Literature analysis of artificial intelligence in biomedicine

Tim Hulsen^

Department of Hospital Services & Informatics, Philips Research, Eindhoven, The Netherlands

Correspondence to: Tim Hulsen. Department of Hospital Services & Informatics, Philips Research, High Tech Campus 34, 5656AE Eindhoven, The Netherlands. Email: tim.hulsen@philips.com.

Abstract: Artificial intelligence (AI) refers to the simulation of human intelligence in machines, using machine learning (ML), deep learning (DL) and neural networks (NNs). AI enables machines to learn from experience and perform human-like tasks. The field of AI research has been developing fast over the past five to ten years, due to the rise of 'big data' and increasing computing power. In the medical area, AI can be used to improve diagnosis, prognosis, treatment, surgery, drug discovery, or for other applications. Therefore, both academia and industry are investing a lot in AI. This review investigates the biomedical literature (in the PubMed and Embase databases) by looking at bibliographical data, observing trends over time and occurrences of keywords. Some observations are made: AI has been growing exponentially over the past few years; it is used mostly for diagnosis; COVID-19 is already in the top-3 of diseases studied using AI; China, the United States, South Korea, the United Kingdom and Canada are publishing the most articles in AI research; Stanford University is the world's leading university in AI research; and convolutional NNs are by far the most popular DL algorithms at this moment. These trends could be studied in more detail, by studying more literature databases or by including patent databases. More advanced analyses could be used to predict in which direction AI will develop over the coming years. The expectation is that AI will keep on growing, in spite of stricter privacy laws, more need for standardization, bias in the data, and the need for building trust.

Introduction

Artificial Intelligence (AI) refers to the simulation of human intelligence in machines. It includes techniques such as machine learning (ML), deep learning (DL) and neural networks (NN). AI makes it possible for machines to learn from experience and perform human-like tasks (1). The term 'artificial intelligence' was first used by John McCarthy at a workshop at Dartmouth College in 1956 (2). Progress was slow at first, until in the late 1990s and early 2000s, AI began to be used for logistics, data mining, medical diagnosis and other areas (3). Over the past 5 to 10 years this field has been developing fast due to computing power, especially in areas where large datasets are available to be mined, such as astronomy (4) or meteorology (5). Biomedicine is such an area as well: many 'big data' projects are taking place (6), with an abundance of clinical data from electronic health record (EHR) systems (7), genomics data from whole genome sequencing (WGS) studies (8), and digital imaging data from magnetic resonance (MR) (9), ultrasound (US) (10) scanners and digital pathology systems (11). This kind of 'wide' (i.e., many parameters) as well as 'long' (i.e., many subjects)

^ ORCID: 0000-0002-0208-8443.

data (12) is particularly useful for AI algorithms, which need a lot of diverse data to avoid overfitting. Therefore, AI has many interesting applications in the medical area, whether it is for diagnosis, prognosis, treatment, surgery, drug discovery, or for other applications. For example, last year an antibiotic (halicin) was discovered using a DL algorithm (13). Robots and telemedicine are being used to keep health workers safe during the COVID-19 pandemic (14). And people with mental health issues can use therapeutic chatbots (15). AI offers a broad range of possibilities in both academia and industry. Top hospitals such as the Mayo Clinic and the Cleveland Clinic (16) as well as large health-tech companies such as Philips (17), Siemens (18) and GE (19) invest a lot of time and money in data science and AI. Biotech companies such as BERG and pharma companies such as Takeda and are also using AI in their biomarker discovery (20) and drug discovery (21) pipelines. Google (by acquiring Fitbit) and Apple invest in AI in wearables that can track the health and lifestyle of the customers (22). Microsoft has built a cloud service especially for storing and sharing healthcare data (23), just like Amazon (24) and Google (25). Facebook has developed a tool called Preventive Health which connects people to health resources and checkup recommendations from health organizations (26). This means that all five of the American "Big Tech" or "GAFAM" companies now have extended their products and services into the healthcare domain. And these products and service all include AI in some form, changing healthcare in a way that has never been seen before. Some excellent reviews have been written already about how AI is changing medicine and healthcare (27-30). In this review we will take another approach: we will look at the current status of AI using bibliographical data from the biomedical literature. We will look at more generic questions: what trends can be seen over time? What countries are leading the way? What universities publish the most AI papers? Which AI algorithms are most popular over time? We will also investigate in which ways AI is applied in the biomedical domain: what diseases are studied using AI? What application areas are being researched? What drugs are studied with AI? What drug or device manufacturers are using AI? We will also place the results in context: what do these trends mean and what developments could cause AI to stop growing at the current rate?

Methodology

The following sections will present several figures based on analysis of biomedical literature data. The first section (containing *Figure 1*) uses the PubMed database; the other sections (containing *Figures 2-9*) use the Embase database. All search queries were carried out on October 10, 2021. *Table 1* shows the search queries that were done to obtain the data presented in the figures.

AI over the years

AI has, for a long time, been a promise of the future. When looking at the occurrence of the search term 'artificial intelligence' in the PubMed database (*Figure 1*), there is a steady rise visible from 1980 to a plateau in the period 2005–2008, followed by a few years of decline. This decline might have been caused by the limited computation power at that time compared to now. According to Moore's law (32), researchers nowadays have 2^7=128 more computation power available than in 2007. In the 2000s, researchers could see the potential of AI, and were performing research in that area, but were still limited in its use. Since 2018, the occurrence of the search term 'artificial intelligence' shows a steep incline again. Apparently, computers are now fast enough to really make use of AI and to fulfill its promise. Related search terms such as 'machine learning', 'deep learning' and 'neural network' show an incline as well, especially since 2016.

