CENTER FOR THE GOVERNANCE OF CHANGE

ARTIFICIAL INTELLIGENCE IN HEALTHCARE: A TECHNOLOGICAL PERSPECTIVE

Marcos Gallego Llorente, IE School of Global and Public Affairs
Adam Berman, University of Cambridge
Mireia Crispin, University of Cambridge

Innovation, Sustainability, and the Future of Healthcare
Chapter 1

Artificial intelligence (AI) is one of the key technological innovations shaping our modern world. The growing **availability of digital health data**, combined with the **newest technical advancements** in computing and data science, form the perfect environment for data-driven, AI-powered medicine to flourish.

Many of the challenges presented by AI arise from its very foundation, from biases to transparency or generalizability. This chapter therefore focuses on presenting the **basic ideas behind AI algorithms**. In particular, it focuses on the different types of learning (supervised vs. unsupervised, shallow vs. deep), and discusses some of the most popular ways in which AI methods have been applied in healthcare. We then discuss the **main technological pitfalls**, and summarize them into **four key questions** on transparency, overfitting, performance and robustness that AI models with translational ambitions should be able to address adequately.

Finally, AI has the potential to go beyond simple applications, and become one of the key tools to connect medical research and clinical practice. We suggest that, if the field progresses steadily despite all the complexities, it could become the backbone of a **knowledge-generating feedback loop** between the two.

Reference to this paper should be made as follows:
Gallego, M., Berman, A., Crispin, M. (2020) "Artificial intelligence in healthcare: A technological perspective", Innovation, Sustainability, and the Future of Healthcare.

1. INTRODUCTION

Recent advances in digital technologies, computing and data analysis methods have triggered the advent of a new data-rich world. In addition, recent years have seen an unprecedented growth in the quantity and complexity of quantitative data. In biomedicine and healthcare this includes, for example, blood biomarkers, DNA and RNA sequencing, digital medical images, electronic health records, or digital recordings from a variety of new medical devices.

One of the critical advancements that have dramatically changed the landscape of data-driven technologies has been the development of a new generation of machine learning algorithms which exhibit cognitive behaviour, broadly called Artificial intelligence (AI). AI has the potential to bring a paradigm shift to healthcare, powered by increasing availability of healthcare data and rapid progress of analytic techniques [1].

There is a growing interest towards the possibility of using AI for patient diagnosis, treatment selection and disease tracking in the near future. New-generation algorithms are becoming increasingly competent at extracting complex patterns from large amounts of data, and using them to make decisions. This, coupled with their ability to improve the quality of their prediction over several iterations, makes AI algorithms an attractive tool for optimizing medical decisions in healthcare settings based on patient data.

