Chapter 33

Artificial intelligence in gastroenterology

James K. Ruffle[1,2], Adam D. Farmer[1,3] and Qasim Aziz[1]

[1]*Centre for Neuroscience and Trauma, Blizard Institute, Wingate Institute of Neurogastroenterology, Barts and the London School of Medicine & Dentistry, Queen Mary University of London, London, United Kingdom;* [2]*Department of Radiology, University College London NHS Foundation Trust, London, United Kingdom;* [3]*Department of Gastroenterology, University Hospitals of North Midlands NHS Trust, Stoke on Trent, United Kingdom*

Introduction

Artificial intelligence (AI) frequents mainstream media news, from such areas as driver-less cars, recommendation systems of television streaming services, or advanced robotics [1,2]. Moreover, it has a number of potential applications in science and technology. Nevertheless, this enthusiasm and positivity have also been tempered with concerns, including the notion of "robots" replacing workforces, or even the end of mankind, as the late Stephen Hawking hypothesized back in 2014 [3].

Furthermore, as the discussion of AI becomes evermore common, the novelty and excitement it affords generates unsustainable hype, referred to by Gartner as the "*peak of inflated expectation*," a preceding state before this overexaggerated excitement falls to the "trough of disillusionment," when the ideas and hype generated simply do not live up to reality [4]. Needless to say, the prospective utility of AI in healthcare is an enticing and rapidly developing field [5]. Will it revolutionize healthcare, damage it, or perhaps have minimal effect at all?

Artificial intelligence and data-driven decision making

Artificial intelligence

AI*,* in essence, refers to a computer performing a task associated with an intelligent being. Included are "cognitive" functions not dissimilar to the human mind, such as the ability to "learn" [6]. Despite interest in AI seeming relatively recent by virtue of media attention, its concept is in-fact far more developed, such as the works of Alan Turing (1912–54), a pioneer of computer science, or even ancient such as in Aristotle's study of logic (384–322 BC) [6]. One could consider AI as the subject field with many subfields, in the same light that gastroenterology engenders inflammatory bowel diseases, endoscopy, nutrition, *etc*. These areas of gastroenterology all interlink in multiple ways, to which AI subfields are no different (Fig. 33.1).

Machine learning(ML)

ML is defined as an area of computer science, wherein a system can develop an ability to "learn" with data, without any explicit programming. Historically, it was coined by Arthur Samuel in 1959, a pioneer of computer games, while at IBM. Machine learning originates from computational learning theory with pattern recognition, where a computer can develop an algorithm to make a prediction, based upon recognition of patterns or features in the data. The developed predictive *model* can then adapt further to new circumstances, such as improvement in accuracy with more practice or time.

As a simple example in gastroenterology, Kinar et al. trained a machine learning model to detect the presence or absence of colorectal cancer (the target), by complete blood count data (the input predictors), with an overall accuracy of 82% [7]. This particular paradigm is that of a binary classifier, disease positive or disease negative, and used a specific model known as a decision tree ensemble. In training a classification model, there are a number of algorithmic methods: decision trees, an ensemble of trees, support vector machines, K-nearest neighbors, to name but a few. The details on the "types of learning" is beyond the scope of this chapter, but we would recommend Dey's "*Machine learning algorithms: a review*" for further reading [8].

There is often some uncertainty as to the best or "gold standard" methodological approach, to be adopted for a given research question, and thus, some machine learning toolboxes now automatically train a model using a variety

Precision Medicine for Investigators, Practitioners and Providers. https://doi.org/10.1016/B978-0-12-819178-1.00033-2

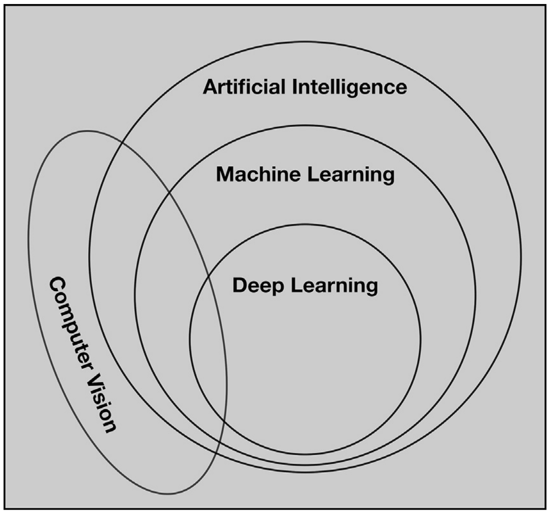

FIGURE 33.1 Nomenclature in artificial intelligence. *Adapted from Ruffle JK, Farmer AD, Aziz Q. Artificial intelligence assisted gastroenterology − promises and pitfalls. Am. J. Gastroenterol 2019;114(3):422−8. https://doi.org/10.1038/s41395-018-0268-4.*

of learning techniques, and depict the receiver operator characteristic or performance for each, enabling the investigator to determine the most effective model to use [9].