Diseases studied with AI

When we look into the Embase database (Excerpta Medica database; a biomedical and pharmacological bibliographic database) and search for 'artificial intelligence', we get 48,141 results. When searching for 'machine learning', we get 76,765 results. Embase offers the 'explosion' option, which includes all narrower (more specific) terms in the search query, so that they do not have to be entered individually (33). The search term "'machine learning'/exp OR 'artificial intelligence'/exp" gives 289,055 results. The full list of search terms can be found using Emtree search (34,35). Embase also offers the option to look at the diseases mentioned in the publications. The top-20 of results is displayed in *Figure 2*. This shows that COVID-19 is already on third place, with 'breast cancer' on first place and 'alzheimer disease' on second place. There are 9 oncological diseases in the top-20, which might point to the complexity and heterogeneity of cancers, which need DL algorithms for proper classification. There are also 4 brain disorders, 3 diabetic diseases and 2 cardiovascular diseases, as well as
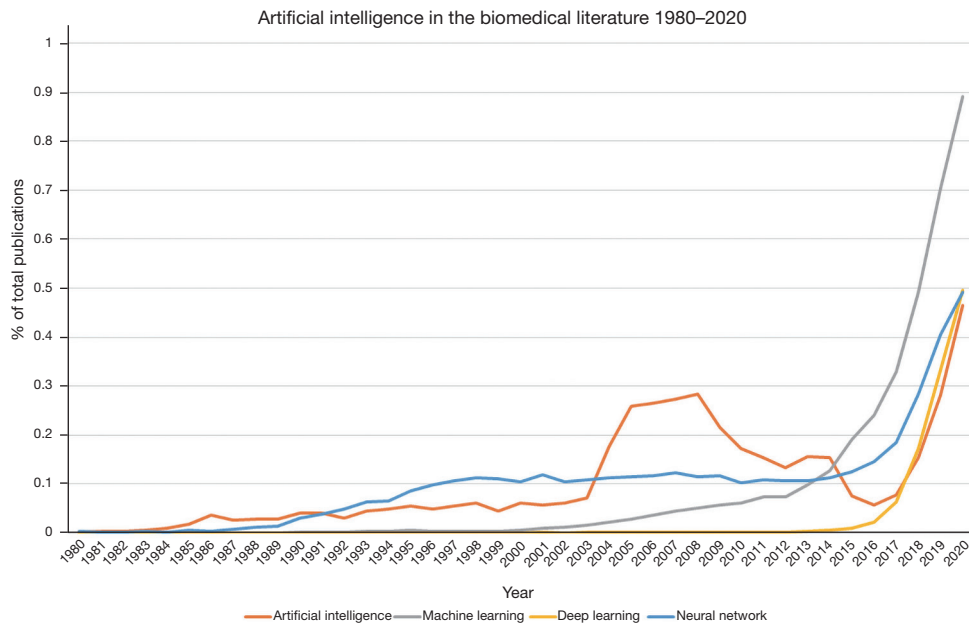
Figure 1 Occurrence of four search terms related to artificial intelligence in the titles and/or abstracts in the PubMed database, as a percentage of the total of publications, for the period 1980–2020. Generated by the R package 'easyPubMed' (31).
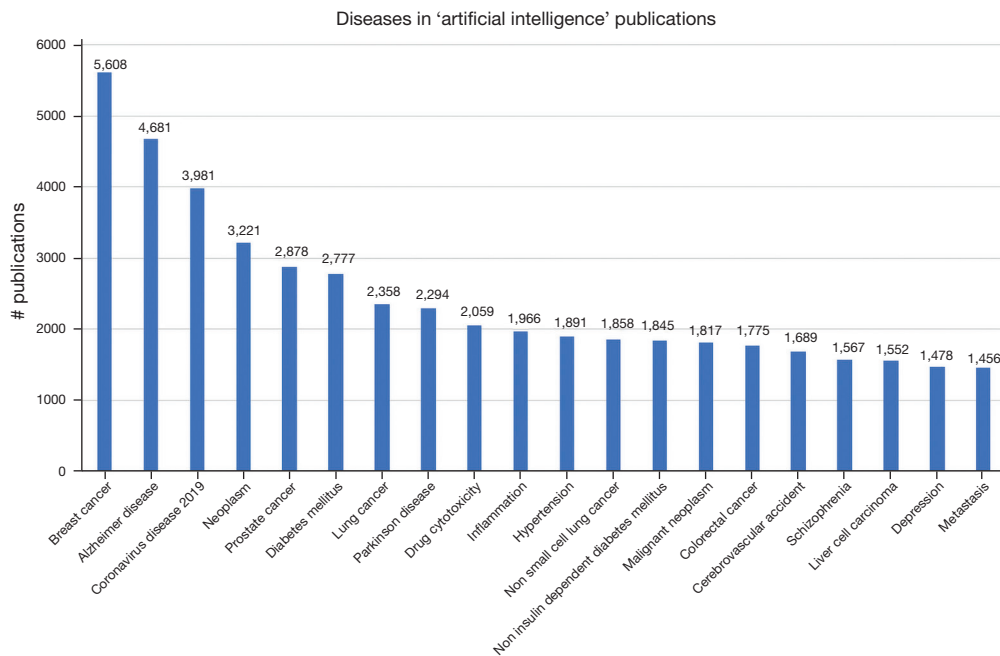

Figure 2 Top 20 of diseases found in publications in the Embase database (until 2021-10-10) which also contain terms related to 'artificial intelligence' and/or 'machine learning'.
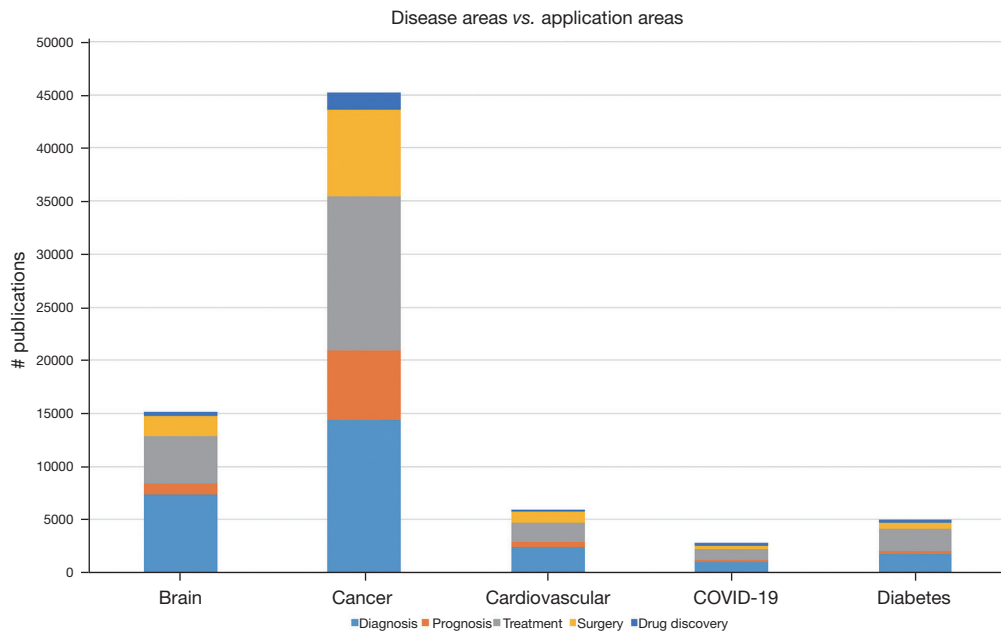
Ann Transl Med 2022;10(23):1284 | https://dx.doi.org/10.21037/atm-2022-50