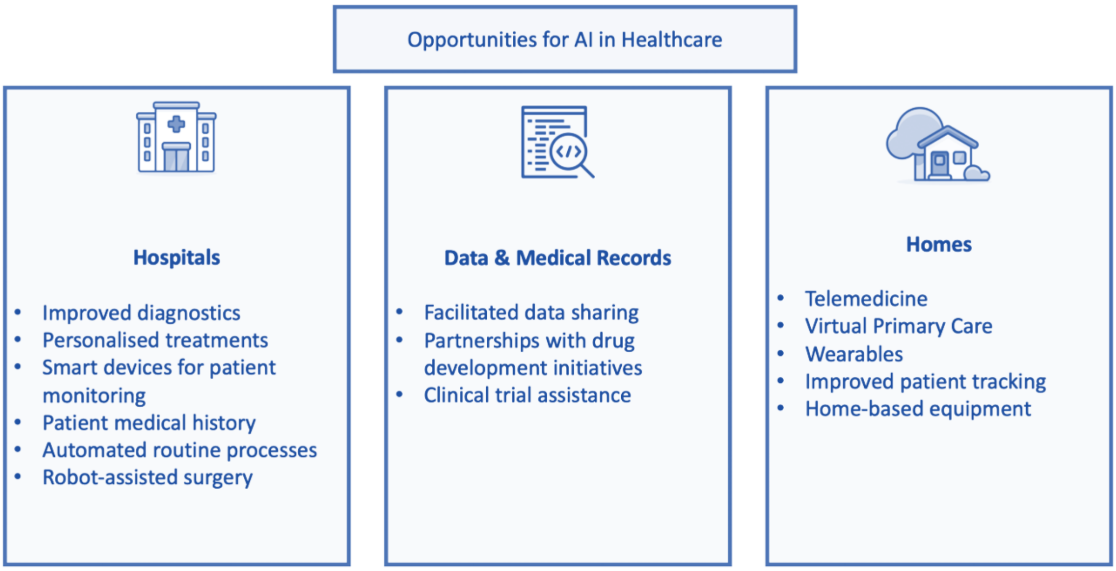

Figure 1. ***Chart on the uses of AI algorithms in different settings.*** *The use of AI has the potential to benefit patient care in hospitals and homes, as well as streamlining the use and sharing of relevant medical data between patients and doctors, between different institutions, stakeholders, and health systems.*

The use of AI in the clinical setting brings forth a set of technical, logistical, regulatory and ethical challenges. Although many of them are general and relevant to all AI applications (for example, data sharing or data privacy), in reality there is a wide variety of scenarios and tools, each of which present their own questions and opportunities [2]. The relative vagueness and suggestive capacity of the term AI sometimes results in unrealistic expectations that are out of reach for current technologies [3]. The aim of this chapter is to lay the foundation for the rest of the report by discussing the main types of AI algorithms, what they can actually accomplish, and how they are being developed within the context of healthcare.

We first discuss the main ideas behind the most widely used algorithms from a computational point of view. Then we discuss the main types of medical scenarios that are being targeted by AI techniques and how they could potentially have an impact on the patient pathway, giving a broad overview of the specific domains in which there has already been progress. Next, we look at the main technological pitfalls, condensing them into a concise set of questions that every AI application ought to be able to answer adequately. Finally, we discuss how AI can also help to streamline medical research from academia into clinical practice.

2. WHAT IS AI?

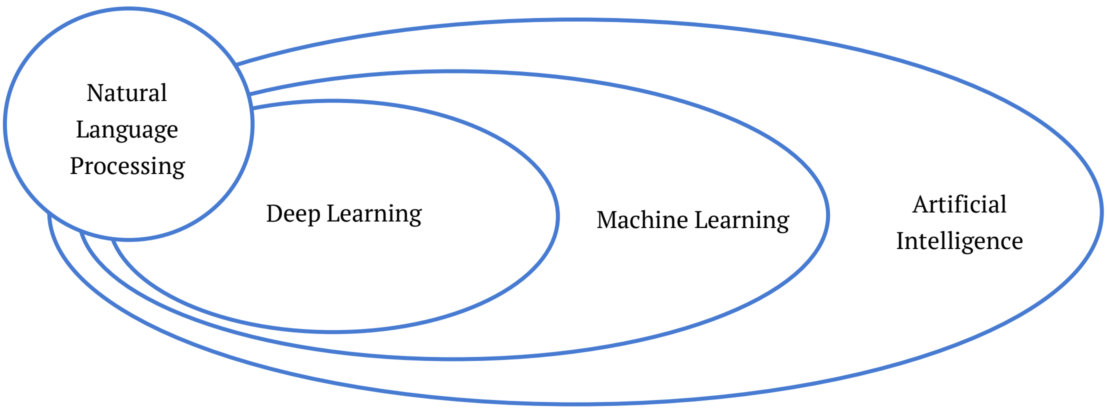

Figure 2. ***Venn diagram on types or sub-types of AI.*** *Machine learning algorithms can be considered as a sub-group within artificial intelligence, while deep learning algorithms are a special class of machine learning algorithms.*

Artificial Intelligence is defined as the collection of techniques that confer an artificial entity with the ability to perceive the environment and take actions that maximise a certain goal [4]. As such, it theoretically encompasses not only the algorithms used for decision-making, but also the techniques used for all perceptual and effectual tasks. Commonly, however, the term is used as a synecdoche to refer to the learning algorithm only. Most of the algorithms used for medical applications are based on **Machine**

Learning (ML) techniques. Machine learning refers to programmes that are able to automatically generate rules and discern patterns based on data and experience with the aim to achieve a desired objective [5].