Data partitioning: training, validation, testing

Data partitioning is essential when developing a predictive model. The essential elements to partition in the data are referred to as *training, validation,* and *testing*. This follows a simple rationale, namely that if one is to develop a predictive model upon a specific dataset, how could it be subsequently tested on unseen data or samples? For achieving this, it is common to use randomized partitioning of data, to an initial training set to train the model, frequently in the region of 70% of the available samples. The remaining samples would then be allocated for model validation and testing. The function of the validation partition, is to ensure that a model does not "overfit" to a training dataset—a phenomenon which means that the machine will not develop an algorithm to make predictions too close to that of the training data, and then does not generalize well, performing poorly on the never before seen testing samples. The testing dataset is used to test the performance of the model.

Deep learning

Deep learning is an area of AI that uses multiple layers of nonlinear processing to extract features of the data and transform them in order to predict a desirable target [10]. *Feature extraction* is the automated selection of variables, that have some degree of predictive power to predict a target, while *transformation* is the augmentation of data, into a means more effective in building the model. Simply put, physicians intrinsically perform feature extraction on a regular basis, such as in taking a medical history and "extracting" the salient points, in predicting a patient's diagnosis. The difference in deep learning, however, is that this feature extraction is automated computationally.

Using the previously discussed colorectal cancer example [7], by providing complete blood count data, feature extraction can select measures of the full blood count, that yield the most "predictive power" in predicting a diagnosis of colorectal cancer, and furthermore assign weights to each of the variables, dependent on their usefulness. For instance, in this model, hemoglobin, mean corpuscular hemoglobin, mean cell volume, and hematocrit were found to have the highest predictive powers, while other measures (such as monocytes) held a lesser contribution. Deep learning has notably formed a key methodology in the research of medical image analysis, and a well-known framework to develop with here is "TensorFlow" [11].

Neural networks

In recent years, *Neural networks* are also frequently discussed and utilized in machine learning. It is named as of its similarity to the brain's neural network. Neural networks incorporate a network of elements called *neurons*, which receive inputs (such as patient data) and subsequently change their internal state (such as activating), depending on the input, which then produces an output contingent on this (such as predicting a diagnosis). By using the data provided as input vectors, the neural network acts as an interconnected group of nodes, to predict a response. In healthcare research, much of the current AI models seemingly use neural networks [12,13].

Below, we provide a hypothetical example using a neural network, in predicting both a diagnosis from a differential (binary classifier) and also a predicted treatment efficacy (continuous variable) in (Fig. 33.2), using inflammatory bowel disease (IBD) as the example. Similar to the aforementioned classification models, the neural network is a partly all-encompassing terminology, which incorporates a variety of different learning types, examples of which are Scaled-Conjugate Gradient, Levenberg−Marquardt and Bayesian Regularization, to name but a few. Further reading on learning types can be found here [6,8].

Patterns and interactions in data

An additional use that machine learning could provide in developing predictive models is the automated analysis of interaction factors in predicting a target. By and large, a large proportion of epidemiological studies in predicting an outcome, rely on the principal analysis of one variable and relating that to an outcome, for instance, the risk of lung