Figure 3 Occurrences of disease areas in combination with application areas, together with the search term "'artificial intelligence'/exp OR 'machine learning'/exp", in the Embase database (date of search query 2021-10-10).
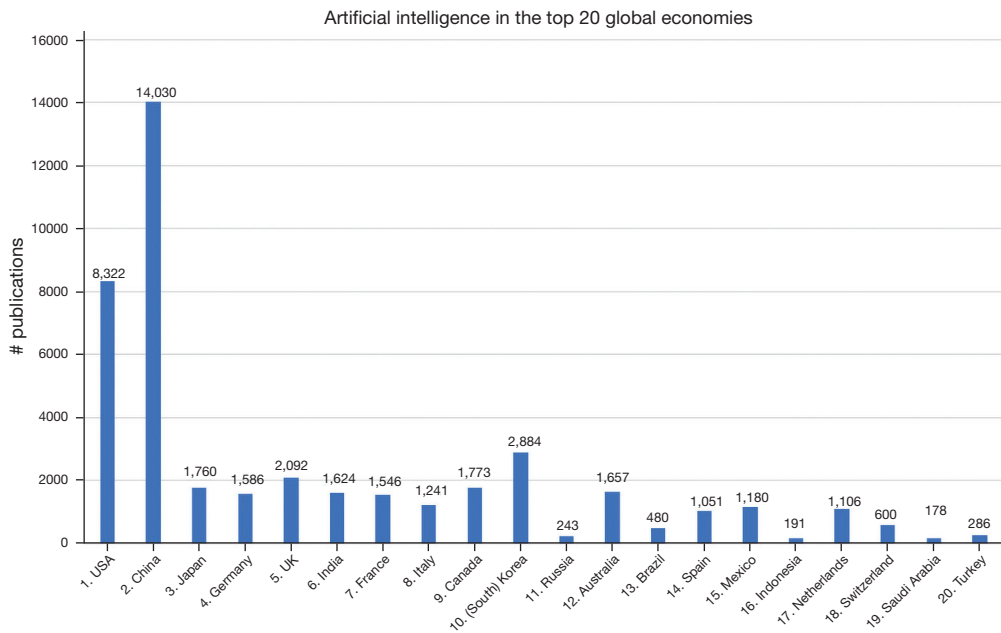

Figure 4 Occurrence of the search term "'artificial intelligence'/exp OR 'machine learning'/exp" in combination with any of the top 20 global economies in the 'affiliation' field, in the Embase database (date of search query 2021-10-10).
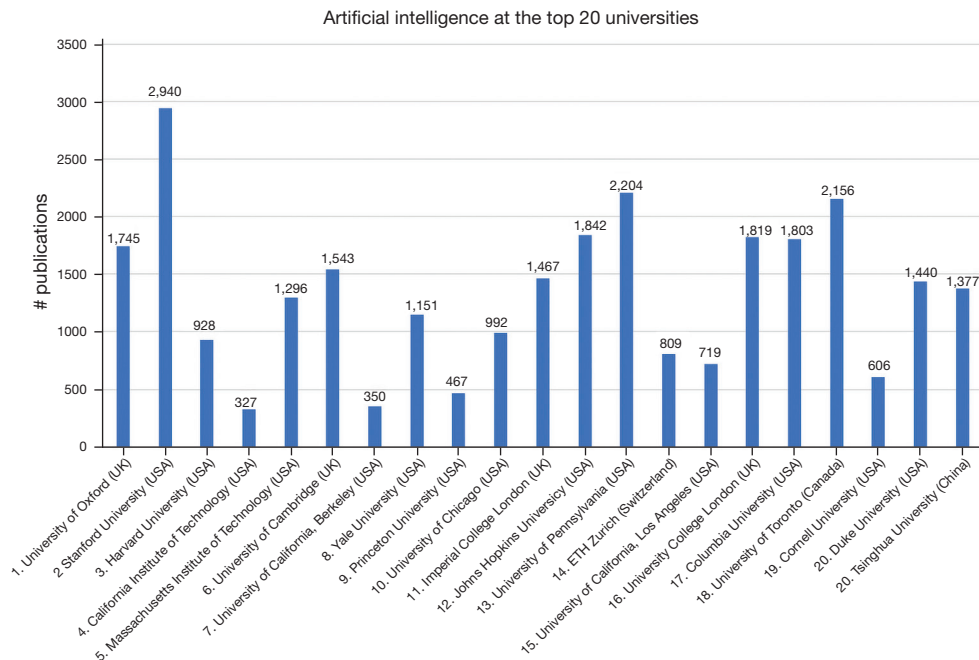
Figure 5 Occurrence of the search term "'artificial intelligence'/exp OR 'machine learning'/exp" in combination with any of the top 20 universities in the 'affiliation' field, in the Embase database (date of search query 2021-10-10).
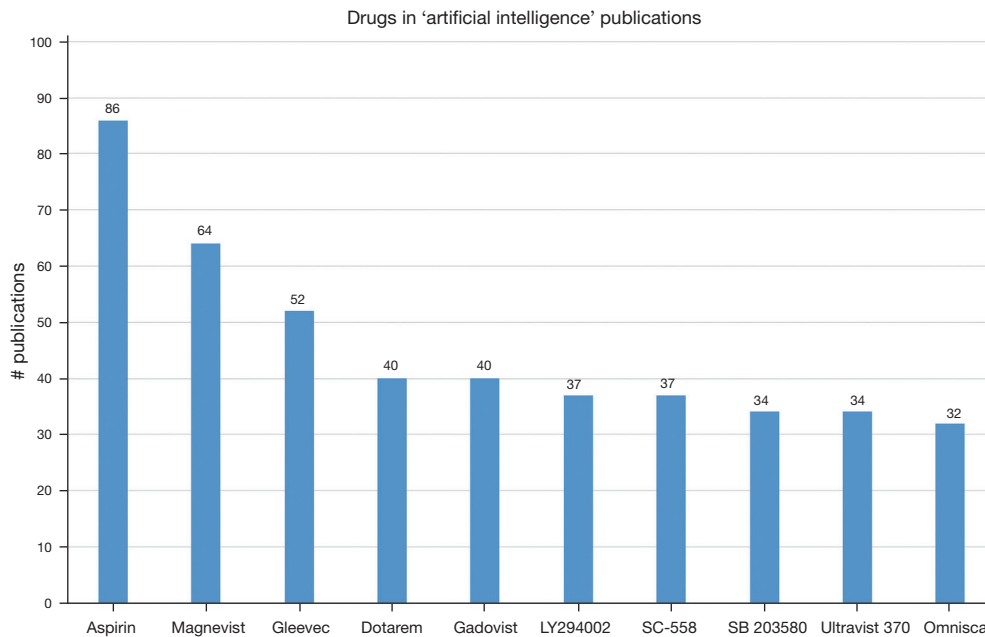

Figure 6 Top 10 of drug trade names found in publications in the Embase database (until 2021-10-10) which also contain the term "'artificial intelligence'/exp OR 'machine learning'/exp".

Drug manufacturers in 'artificial intelligence' publications

Figure 7 Top 10 of drug manufacturers found in publications in the Embase database (until 2021-10-10) which also contain the term "'artificial intelligence'/exp OR 'machine learning'/exp".