Ultimately, machine learning models need to be able to generalize and apply the rules they have learnt to new, unseen data. This process is known as "testing" or "validation", and it is crucial to assess whether a model can be used widely or if it has important biases.

Supervised vs. unsupervised learning

Broadly speaking there are two main types of machine learning algorithm which are distinguishable by the way they are trained and the ultimate task they accomplish:

Supervised learning algorithms are trained by providing a set of examples for which the association one wishes the model to discern has been provided as a *ground truth* that the algorithm can learn from. Supervised learning from ground-truth training has proven to be highly successful across multiple domains, from automated identification of image features to optical text recognition [6]. Algorithms can be further classified into whether the task deals with classification into different categories (for example, disease subtypes), or regression of a continuous variable (for example, predicting the volume shrinkage of a tumour as a result of therapy).

Supervised learning is highly dependent on the data available for training. If the dataset is too small, or the number of input features the algorithm can learn from too large, the resulting model may be *overfitted*. An overfitted model is based on spurious patterns that are only relevant for the data it learnt from, therefore becoming unable to generalize. This is a common problem that will be discussed in more detail in the next section as it can have serious repercussions for healthcare applications.

Unsupervised learning algorithms are trained without ground truth training examples. Instead, these algorithms aim to find separations or clusters in the inputted data based on the data's natural patterns across multiple dimensions. Unsupervised clustering is used widely to automatically identify patterns of disease response based on clinically relevant features, for example [7].

A third type of learning method known as **reinforcement learning** has been receiving more attention lately [8]. Reinforcement learning algorithms take sequences of actions as input and learn to follow the sequences that maximize a cumulative reward. They have been proposed to manage situations such as sepsis or epilectic seizures, among others [9], [10].

Traditional vs. deep learning

In addition to the supervised-unsupervised split, another important division within ML is the distinction between shallow (or "traditional") ML and their counterpart, the newer deep learning (DL) algorithms.

Traditional (shallow) ML algorithms are all characterized by having a simpler architecture and fewer parameters than DL models. They have the advantage of being generally quicker to implement and train than their deep counterparts. Despite their speed and relative accessibility, they are not able to learn as complex relationships as deep learning models can, rendering them often insufficient for classifying large, unstructured data. Despite their limitations, traditional ML approaches remain very effective in many domains and have been widely used to great success [11].

Artificial neural networks, which DL algorithms are based on, are in fact a class of traditional ML algorithm that has been used for decades [12]. Artificial neural networks are composed of neurons, computational units that provide an activation signal upon receiving some input. Neurons are connected such that a neuron's output becomes another one's input, thus creating a layered network.

Deep Learning (DL) algorithms are a sub-class of machine learning techniques primarily comprising neural networks with large numbers of hidden neuron layers controlled by complex architectures with a large number of parameters. DL uses multiple layers in order to progressively extract higher level features from the raw input [13].

One of the most widespread types of Deep Learning algorithms is the **Convolutional Neural Network (CNN)**. CNNs are a subtype of Deep Neural Networks that have attracted a high level of interest due to the fact that they are extremely well-suited to analysing image data, including medical images such as MRI scans, x-rays and histopathologically stained tissue images. CNNs have achieved notoriety as some studies have claimed that they were able to achieve superior accuracy compared to humans, for example when analysing dermatology images for some specific diagnostic purposes [14]. These methods still pose unsolved challenges related to overfitting and dataset size restrictions, as deep neural networks often rely heavily on a large and robust training dataset. However, progress is being made to minimize these challenges [15].

In addition to CNNs, a class of DL algorithm that is gaining attention is the **Generative Adversarial Network (GAN)**. These algorithms have gained importance in the image analysis community due to the fact that they learn to generate new pseudo-data, which can be used to augment the data and therefore improve training, or for image-to-image translation, among others [16].