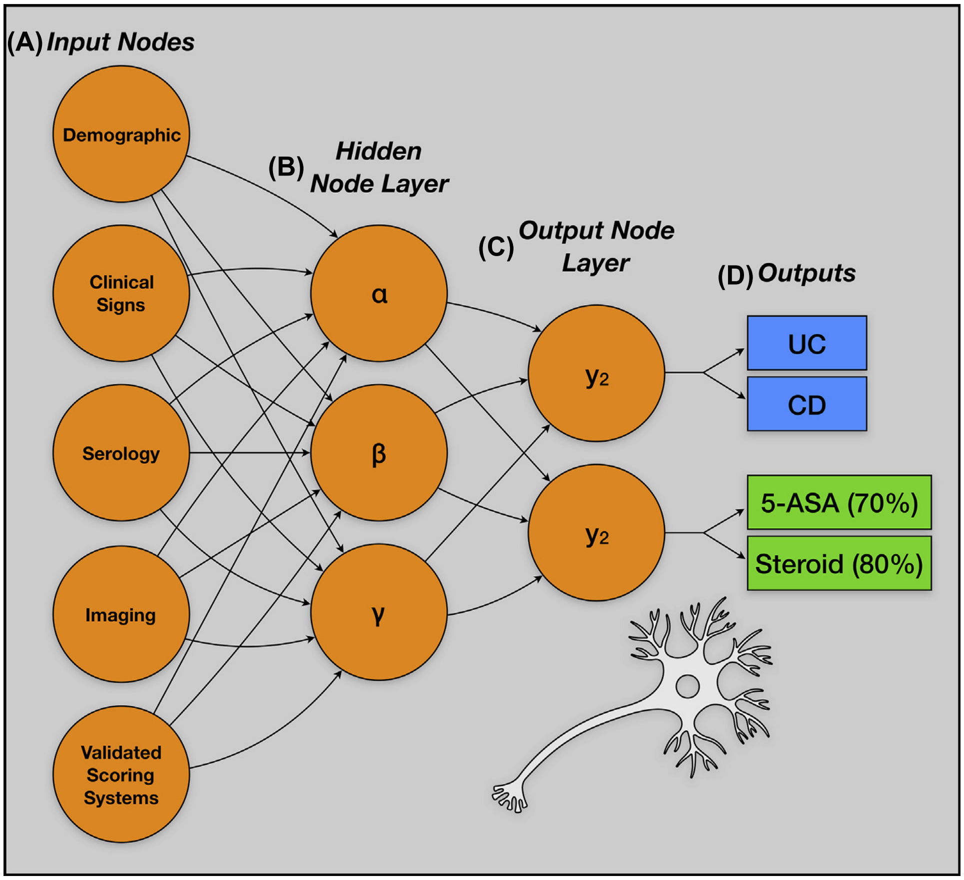

FIGURE 33.2 Example of a neural network to predict the diagnosis of ulcerative colitis (UC) or Crohn's disease (CD), and moreover whether treatment with a 5-aminosalicylic acid (5-ASA) or a corticosteroid, would be more effective. (A) Example input vectors about the patient, to be inputted to the neural network include history, examination findings, an investigation, and so on. (B, C) In neural network design, the utility of these parameters in (A) are combined, weighted and biased, depending on their predictive power at a hidden nodal layer, to provide an output layer. (D) After the internal steps of (B, C), the algorithm determines a response − such as a diagnosis from a differential (output layer y_1) − a binary classifier, or a predicted response to a treatment (output y_2) − a continuous variable. In this hypothetical example, this data could then be taken into consideration between doctor and patient, to determine what may be best for the patient. For instance, the doctor/patient may elect to choose the corticosteroid, as of higher predicted symptom relief, given this particular patient's input data. However, while the steroid in this example displays a 10% greater symptomatic response, the doctor/patient may feel this 10% is "not worth" the adverse effect profile, that accompanies treatment with steroids. Either way, this provides a powerful tool to make a data-driven decision in healthcare, at the individual level. *Figure from Ruffle JK, Farmer AD, Aziz Q. Artificial intelligence assisted gastroenterology − promises and pitfalls. Am. J. Gastroenterol 2019;114(3):422−8.* https://doi.org/10.1038/s41395-018-0268-4.

cancer in smokers. A scientist can then introduce other variables, and interrogate their relationship to the predictor (smoking), or the target (lung cancer), either methodologically, such as with the Bradford Hill causality criteria (a set of viewpoints to determine if an observed epidemiological association is causal) [14,15], or even statistically, such as with logistic regression.

In machine learning, however, this process can be a highly efficient process of recognizing interactions between variables, and how they relate to the target predictor, in an automated way. For example, let us say, a hepatologist wanted to develop a model to predict alcoholic liver disease, by using serological liver function tests results. The presence or absence of alcoholic liver disease would be the target to predict (a binary classifier), and the liver function test values would be the input vectors, on which the model would base its prediction.