Device manufacturers in 'artificial intelligence' publications

Figure 8 Top 20 of device manufacturers found in publications in the Embase database (until 2021-10-10) which also contain the term "'artificial intelligence'/exp OR 'machine learning'/exp".
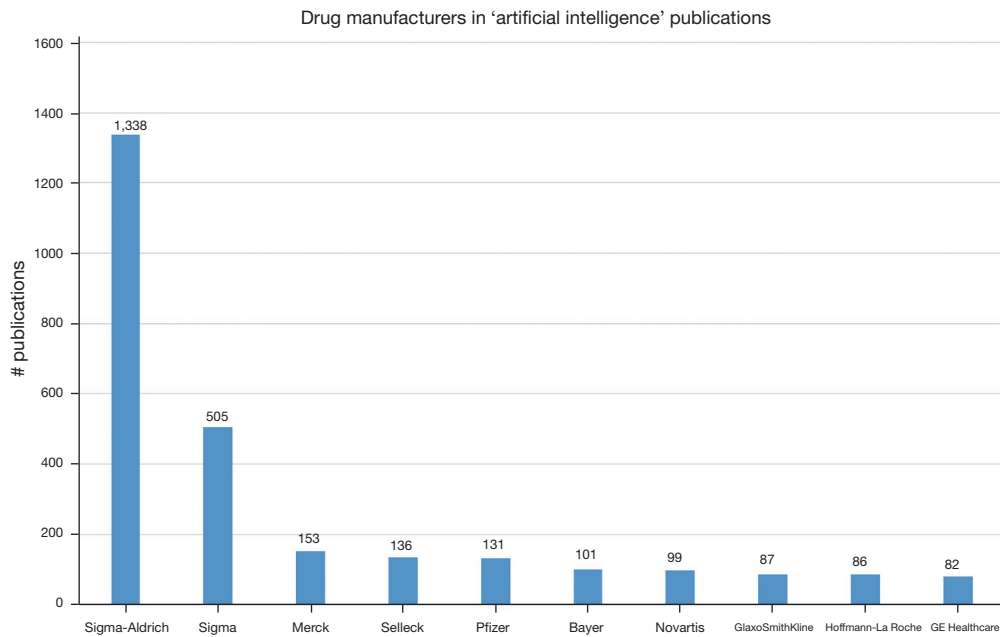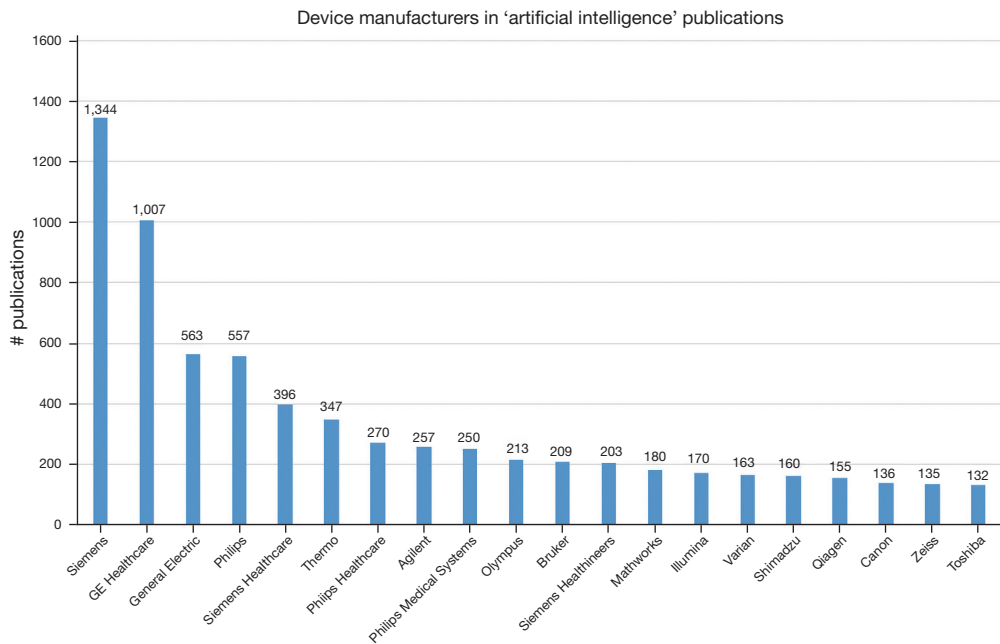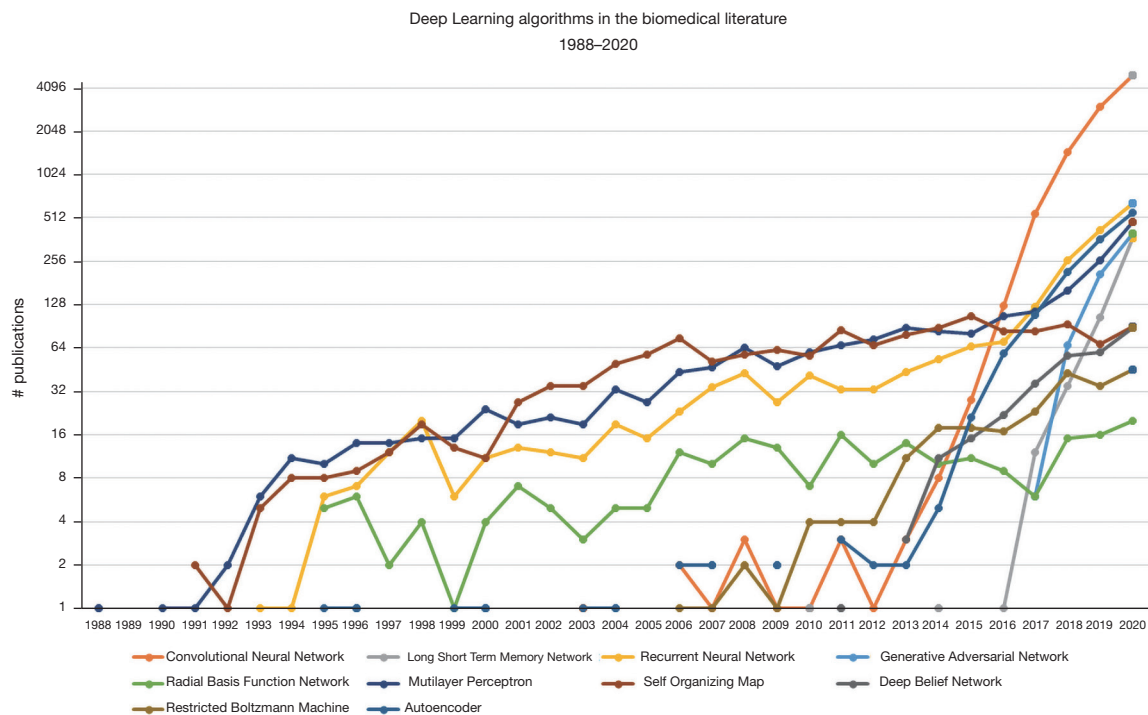
Figure 9 Occurrence of ten deep learning algorithms in the titles and/or abstracts in the Embase database. The vertical axis is displayed on a base-2 log scale.

drug cytotoxicity and inflammation.

