J.; Molenaar, D.; Siezen, R.J. Visualization for genomics: The microbial genome viewer. *Bioinformatics* **2004**, *20*, 1812–1814. [[CrossRef](#)] [[PubMed](#)]
34. Russell, R.A.; Russell, J.R.; Benke, K.K. *Subjective Factors in Combat Simulation: Correlation between Fear and the Perception of Threat*; DSTO-TR-0410; Department of Defence: Canberra, Australia, 1996.
35. Palisade. Decision Tools Suite Ver. 7.5, Palisade Corporation. 2018. Available online: www.palisade.com (accessed on 10 September 2018).
36. SYSTAT Software Inc. TableCurve 3D Ver. 4. 2002. Available online: systatsoftware.com (accessed on 10 September 2018).
37. Benke, K.E.; Benke, K.K. Automation of diagnostics by new disruptive technologies supports local general practice and medical screening in the third world. *Aust. Med. J.* **2015**, *8*, 174–177. [[CrossRef](#)] [[PubMed](#)]