Finally, DL techniques are also becoming more widespread for **Natural Language Processing (NLP)**. These algorithms are used to extract information from large amounts of unstructured data from various textual sources, such as medical records, clinical notes, scientific publications, etc. NLP models can generate structured medical data that can in turn be used by ML algorithms to aid diagnosis and treatment choices.

3. AI IN HEALTHCARE

AI applications are being developed to tackle most stages of the process of going from bench to bedside and beyond. Not all of them will have the same disruptive impact. Some may be used as more efficient versions of previously automated technologies, whereas others may have the potential of completely changing the patient care pathway.

Some of the most widely discussed applications include the following:

Pre-clinical applications. A number of AI systems have been developed to optimize pre-clinical drug discovery processes, such as DeepBind by the University of Toronto, DeepSEA at Princeton, or companies such as BenevolentAI or HealX in the UK. While these technologies have the potential to accelerate drug development dramatically, and they are a key component of the future of AI-powered biomedicine, they are unlikely to affect the patients' direct experience. This type of application will be discussed in more detail in the next section.

Diagnostic and non-patient-facing clinical applications. The types of tasks that are most amenable to AI automation are those based on perceptual recognition. As such, disciplines such as radiology, pathology or some aspects of cardiology, where physicians' tasks are to a large extent based on assessing data visually, have received more attention. These types of advancements may be key to relieve some of the pressure on physicians given workforce shortages, and in some cases to improve the assessments by gaining in robustness and quantitative value. In many cases they replace other computational tools, and may not affect the patients' experience directly.

In **radiology** a large proportion of the work is based on image analysis by deep neural networks, including applications for X-rays, CT, MRI and PET scans, and focus on improving image acquisition or reconstruction, automated detection or interpretation of findings, image segmentation and registration, reporting and correlation of other data sources of interest [17]. Although the field is broad and varied, image modalities with large available datasets tend to be the one with more mediatic impact. As an example, chest X-rays have been the focus of popular studies by Stanford [18] and Google [19], among others [20], [21]. Some of these studies claimed superhuman performance, although these claims have received some criticism [22].

In **pathology** the work is also dominated by image analysis based on a combination of traditional and deep learning approaches to identify and characterize cell and tissue features that can help with automatic feature detection and segmentation, as well as more complicated tasks such as disease grading or subtyping [23]. These applications all rely on a digital pathology platform, which is still not the norm in most hospitals [24].

Other popular studies include skin malignancies classification in **dermatology** [14], [25], and the analysis of electrocardiograms [26], [27] for **cardiology** applications. AI can also be used to provide automated and editable ventricle segmentations based on conventional cardiac MRI images. This method has currently received clearance from the FDA in the United States [28], [29]. In **neurology**, an AI system has been proposed in order to restore the control of movement to a certain extent for quadriplegic patients [30].

Patient management and treatment decision applications. There are some AI applications that may have a more dramatic impact on patients' experience, as they affect their interaction with human doctors, or rely on the patients disclosing personal data.

A well-known example is that of chatbots used for symptom checking and triaging, such as the UK's Babylon Chatbot or Sweden's Aitopya (both covered as case studies in this report). These applications are being scrutinized regarding performance, regulation and patients' engagement [31], [32]. Another extreme example is the development of ML models to predict the onset of depression from the patient's Facebook posts [33].

Many healthcare institutions are transitioning to Electronic Health Records (EHRs): digital versions of a patient's record including medical history, diagnoses, medications, treatment plans, immunization dates, allergies, radiology images and laboratory and test results [34]. The digitalization enables automation and streamlining of the decision-making workflow, as well as a more comprehensive treatment personalization strategy [35].

These are only some examples of the large number of applications that are being developed, especially in the academic setting. As the technology evolves and the field matures, it is likely that new opportunities will arise, at the same time as the attention focuses on the most successful avenues found in this initial exploration phase. Overall, these developments are affecting the entire pipeline of medical practice: from pre-clinical research, early detection and diagnosis, through outcome prediction and prognosis, to the course of treatment and discharge.