Disease areas and application areas using AI

Embase offers more functionality than just the study of disease areas. We can also look at disease areas in relation to application areas such as diagnosis, prognosis, treatment, surgery and drug discovery. Diagnosis {according to MeSH [Medical Subject Headings (36)]} is "the determination of the nature of a disease or condition, or the distinguishing of one disease or condition from another. Assessment may be made through physical examination, laboratory tests, or the likes. Computerized programs may be used to enhance the decision-making process" (37). Prognosis is "a prediction of the probable outcome of a disease based on an individual's condition and the usual course of the disease as seen in similar situations" (38). *Figure 3* shows the main five disease areas derived from the analysis of *Figure 2*, subdivided into these 5 application areas. We used the 'explosion' option here as well, including all narrower terms. AI seems to be used mostly within the field of oncology. In all disease areas, there seem to be more AI applications for diagnosis than

for the other application areas. Publications around brain diseases are relatively often related to diagnosis (48.7% of the total), whereas publications around cancer are relatively often related to treatment (32.3% of the total). Similar relationships can be found between diabetes and treatment (42.4%), cardiovascular diseases and diagnosis (40.8%) and COVID-19 and treatment (38.3%).

Countries performing research in AI

Another possibility with the Embase database is to find out which countries perform the most research in the field of AI. *Figure 4* shows all combinations of the search term 'artificial intelligence' and/or 'machine learning' with each of the names of the top-20 global economies in the 'affiliation' field. Not surprisingly, the top spots are taken by China and the USA (the #2 and #1 economies), followed at a distance by South Korea, the UK and Canada (the #10, #5 and #9 economies). The bottom spots are taken by Saudi Arabia (178 publications) and Indonesia (191 publications). These findings are not much different from the Nature Index analysis of 2020 (39), in which the USA and China

Table 1 Search queries used to create each figure in this manuscript

Figure	Database	Queries	Variable(s)
1	PubMed	'artificial intelligence' AND <year> [dp]	<year> = 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020
		'machine learning' AND <year> [dp]	
		'deep learning' AND <year> [dp]	
		'neural network' AND <year> [dp]	
2	Embase	'artificial intelligence'/exp OR 'machine learning'/exp	–
		Embase menu → 'Diseases'	
3	Embase	('artificial intelligence'/exp OR 'machine learning'/exp) AND '<disease_area>' AND '<application_area>'	<disease_area> = brain, cancer, cardiovascular, COVID-19, diabetes <application_area> = prognosis, diagnosis, treatment, surgery, drug discovery
4	Embase	('artificial intelligence'/exp OR 'machine learning'/exp) AND '<country>':ff	<country> = USA, China, Japan, Germany, UK, India, France, Italy, Canada, Korea, Russia, Australia, Brazil, Spain, Mexico, Indonesia, Netherlands, Switzerland, Saudi Arabia, Turkey
5	Embase	('artificial intelligence'/exp OR 'machine learning'/exp) AND '<university>':ff	<university> = University of Oxford, Stanford University, Harvard University, California Institute of Technology, Massachusetts Institute of Technology, University of Cambridge, University of California Berkeley, Yale University, Princeton University, University of Chicago, Imperial College London, Johns Hopkins University, University of Pennsylvania, ETH Zurich, University of California Los Angeles, University College London, Columbia University, University of Toronto, Cornell University, Duke University, Tsinghua University
6	Embase	'artificial intelligence'/exp OR 'machine learning'/exp	–
		Embase menu → 'Drug trade names'	
7	Embase	'artificial intelligence'/exp OR 'machine learning'/exp	–
		Embase menu → 'Drug manufacturers'	
8	Embase	'artificial intelligence'/exp OR 'machine learning'/exp	–
		Embase menu → 'Device manufacturers'	
9	Embase	'convolutional neural network'	–
		'long short term memory network'	
		'recurrent neural network'	
		'generative adversarial network'	
		'radial basis function network'	
		'multilayer perceptron'	
		'self organizing map'	
		'deep belief network'	
		'restricted boltzmann machine'	
		'autoencoder'	
		Embase menu → 'Publication years'	

are leading the way as well. There are some differences though: in the Nature Index, the USA is #1 and China is #2. Moreover, the #3 and #5 spot are taken by Germany and Japan, whereas they are only #9 and #6 in Embase. This might be caused by the smaller number of journals covered by the Nature Index analysis (only 82 top journals), the shorter time period (2015–2019) and/or a different query used to search for AI publications.

Universities performing research in AI

Besides studying AI research within countries, we can also investigate which universities perform the most research in the field of AI. *Figure 5* shows all combinations of artificial intelligence' and/or 'machine learning' with each of the names of the top-20 universities [taken from the Times Higher Education World University Rankings (40)] in the 'affiliation' field. The #1 spot is taken by Stanford University with 2940 publications. Indeed, Stanford University is the employer of some of the world's AI experts such as Andrew Ng, who is also one of the co-founders of Coursera and DeepLearning.ai (41). Second place is for the University of Pennsylvania with 2,204 publications, which has a large AI research community (42). Third place is for the University of Toronto (UofT) with 2,156 publications. UofT is located in the region with the highest concentration of AI startups in the world, and has started the University of Toronto Artificial Intelligence (UofT AI) Group (43). The world's top university according to this list, the University of Oxford, is only on the seventh spot when it comes to AI. The California Institute of Technology (Caltech) has only 327 publications related to 'artificial intelligence' and/or 'machine learning'.

Drugs studied with AI

Embase also includes drug trade names in their database. *Figure 6* shows the prevalence of these drug names together with the search term 'artificial intelligence' or 'machine learning'. The number one spot is for Aspirin, the popular medication used to reduce pain, fever and inflammation. The number two spot is for Magnevist, a contrast agent for MRI. The number three spot is for Gleevec, a chemotherapy medication used to treat cancer. Dotarem, Gadovist and Omniscan are all MRI contrast agents. LY294002 is a PI3-kinase inhibitor. SC-558 is a selective inhibitor of COX2. SB 203580 is a specific inhibitor of p38α and p38β. Ultravist 370 is a contrast medium for X-rays. It

should be noted that overall numbers for drug trade name data in Embase are quite low.

Drug manufacturers using AI

We can also have a look at the manufacturing companies of the drugs mentioned in the previous section (*Figure 7*). This shows us that Sigma-Aldrich, currently known as MilliporeSigma and part of Merck, is on top of the list. The number 2, Sigma, is MilliporeSigma's main biochemical supplier. MilliporeSigma's chemicals are apparently being used a lot in scientific research; in search queries without 'artificial intelligence', 'Sigma' and 'Sigma-Aldrich' are on top of the list as well. Sigma is followed by larger companies such as Merck, Pfizer, Bayer, Novartis, GlaxoSmithKline (GSK), Hoffmann La-Roche and GE Healthcare. Another company in the top-10 is Selleck (#4), an American chemicals supplier. Some of the largest pharma companies are missing on this list, such as Johnson & Johnson, Sanofi, AbbVie and Takeda.