In feature extraction, the model would assign particular weights to predict a target. For example, aspartate transaminase (AST) would likely have more predictive power than lactate dehydrogenase (LDH). However, the model would also *likely* determine that, in fact, the interaction of AST with alanine transaminase (ALT) also holds predictive power. Notably, as this is similar to how a physician may calculate the AST/ALT ratio in alcoholic liver disease [16,17]. However, what is also important to appreciate here is our stipulating what the model would *likely* do. There exists an inherent limitation in current AI technology where, despite a model being hugely effective at a prediction, its means of how it achieves this precisely often remains unknown, a "black box" technology of sorts.

Image analysis and computer vision

Computer vision refers to the AI subfield of using computer systems in the processing of visual data, images or videos, for instance how a computer may ascertain information from this (and make a prediction accordingly). This is a frequently discussed area in the media, for example, driverless cars, which has featured prominently on Gartner's recent predictions of technological advances [4,18].

Historically, radiology has been a specialty of humans undertaking pattern-recognition of images to predict a diagnosis, such as spotting polyps on CT colonography [19]. The use of computer vision in radiological investigations is a hot topic, as to how AI could augment and improve the field. Moreover, a key message from radiological academic meetings has been that "*Rads [Radiologists] who use AI will replace rads who don't*" [20]. Perhaps the question could also be asked for certain areas of gastroenterology?

Guideline specific treatment algorithms

AI assisting treatment allocation

A central facet of contemporaneous medical practice (not just in gastroenterology), is the use of treatment guideline pathways or algorithms, typically determined by evidence and expert consensus. The physician identifies the diagnosis and follows the specific investigative and treatment algorithm for that particular disorder, regardless of the individual patient. For instance, in upper gastrointestinal (GI) bleeding, physicians follow a sequence of treatment algorithm of steps X, Y, and Z, which were all determined to be best practice.

In the upper GI bleeding example, one point of outstanding controversy in management is the use of proton pump inhibitors (PPI) prior to endoscopy [21]. There is without question a plethora of research studies investigating upper GI bleeding over recent decades (see Cochrane), and with that considerable patient data from clinical trials. The nature of this data might be demographics, patient comorbidity, GI-bleed score systems, *etc.*, before intervention but also that of patient outcome. This serves as a perfect example of how this data could be used with machine learning, wherein this preinterventional data could be used for predictive vectors, and the predictive outcome being a clinical outcome with or without PPI. Furthermore, this could then be developed further to build a model to ascertain whether a specific patient would benefit from preendoscopic PPI or not.

AI assisting treatment choice

Often enough, there might be a handful of possible treatments, and it might not be a simple "black or white" decision, in choosing what a patient would best benefit from. The quantification of a patient's response to treatment has an important role here, whether that is by the quality of life, symptom relief (e.g., pain, nausea, bowel habit), or other markers of improvement, such as inflammatory serological markers, endoscopic, or radiological appearances. Outside of AI, the use of quality-adjusted life years (QALY) forms a vital area in health outcomes and decisions in treatments, including with an economic role in the UK's National Health Service [22].

Let us consider irritable bowel syndrome (IBS), an enigmatic disorder which has frequently been reclassified, and with that its recommended treatment algorithms [23−25]. The choice of X or Y might not be clear cut, it might frankly feel more like a "gut feeling," or perhaps the *a priori* knowledge that the physician might have about the specific patient, and the condition might lead to choosing one treatment over the other.

There is without question vast amounts of patient data from clinical trials, of different treatment options in IBS, with associated demographics, comorbidities, investigation results, *etc*., but also response to treatments (be that an objective measure or a patient-quantified response), such as "*on a scale of 1-10, how much improvement did you feel?*." Machine learning models can be trained to fit continuous variable data (such as symptomatic response following treatment). An example here is the neural network "fitting" algorithms. Could machine learning algorithms be used to predict a percentage improvement to multiple therapies? This could enable the gastroenterologist to make a data-driven decision at the individual patient level, i.e., *personalized medicine* (Fig. 33.3).

Computer vision in endoscopy

The majority of current AI research in gastroenterology focuses on computer vision in endoscopy [26,27], though capsule endoscopy has also recently been studied, such as automatic detection of GI angioectasia [28]. For example, in polyp surveillance in endoscopy, a real-time video is depicted on a screen for the gastroenterologist to interpret, while simultaneously facing the difficulty of navigating the mobilizing colon, through a two-dimensional representation of what is actually a three-dimensional structure.