4. TECHNOLOGICAL CHALLENGES

Some of the most important aspects to be assessed during the evaluation of an AI tool for practical use will be discussed later on in this same report, within the context of regulation. Nevertheless, on a fundamental technical level, all AI algorithms aiming to have translational impact should be able to address the following questions.

Explainability and transparency: Can the decision process be rationalized? The lack of understanding of how AI algorithms work internally is one of the most common concerns. This is an active field of research, and various solutions have been proposed in different contexts. For example, Class Activation Maps (CAMs) are one of the methods to be able to know what the CNN-based models are observing as they train [36]. CAMs enable researchers to see which areas of the image were most important for the CNN's predictive process, thereby laying bare any biases or lack of generalizing power in the model.

Overfitting and biases: Is the training data representative and relevant? Has the model been extensively validated on large external datasets? As described previously, overfitting is the key problem when working with small datasets. Overfitting occurs when a model corresponds too closely to a particular set of data, and may therefore fail to predict new observations reliably. Despite the growing availability of medical data, it is still an important problem as specific tasks often can only be trained on a small subset with the relevant characteristics. To combat overfitting, data splitting techniques such as cross-validation can be used, as well as a technique known as data augmentation, whereby training data are randomly altered in order to generate a larger training set.

Performance assessment: Is the performance being compared to a robust ground truth? Are the models consistent with the desired specifications? Many AI tools are presented as being equal or better than humans at a given task. However, particularly in healthcare applications, defining human performance is complex as it can be influenced by multiple factors [37]. Using multiple assessments to define the ground truth can help [38], with recent radiological studies including assessments from over 100 radiologists [39]. In addition, the results should be tested against the list of desired specifications, including, for example, fairness [40].

Robustness and adversarial attacks: Are predictions sensitive to subtle variations in the input data? Some ML models, especially deep neural networks, are sensitive to small changes in the input data, potentially leading them to produce an entirely different prediction [41]. This type of vulnerability can be used to carefully engineer changes in the input data to force the system to produce the wrong result [42]. In deep learning, these changes could be as subtle as the application of a noisy-looking filter to an image, producing alterations that are almost imperceptible to the human eye.

5. BEYOND THE CLINIC: EMERGING WAYS OF UNDERSTANDING MEDICAL RESEARCH

AI has the potential to build on and improve the interaction between biomedical research, clinical research and medical practice.

Firstly, the constant evolution of new tests, therapies and biomarkers needs a framework that enables a rapid assessment of their clinical validity and utility. Here, new digital technologies have the opportunity to play a key role: in-silico identification of biomarkers can, for example, save a significant amount of resources in the development of personalized compound diagnostics, as biomarkers are normally identified late during the drug-development process. Additionally, new developments in machine learning approaches make it possible for algorithms to scan through databases and published literature to help scientists identify targets to then test in a laboratory setting, saving time and money while increasing efficacy and probability of success [43].

Through NLP algorithms capable of reading and understanding millions of pages, it is now also possible to parse content at a completely new scale and speed. Predictive analytics applies deep learning algorithms for reading and understanding large amounts of data, including scientific articles, clinical trial data, molecular structures, off-target prediction, toxicity estimations and more, thereby enabling scientists to develop in-silico assays on a large scale while reducing the risk of real-life assays performed in laboratory conditions [44], [45]. Additionally, AI has also been successfully used to combine large pharmaceutical company datasets and discover previously unidentified drug interactions [46].

Secondly, research and clinical practice are and will be increasingly compatible. Initiatives to aggregate and combine research and clinical data aim to improve all the stages of the care pathway: from early detection, to disease management, monitoring and follow up. Most of these areas are likely to be the targets of new developments in machine learning and big data. Machine learning algorithms on the patient side of the healthcare system have the ability to carry out an in-depth, personalized monitoring of the patient and to identify prognostic patterns. Coupled with the pre-clinical advances, this could effectively create an AI-supported knowledge-generating healthcare system [47].