Device manufacturers using AI

When looking at device manufacturers using AI in biomedical publications (*Figure 8*), Siemens takes the top spot with 1,344 publications. Siemens is also on the #5 ('Siemens Healthcare') and #12 ('Siemens Healthineers') spot, and even occurs eleven times in the complete list. GE takes the #2 ('GE Healthcare') and #3 ('General Electric') spots and is mentioned sixteen times in total. Philips takes the #4 ('Philips'), #7 ('Philips Healthcare') and #9 ('Philips Medical Systems') place and is included in the list for ten times. Other companies include genomics companies such as Illumina and Qiagen, but also imaging companies such as Olympus, Canon and Toshiba. Thermo Fisher Scientific Agilent, Bruker and Shimadzu are producers of scientific instrumentation and software. Zeiss is an Ophthalmology company. Varian Medical Systems is a radiation oncology treatments and software maker. MathWorks is the creator of software such as MATLAB and Simulink, which support data analysis and simulation.

AI algorithms

The Embase database also enables us to have a look at the use of several DL algorithms over time. We adapted the top-10 list found at (44) and checked their occurrence in biomedical publications. These 10 algorithms are:

(I) Convolutional Neural Network (CNN, or ConvNet): a DL algorithm which can take in an input image, assign importance to various aspects/objects in the image and be able to differentiate one from the other. The architecture of a CNN is analogous to the neurons in the human brain. The first ever CNN was the 'Neocognitron' by Kunihiko Fukushima in 1980 (45);

(II) Long Short-Term Memory Network (LSTM): an artificial recurrent NN architecture (see #3) with feedback connections. It was proposed by Hochreiter and Schmidhuber in 1997 (46);

(III) Recurrent Neural Network (RNN): a class of artificial NNs where connections between nodes form a directed graph along a temporal sequence. This allows it to exhibit temporal dynamic behavior. RNNs are based on David Rumelhart's work from 1986 (47);

(IV) Generative Adversarial Network (GAN): a class of machine learning frameworks designed by Ian Goodfellow *et al.* in 2014 (48). Two NNs contest with each other in a zero-sum game. Given a training set, this technique learns to generate new data with the same statistics as the training set.

(V) Radial Basis Function Network (RBFN): an artificial NN that uses radial basis functions as activation functions. The network's output is a linear combination of radial basis functions of the inputs and neuron parameters. RBFNs were introduced by Broomhead and Lowe in 1988 (49);

(VI) Multilayer Perceptron (MLP): a class of feedforward artificial NN, composed of multiple layers of perceptrons. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron, which was invented by Frank Rosenblatt as early as 1957 (50);

(VII) Self-Organizing Map (SOM): a type of artificial NN that is trained using unsupervised learning to produce a low-dimensional, discretized representation of the input space of the training samples (called a map) and is therefore a method to do dimensionality reduction. Invented by Teuvo Kohonen in 1982 (51);

(VIII) Deep Belief Network (DBN): a generative graphical model, or alternatively a class of deep NN, composed of multiple layers of latent variables, with connections between the layers but not between units within each layer. DBNs were invented by Larochelle *et al.* in 2007 (52);

(IX) Restricted Boltzmann Machine (RBM): a generative stochastic artificial NN that can learn a probability distribution over its set of inputs. RBMs were initially invented under the name 'Harmonium' by Paul Smolensky in 1986 (53). Named after the Austrian physicist and philosopher Ludwig Boltzmann.

(X) Autoencoder: an unsupervised artificial NN that learns how to efficiently compress and encode data, and then learns how to reconstruct the data back from the reduced encoded representation to a representation that resembles the original input as much as possible. Autoencoders were introduced by Rumelhart *et al.* in 1986 (54), just like RNNs.

The first occurrence of any of the terms in the Embase database was 'Multilayer Perceptron' (MLP) in 1988, in the paper 'Auto-association by multilayer perceptrons and singular value decomposition' (55) by Bourlard & Kamp from the Philips Research Laboratory in Brussels, Belgium. The following years there was some use of RNNs, RBFNs, MLPs and SOMs, but never more than 100 publications per year in total. Only since 2015 has the use of these DL algorithms really taken off, which can be seen in *Figure 9*. Especially CNNs are getting very popular, with 5,023 publications in 2020, followed by RNNs (643 publications in 2020). CNNs are mostly used for image analysis and classification (56). RNNs can be used for predictions based on time series (longitudinal data) (57). Autoencoders are on third place in 2020 with 553 publications. MLPs are still popular as well (482 publications in 2020). GANs have only been around in the biomedical literature since 2017, despite its first publication in 2014 (48). They can be used to create synthetic data based on real data (58), which is particularly useful in cases where the real data cannot be analyzed because of privacy regulations. GANs (and its application to create synthetic data) are gaining popularity rapidly since the enforcement of the GDPR in May 2018.

Discussion & conclusions

This review shows the wide range of AI in biomedicine, and it shows the exponential rate at which it is growing. AI seems to be used mostly for diagnosis. COVID-19 is already in the top-3 of diseases studied using AI, which shows the large influence the COVID-19 pandemic has on medical science (59). China, the United States, South Korea, the United Kingdom and Canada are publishing the

most articles in AI research. China, USA and UK are not surprising because these countries host many of the world's top universities, but South Korea and Canada appear to be very productive in AI research as well. Stanford University is the world's leading university in AI research at this moment. This literature review also shows that CNNs are by far the most popular DL algorithms at this moment. Possible reasons for this are the wide range of applications (image classification and segmentation, object detection, video processing, natural language processing, speech recognition, etc.) and the ability of CNNs to do feature extraction (60). This review only shows results from the PubMed and Embase databases, which is a (large) subset of the complete biomedical literature. PubMed and Embase include mostly English publications, and they do not contain patent data. Therefore, trends and findings mentioned in this review could be studied in more detail, by studying more literature databases and by including patent databases. More advanced analyses could be used to predict in which direction AI will develop over the coming years. Now that computers are fast enough to execute complex algorithms, AI is expected to keep on growing. Heterogeneous diseases such as cancer can profit enormously from DL algorithms to properly stratify patient cohorts (61), and drugs and biomarkers will be found much more easily using AI (62). AI will therefore be a crucial component in enabling precision medicine. AI-powered algorithms have been met with enthusiasm by the general population (63), partly because it enables a 4P (Predictive, Preventive, Personalized, and Participatory) (64) model of medicine. However, there are also recent developments that might hinder the rise of AI. For example, stricter privacy regulations such as the GDPR (65) make it more difficult to store and analyze personal data, as it requires researchers to state the purpose of the data collection upfront, gives data subjects the right to withdraw their data at any time, etc. A possible solution here is the use of a federated setup, in which data stay at the source, AI algorithms are run locally on the data, and only the results are then brought together (66). An example of this is the Personal Health Train initiative in the Netherlands. Such a federated setup can be combined with multiparty homomorphic encryption, which is a synthesis between two novel advanced privacy-enhancing technologies: homomorphic encryption and secure multiparty computation (67). This will provide a mathematical guarantee of privacy, ensuring that data can be considered anonymized. Moreover, federated learning is especially useful for the study of rare diseases, where each separate source might have too little cases to setup a study