In addition to these difficulties, there are the distractions from the procedure, such as if the patient is in discomfort or pain, or if the bowel preparation is suboptimal . This could result in a polyp being missed. Perhaps AI computer vision could aid the gastroenterologist here, identifying suspicious lesions from the video feed, in *real-time*. Notably, this

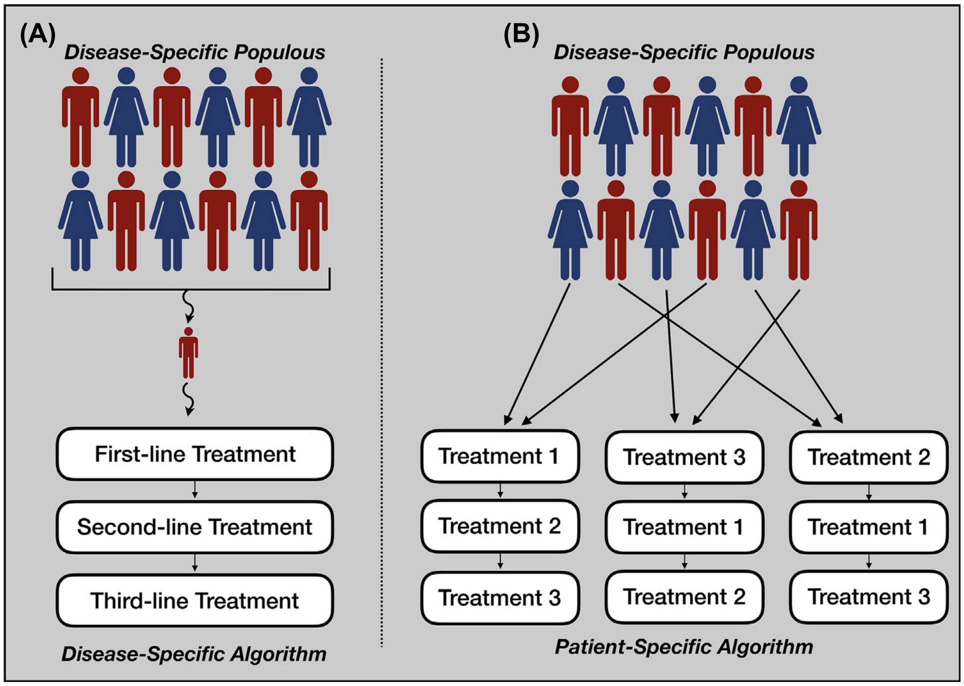

FIGURE 33.3 From disease-specific algorithms to patient-specific with machine learning. (A) Conventional healthcare provision of a given disease, focuses on the definition of a treatment pathway determined by best clinical practice, cost efficiency, scientific literature, *etc*. (B) With machine learning, perhaps models could be developed to stratify patients to treatments, most likely to benefit them specifically, by personalizing medicine. *Adapted from Ruffle JK, Farmer AD, Aziz Q. Artificial intelligence assisted gastroenterology — promises and pitfalls. Am. J. Gastroenterol 2019;114(3):422—8. https://doi.org/10.1038/s41395-018-0268-4.*

would not be dissimilar to if there were a second gastroenterologist watching the endoscopy video-feed, highlighting areas they thought were of concern. We provide a hypothetical example of this below (Fig. 33.4).

There are a number of promising recent studies, developing machine learning models for polyp detection [29—32], or even a step further, the differentiation between adenomatous or hyperplastic [33]. Referring to two studies in particular, Hirasawa et al., developed a neural network to predict gastric cancer, utilizing a *training* set of more than 13,000 endoscopic images, and reported an overall sensitivity of 92%, when *testing* it on a further 2200 images [29]. Meanwhile, Byrne et al. built a neural network using a *training* set of 223 videos (split into >60,000 image frames), and, when tested on a further 125 endoscopy videos, report a 94% accuracy rate, in differentiating diminutive adenomas from hyperplastic polyps [33]. A key learning point to note is the large sample size required to produce robust algorithms.