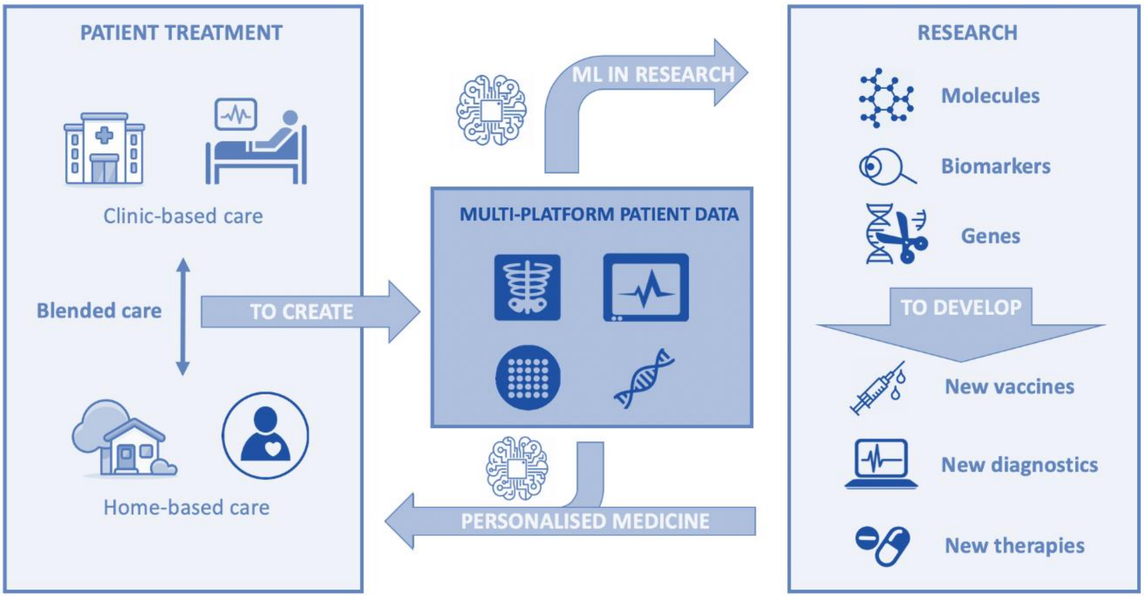

Figure 3. *Chart on the synergies that are currently appearing between research and clinical settings, with the aid of AI-enabled capabilities.* *Multi-platform data obtained by healthy individuals and patients alike can be used by ML-based algorithms to aid drug development, biomarker identification and accompanying diagnostic development. NLP algorithms, in turn, can also be fed information from an extensive body of published literature. The ML-enabled approach to disease treatment would create a body of real-time patient data, which together with NLP algorithms for therapeutic context, would enable an expansive development and use of personalized therapies.*

6. CONCLUSION

Artificial intelligence advancements in healthcare include a variety of applications, which are increasingly relying on heavily parameterized deep learning techniques. The proposed tools range from pre-clinical drug discovery assistants all the way to image-based diagnostics, treatment decision support or follow-up, and therefore may have different degrees of impact on patient care pathways. However, all of them are subject to the same technical challenges, which we have summarized in four key questions that should always be asked: can the decision process be rationalized? Is the training data representative and relevant, and has the model been extensively validated on large external datasets? Is the performance assessment robust, and are specifications satisfied? And finally, are predictions sensitive to subtle variations in the input data? If these and other quality criteria are met, and the field manages to make steady progress, it may be possible to create eventually a knowledge-generating feedback loop to improve patient care through data integration and model development and re-training.

REFERENCES

[1] F. Jiang *et al.*, "Artificial intelligence in healthcare: past, present and future," *Stroke Vasc. Neurol.*, vol. 2, no. 4, pp. 230–243, Jun. 2017.

[2] T. Davenport and R. Kalakota, "The potential for artificial intelligence in healthcare," *Futur. Healthc. J.*, vol. 6, no. 2, pp. 94–98, Jun. 2019.

[3] E. Strickland, "IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care," *IEEE Spectr.*, vol. 56, no. 4, pp. 24–31, Apr. 2019.

[4] D. L. (David L. Poole, A. K. Mackworth, and R. Goebel), *Computational intelligence : a logical approach*. Oxford University Press, 1998.