with enough statistical power (68). Recently, the European Commission has also announced a set of regulations around the use of AI (69), defining some high-risk application areas of AI, under which some medical areas such as robot-assisted surgery. These high-risk application areas need to conform to a number of strict obligations around data quality, security, documentation, etcetera. Increasing use of AI also means that data should be machine-readable: data should adhere to standards around reproducibility and interoperability, such as the FAIR Guiding Principles (70) and other guidelines (71). Data should be standardized using medical ontologies and vocabularies. Another issue that might affect the use of AI in biomedicine and healthcare is bias (72): the output of an algorithm is shaped by the data that is fed into it. Therefore, if this data comes from only white males in the range of 50–65 years old, the algorithm will not create reliable results for Asian females in the range of 35–50 years old. Initiatives where data from all over the world (from both genders, many ethnicities, and many age groups) are combined into a central database produce results that are less biased. For example, in the field of prostate cancer such projects are Movember GAP3 [(73), combining 25 different cohorts] and PIONEER [(74), with 32 public and private partners]. Next to privacy regulations, machine-readability and bias, scientists should also work on building trust in AI. For non-experts, AI often seems like a black box where a lot of data go in and through some 'black magic' something comes out. This is only a logical effect of the mathematical complexity of most AI algorithms. This complexity also makes it difficult for AI algorithms to be approved by regulatory instances such as the Food and Drug Administration (FDA) and the European Medicines Agency (EMA) (75). Possible solutions lie in explaining AI in a simple way (76), and also in anthropomorphizing AI algorithms and robots (77). Finally, it is important to realize that AI is not meant to completely replace clinicians (at least, not in the near future), but to help them. Clinicians should be supported by AI (78,79) in patient-centric solutions that can collaborate to reach precise diagnoses and optimal treatment pathways.

Acknowledgments

Footnote

Provenance and Peer Review: This article was commissioned

by the editorial office, *Annals of Translational Medicine* for the series "Big Data in Precision Medicine". The article has undergone external peer review.

Peer Review File: Available at https://atm.amegroups.com/article/view/10.21037/atm-2022-50/prf

Conflicts of Interest: The author has completed the ICMJE uniform disclosure form (available at https://atm.amegroups.com/article/view/10.21037/atm-2022-50/coif). The series "Big Data in Precision Medicine" was commissioned by the editorial office without any funding or sponsorship. The author is an employee of Philips Research; as an expert in the area of big data and precision medicine, the author writes publications, book chapters and patents, presents at conferences and serves as guest editor for scientific journals; and served as an unpaid Guest Editor of the series. The author has no other conflicts of interest to declare.

Ethical Statement: The author is accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Negnevitsky M. Artificial intelligence: a guide to intelligent systems. Pearson education; 2005.
2. Solomonoff RJ. The time scale of artificial intelligence: Reflections on social effects. Human Systems Management 1985;5:149-53.
3. Russell S, Norvig P. Artificial Intelligence - A modern approach. Prentice Hall Upper Saddle River. NJ, USA, 2002.
4. Chen Y, Kong R, Kong L. 14 - Applications of artificial intelligence in astronomical big data. In: Kong L, Huang T, Zhu Y et al., editors. Big Data in Astronomy. Elsevier; 2020. p. 347-75.
5. McGovern A, Elmore KL, Gagne DJ, et al. Using Artificial Intelligence to Improve Real-Time Decision-Making for High-Impact Weather. Bulletin of the American Meteorological Society 2017;98:2073-90.
6. Hulsen T, Jamuar SS, Moody AR, et al. From Big Data to Precision Medicine. Front Med (Lausanne) 2019;6:34.
7. Ross MK, Wei W, Ohno-Machado L. "Big data" and the electronic health record. Yearb Med Inform 2014;9:97-104.
8. Suwinski P, Ong C, Ling MHT, et al. Advancing Personalized Medicine Through the Application of Whole Exome Sequencing and Big Data Analytics. Front Genet 2019;10:49.
9. Tahmassebi A, Gandomi AH, McCann I, et al. Deep learning in medical imaging: fmri big data analysis via convolutional neural networks. Proceedings of the Practice and Experience on Advanced Research Computing. 2018. p. 1-4.
10. Van Sloun RJ, Cohen R, Eldar YC. Deep learning in ultrasound imaging. Proceedings of the IEEE 2019;108:11-29.
11. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: Challenges and opportunities. Med Image Anal 2016;33:170-5.
12. Hulsen T. Challenges and Solutions for Big Data in Personalized Healthcare. In: Moustafa A, editor. Big Data in Psychiatry, Neurology, and Personalized healthcare. 1 ed.: Elsevier; 2021.
13. Stokes JM, Yang K, Swanson K, et al. A Deep Learning Approach to Antibiotic Discovery. Cell 2020;181:475-83.
14. Yang G, Lv H, Zhang Z, et al. Keep Healthcare Workers Safe: Application of Teleoperated Robot in Isolation Ward for COVID-19 Prevention and Control. Chinese Journal of Mechanical Engineering 2020;33:47.
15. Yang M. Painful conversations: Therapeutic chatbots and public capacities. Communication and the Public 2020:2057047320950636.
16. Sennaar K. How America's 5 Top Hospitals are Using Machine Learning Today. Emerj. 2020. Available online: https://emerj.com/ai-sector-overviews/top-5-hospitals-using-machine-learning/, accessed 2021/05/26.
17. Philips. Artificial Intelligence. 2021. Available online: https://www.philips.com/a-w/about/artificial-intelligence.html, accessed 2021/05/26.
18. Healthineers S. Artificial intelligence in healthcare. 2021. Available online: https://www.siemens-healthineers.com/digital-health-solutions/artificial-intelligence-in-healthcare, accessed 2021/05/26.
19. GE. Edison - Intelligence to enable better patient care. 2021. Available online: https://www.gehealthcare.com/

products/edison, accessed 2021/05/26.

20. Tolstikov V, Moser AJ, Sarangarajan R, et al. Current Status of Metabolomic Biomarker Discovery: Impact of Study Design and Demographic Characteristics. Metabolites 2020;10:224.

21. Smalley E. AI-powered drug discovery captures pharma interest. Nat Biotechnol 2017;35:604-5.

22. Sutner S. Google, Fitbit, startups storm into healthcare AI. 2021. Available online: https://www.limeade.com/en/press/google-fitbit-startups-storm-into-healthcare-ai/, accessed 2021/05/26.

23. Microsoft. Microsoft Cloud for Healthcare. 2021. Available online: https://www.microsoft.com/en-us/industry/health/microsoft-cloud-for-healthcare, accessed 2021/05/26.

24. Amazon. Healthcare & Life Sciences. 2021. Available online: https://aws.amazon.com/health/, accessed 2021/05/26.

25. Google. Google Cloud for healthcare and life sciences. 2021. Available online: https://cloud.google.com/solutions/healthcare-life-sciences, accessed 2021/05/26.

26. Facebook. Preventive Health. 2021. Available online: https://preventivehealth.facebook.com/, accessed 2021/05/26.

27. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019;6:94-8.

28. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017;2:230-43.

29. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018;2:719-31.

30. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. Artificial Intelligence in Healthcare. Elsevier, 2020:25-60.

31. Fantini D. easyPubMed: Search and Retrieve Scientific Publication Records from PubMed. 2019. Available online: https://CRAN.R-project.org/package=easyPubMed, accessed 2021/05/26.

32. Moore GE. Cramming more components onto integrated circuits. New York: McGraw-Hill, 1965.

33. Embase. What is Emtree and how do I use it? 2020. Available online: https://service.elsevier.com/app/answers/detail/a_id/16389/supporthub/embase/kw/explosion/, accessed 2021/05/26.

34. Embase. Emtree 'machine learning'. 2021. Available online: https://www.embase.com/#emtreeSearch/search/51317::machine%20learning

35. Embase. Emtree 'artificial intelligence'. 2021. Available online: https://www.embase.com/#emtreeSearch/search/10089::artificial%20intelligence

36. Lipscomb CE. Medical Subject Headings (MeSH). Bull Med Libr Assoc 2000;88:265-6.

37. NCBI. MeSH term 'Diagnosis'. 2021. Available online: https://www.ncbi.nlm.nih.gov/mesh/?term=diagnosis

38. Baston A, Gerhardt C, Zandi S, et al. Visual Outcome after Intravitreal Anti-VEGF Therapy for Macular Neovascularisation Secondary to Sorsby's Fundus Dystrophy: A Systematic Review. J Clin Med 2021;10:2433.

39. Nature Index. The race to the top among the world's leaders in artificial intelligence. 2020. Available online: https://www.nature.com/articles/d41586-020-03409-8, accessed 2021/05/26.

40. Times Higher Education. World University Rankings 2021. 2021. Available online: https://www.timeshighereducation.com/world-university-rankings/2021/world-ranking, accessed 2021/05/26.

41. Coursera. Andrew Ng. 2021. Available online: https://www.coursera.org/instructor/andrewng, accessed 2021/05/26.

42. Pennsylvania Uo. AI@Penn. 2021. Available online: https://aipenn.org/.

43. Toronto Uo. University of Toronto Artificial Intelligence Group. 2021. Available online: https://www.uoft.ai/, accessed 2021/05/26.

44. Biswal A. Top 10 Deep Learning Algorithms You Should Know in 2021. Available online: https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-algorithm, accessed 2021/05/26.

45. Fukushima K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern 1980;36:193-202.

46. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9:1735-80.

47. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature 1986;323:533-6.

48. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. arXiv preprint. arXiv:14062661 2014.

49. Broomhead DS, Lowe D. Radial basis functions, multi-variable functional interpolation and adaptive networks: Royal Signals and Radar Establishment Malvern (United Kingdom), 1988.

50. Rosenblatt F. The perceptron, a perceiving and

recognizing automaton Project Para. Cornell Aeronautical Laboratory, 1957.

51. Kohonen T. Self-organized formation of topologically correct feature maps. Biological Cybernetics 1982;43:59-69.

52. Larochelle H, Erhan D, Courville A, et al. editors. An empirical evaluation of deep architectures on problems with many factors of variation. Proceedings of the 24th international conference on Machine learning; 2007.

53. Smolensky P. Information processing in dynamical systems: Foundations of harmony theory: Colorado Univ at Boulder Dept of Computer Science1986.

54. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations. MIT Press, 1986:318-62.

55. Bourlard H, Kamp Y. Auto-association by multilayer perceptrons and singular value decomposition. Biol Cybern 1988;59:291-4.

56. Li Q, Cai W, Wang X, et al., editors. Medical image classification with convolutional neural network. 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV); 2014 10-12 Dec. 2014.

57. Mandic D, Chambers J. Recurrent neural networks for prediction: learning algorithms, architectures and stability. Wiley, 2001.

58. Creswell A, White T, Dumoulin V, et al. Generative Adversarial Networks: An Overview. IEEE Signal Processing Magazine 2018;35:53-65.

59. Naudé W. Artificial Intelligence against COVID-19: An early review. 2020.

60. Khan A, Sohail A, Zahoora U, et al. A survey of the recent architectures of deep convolutional neural networks. Artificial Intelligence Review 2020;53:5455-516.

61. Ronen J, Hayat S, Akalin A. Evaluation of colorectal cancer subtypes and cell lines using deep learning. Life Sci Alliance 2019;2:e201900517.

62. Chen H, Engkvist O, Wang Y, et al. The rise of deep learning in drug discovery. Drug Discov Today 2018;23:1241-50.

63. Briganti G, Le Moine O. Artificial Intelligence in Medicine: Today and Tomorrow. Front Med (Lausanne) 2020;7:27.

64. Orth M, Averina M, Chatzipanagiotou S, et al. Opinion: redefining the role of the physician in laboratory medicine in the context of emerging technologies, personalised medicine and patient autonomy ('4P medicine'). J Clin Pathol 2019;72:191-7.

65. Union TEPatCotE. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union 2016;Vol. L119:1-88.

66. Hulsen T. Sharing Is Caring-Data Sharing Initiatives in Healthcare. Int J Environ Res Public Health 2020;17:3046.

67. Scheibner J, Raisaro JL, Troncoso-Pastoriza JR, et al. Revolutionizing Medical Data Sharing Using Advanced Privacy-Enhancing Technologies: Technical, Legal, and Ethical Synthesis. J Med Internet Res 2021;23:e25120.

68. Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning. npj Digital Medicine 2020;3:119.

69. Commission E. Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. 2021. Available online: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=75788, accessed 2021/05/26.

70. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3:160018.

71. Hulsen T. The ten commandments of translational research informatics. Data Science 2019;2:341-52.

72. Parikh RB, Teeple S, Navathe AS. Addressing Bias in Artificial Intelligence in Health Care. JAMA 2019;322:2377-8.

73. Hulsen T, Obbink H, Van der Linden W, et al. 958 Integrating large datasets for the Movember Global Action Plan on active surveillance for low risk prostate cancer. European Urology Supplements 2016;15:e958.

74. Omar MI, Roobol MJ, Ribal MJ, et al. Introducing PIONEER: a project to harness big data in prostate cancer research. Nat Rev Urol 2020;17:351-62.

75. Kulkarni S, Seneviratne N, Baig MS, et al. Artificial Intelligence in Medicine: Where Are We Now? Acad Radiol 2020;27:62-70.

76. Siau K, Wang W. Building trust in artificial intelligence, machine learning, and robotics. Cutter Business Technology Journal 2018;31:47-53.

77. Glikson E, Woolley AW. Human Trust in Artificial Intelligence: Review of Empirical Research. Academy of Management Annals 2020;14:627-60.

78. Asan O, Bayrak AE, Choudhury A. Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. J Med Internet Res 2020;22:e15154.

79. Verghese A, Shah NH, Harrington RA. What This Computer Needs Is a Physician: Humanism and Artificial Intelligence. JAMA 2018;319:19-20.