Aiding in differential diagnoses

Thus far, we have described AI approaches that form a category referred to as *supervised learning*, a terminology which is defined by the human investigator supplying the definitively correct answer for a machine to build a model to predict, such as a diagnosis of bowel cancer, or the known outcome to a particular treatment. This correct answer is often referred to as the *ground truth*. In contrast to supervised learning is *unsupervised learning*, which is when the ground truth is not known (or not supplied to the model).

This area involves providing a dataset with data vectors (such as patient data, symptoms, blood results, imaging), and employing the machine to identify patterns, classifiers or clusters based upon this, without specifying what groups are expected. For instance, this could be useful in gastroenterology, wherein there is an arguably vague and homogenous disease label, for what is perhaps much more heterogeneous, such as functional constipation and IBS with constipation. Perhaps an unsupervised algorithm could identify patterns in patients with constipation, to provide more useful clusters. Functional dyspepsia would, perhaps, also be another poignant example here.

Perspectives for disease prevention and temporality

Machine learning can also predict the temporal evolution of data and trends. Key examples of temporal AI predictions in the commercial sector is financial forecasting, or developing more accurate weather prediction. As a healthcare example, perhaps machine learning could be used to more accurately predict temporal changes, treatment outcomes,

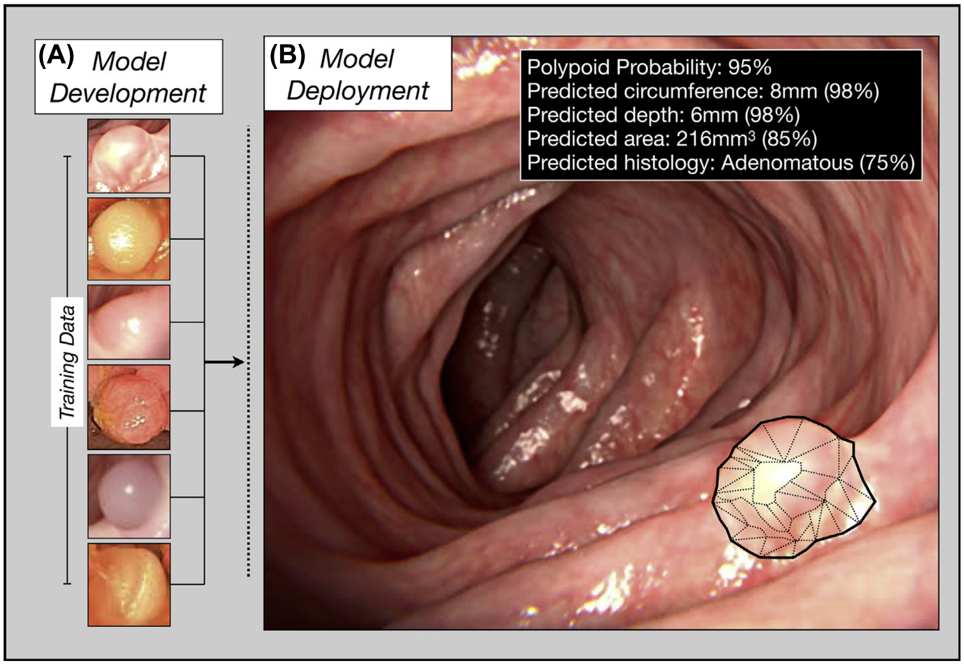

FIGURE 33.4 Computer-vision aided endoscopy. (A) The model is first trained by providing the machine with a large number of training data, images of polyps in this case. (B) A hypothetical example of how the deployment of such a system might take place, whereby in *real-time*, areas of an image that closely relate to the ground truth training data are highlighted to the investigator, and its 3-dimensional geometry recognized, with the probability of it being a polyp illustrated, as well as other metrics. Notably, this is similar to how algorithms of human-testing on the Internet functions: "select all the images with street signs." *Figure from Ruffle JK, Farmer AD, Aziz Q. Artificial intelligence assisted gastroenterology — promises and pitfalls. Am. J. Gastroenterol 2019;114(3):422—8.* https://doi.org/10.1038/s41395-018-0268-4.

or even survival at an individual patient level, as opposed to say the conventional use of a Kaplan-Meir curve, to apply to all patients. For instance, in inflammatory bowel disease (IBD), the physician has to make a decision whether or not to escalate an individual patient to biological therapy. Perhaps temporal AI-analytics could predict the time point in the patient's disease, at which they are likely to necessitate it, benefit from it most, weighted against side effects that accompany it (Fig. 33.5)?