[5] E. Alpaydin, "Machine learning," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 3, no. 3, pp. 195–203, May 2011.

[6] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," *Brief. Bioinform.*, 2017.

[7] X. Wang, F. Markowetz, F. De Sousa E Melo, J. P. Medema, and L. Vermeulen, "Dissecting cancer heterogeneity - An unsupervised classification approach," *International Journal of Biochemistry and Cell Biology*, vol. 45, no. 11, pp. 2574–2579, Nov. 2013.

[8] O. Gottesman *et al.*, "Guidelines for reinforcement learning in healthcare," *Nat. Med.*, vol. 25, no. 1, pp. 16–18, Jan. 2019.

[9] M. Komorowski, L. A. Celi, O. Badawi, A. C. Gordon, and A. A. Faisal, "The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care," *Nat. Med.*, vol. 24, no. 11, pp. 1716–1720, Oct. 2018.

[10] A. Guez, R. D. Vincent, J. Pineau, and M. Avoli, "Adaptive treatment of epilepsy via batch-mode reinforcement learning," in *Proceedings of the National Conference on Artificial Intelligence*, 2008.

[11] P. Afshar, A. Mohammadi, K. N. Plataniotis, A. Oikonomou, and H. Benali, "From handcrafted to deep-learning-based cancer radiomics: Challenges and opportunities," *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 132–160, Jul. 2019.

[12] H. Wang and B. Raj, "On the Origin of Deep Learning," Feb. 2017.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[14] H. A. Haenssle *et al.*, "Man against Machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Ann. Oncol.*, 2018.

[15] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random Erasing Data Augmentation," Aug. 2017.

[16] M. Jennings, L. G. Marcu, and E. Bezak, “PET-specific parameters and radiotracers in theoretical tumour modelling,” *Comput. Math. Methods Med.*, 2015.

[17] G. Choy *et al.*, “Current Applications and Future Impact of Machine Learning in Radiology,” *Radiology*, vol. 288, no. 2, pp. 318–328, Aug. 2018.

[18] P. Rajpurkar *et al.*, “CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning,” Nov. 2017.

[19] A. Majkowska *et al.*, “Chest Radiograph Interpretation with Deep Learning Models: Assessment with Radiologist-adjudicated Reference Standards and Population-adjusted Evaluation,” *Radiology*, vol. 294, no. 2, pp. 421–431, Feb. 2020.

[20] R. Singh *et al.*, “Deep learning in chest radiography: Detection of findings and presence of change,” *PLOS ONE*, vol. 13, no. 10, p. e0204155, Oct. 2018.

[21] J. G. Nam *et al.*, “Development and Validation of Deep Learning{\textendash}based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs,” *Radiology*, vol. 290, no. 1, pp. 218–228, Jan. 2019.

[22] L. Oakden-Rayner, “CheXNet: an in-depth review,” 2014. [Online]. Available: https://lukeoakdenrayner.wordpress.com/2018/01/24/chexnet-an-in-depth-review/. [Accessed: 15-Mar-2020].

[23] T. J. Fuchs and J. M. Buhmann, “Computational pathology: Challenges and promises for tissue analysis,” *Computerized Medical Imaging and Graphics*, vol. 35, no. 7–8. Pergamon, pp. 515–530, Oct 2011.

[24] J. Griffin and D. Treanor, “Digital pathology in clinical use: Where are we now and what is holding us back?,” *Histopathology*, vol. 70, no. 1, pp. 134–145, Jan 2017.

[25] A. Esteva *et al.*, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, Jan. 2017.

[26] A. Y. Hannun *et al.*, “Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network,” *Nature Medicine*, vol. 25, no. 1, pp. 65–69, Jan 2019.

[27] A. Madani, R. Arnaout, M. Mofrad, and R. Arnaout, “Fast and accurate view classification of echocardiograms using deep learning,” *npj Digit. Med.*, vol. 1, no. 1, pp. 1–8, Dec. 2018.

[28] B. Marr, “First FDA Approval For Clinical Cloud-Based Deep Learning In Healthcare,” 2017. [Online]. Available: https://www.forbes.com/sites/bernardmarr/2017/01/20/first-fda-approval-for-clinical-cloud-based-deep-learning-in-healthcare/#644115bb161c. [Accessed: 15-Mar-2020].

[29] S. E. Dilsizian and E. L. Siegel, “Artificial Intelligence in Medicine and Cardiac Imaging: Harnessing Big Data and Advanced Computing to Provide Personalized Medical Diagnosis and Treatment,” *Curr. Cardiol. Rep.*, vol. 16, no. 1, Dec. 2013.

[30] C. E. Bouton *et al.*, "Restoring cortical control of functional movement in a human with quadriplegia," *Nature*, vol. 533, no. 7602, pp. 247–250, Apr. 2016.

[31] T. Nadarzynski, O. Miles, A. Cowie, and D. Ridge, "Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study," *Digit. Heal.*, vol. 5, p. 205520761987180, Jan. 2019.

[32] G. Iacobucci, "Row over Babylon's chatbot shows lack of regulation," *BMJ*, vol. 368, p. m815, Feb. 2020.

[33] J. C. Eichstaedt *et al.*, "Facebook language predicts depression in medical records," *Proc. Natl. Acad. Sci.*, vol. 115, no. 44, pp. 11203–11208, Oct. 2018.

[34] D. Blumenthal and M. Tavenner, "The 'Meaningful Use' Regulation for Electronic Health Records," *N. Engl. J. Med.*, vol. 363, no. 6, pp. 501–504, Aug. 2010.

[35] C. Xiao, E. Choi, and J. Sun, "Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review," *Journal of the American Medical Informatics Association*, vol. 25, no. 10. Oxford University Press, pp. 1419–1428, 01-Oct-2018.

[36] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, Feb. 2020.

[37] J. W. Gichoya, S. Nuthakki, P. G. Maity, and S. Purkayastha, "Phronesis of AI in radiology: Superhuman meets natural stupidity," Mar. 2018.

[38] V. C. Raykar *et al.*, "Supervised learning from multiple experts: Whom to trust when everyone lies a bit," in *ACM International Conference Proceeding Series*, 2009, vol. 382, pp. 1–8.

[39] A. Rodriguez-Ruiz *et al.*, "Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists.," *J. Natl. Cancer Inst.*, vol. 111, no. 9, pp. 916–922, Sep. 2019.

[40] S. Galhotra, Y. Brun, and A. Meliou, "Fairness Testing: Testing Software for Discrimination," *Proc. ACM SIGSOFT Symp. Found. Softw. Eng.*, vol. Part F130154, pp. 498–510, Sep. 2017.

[41] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards Deep Learning Models Resistant to Adversarial Attacks," *6th Int. Conf. Learn. Represent. ICLR 2018 - Conf. Track Proc.*, Jun. 2017.

[42] S. G. Finlayson, J. D. Bowers, J. Ito, J. L. Zittrain, A. L. Beam, and I. S. Kohane, "Adversarial attacks on medical machine learning," *Science (80)*, vol. 363, no. 6433, pp. 1287–1289, Mar. 2019.

[43] D. A. Chambers, W. G. Feero, and M. J. Khoury, "Convergence of Implementation Science, Precision Medicine, and the Learning Health Care System," *JAMA*, vol. 315, no. 18, p. 1941, May 2016.

[44] E. Smalley, "AI-powered drug discovery captures pharma interest," *Nat. Biotechnol.*, vol. 35, no. 7, pp. 604–605, Jul. 2017.

[45] G. Schneider, "Automating drug discovery," *Nature Reviews Drug Discovery*. 2018.

[46] B. Hie, H. Cho, and B. Berger, "Realizing private and practical pharmacological collaboration," *Science (80)*, vol. 362, no. 6412, pp. 347–350, Oct. 2018.

[47] S. H. Beachy, S. Olson, A. C. Berger, and H. S. Policy, *Genomics-Enabled Learning Health Care Systems : Gathering and Using Genomic Information to Improve Patient Care and Research : Workshop Summary*. 2015.