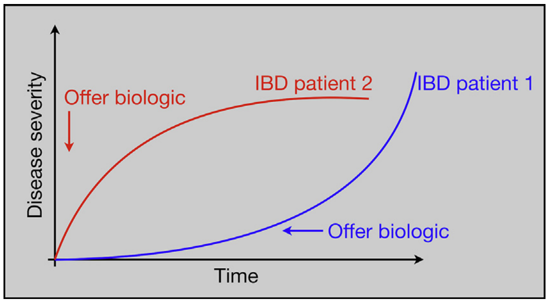

FIGURE 33.5 Predicting disease course temporally to ascertain the effective time-point for treatment intervention. A hypothetical example of two patients with IBD. It is possible that with preexisting patient data, a temporal predictive model could be devised from their disease severity and trajectory, and with that the suggested intervention point with a new therapy. Consider in this example how the gradient of worsening disease severity, differs in both patients significantly, wherein an accurately trained model could purport the point at which an intervention (here, a biological therapy), be introduced to attenuate this gradient.

Research opportunities, limitations, and future directions

Research opportunity

Much of contemporaneous research requires a prospective, expensive/large grant-funded collection of data, over several years, to ascertain a research question as to whether patients benefit most from treatment X or Y. However, data-science and AI methodology offer the utility of using previously acquired large datasets retrospectively to build a model. In fact, they typically rely on this, as the ground truth is therefore established. A salient point in machine learning research is the necessity for large sample sizes, to ensure reproducibility and prevent an overfit; consider the large sample sizes (in the thousands), of the aforementioned computer-vision endoscopy papers as an example.

However, much of this data is archived by research groups, governments, pharmaceutical companies, data repositories, or even made freely available with author's manuscripts. Wherein a dataset was used to determine one clinical question, the dataset can be used for machine learning, to teach a model in a cost-effective, data-effective manner.

Thousands of algorithms to predict thousands of diseases, or one algorithm to predict a thousand diseases?

An "AI-marketplace" has been suggested to become commonplace, by multiple commercial/industrial AI-leaders. Indeed, one for radiology already exists [34]. While this might seem enticing, there lies an inherent limitation to this, which is the differential diagnosis.

If a gastroenterologist reviews a patient, he may conclude with a diagnostic "impression," X, and a series of differentials, Y and Z. In this case, the gastroenterologist needs a test to confirm the impression, rule out differentials, or perhaps even reveal a different diagnosis entirely. These are measures to determine the *ground truth*. Meanwhile, what the gastroenterologist almost certainly does *not* say is that "it's either disorder X or they are completely healthy."

Yet, to date, this is how the vast majority of machine learning models have been developed (and continue to be), where the machine is provided with an array of diseased and healthy controls, and the model needs to identify which is which. Considering the endoscopy-computer vision examples − near all models published thus far function by identifying a singular pathology amongst healthy tissue (e.g., a polyp), rather than identifying *any* pathology, let alone differentiate into distinct pathological diagnoses.

With this in mind, it could be questionable how clinically useful many of these machine learning models actually would be, if deployed in healthcare at present. Rather, we would recommend that models need to be developed that aim to tease out a difficult differential diagnosis in a more effective manner than a physician could.

Algorithm ownership and responsibility

Will hospitals, governments, insurance companies, or even individual physicians, purchase or license the use of algorithms? Legally, who would take responsibility for any erroneous predictions, missed or incorrect diagnoses, for instance, the algorithm purchaser or the designer?

Artificial intelligence does not equate to artificial humanity

The view of "robots replacing the workforce" is not uncommon, forming many a media headline, not least in Hollywood dystopian-themed films. However, while this *may* be a genuine concern in some occupations, we argue the physician/gastroenterologist is not one of them. A principal reason for which is that artificial intelligence is not equivalent to artificial humanity. A machine could be trained to spot a suspicious lesion among healthy tissue on endoscopic images with relative "ease;" however, there are certain areas the machine could not perform or outperform the physician, not least in communication, emotional awareness or empathy.

An AI algorithm may decide that, for a given patient with bowel cancer, he or she would benefit from treatment X over Y, based upon the probability of therapeutic success or survival rate. However, the algorithm would lack an ability to be dynamic, and accommodate the ideas and concerns of the patient (or physician for that matter). Perhaps the agenda of the patient's treatment preference would be vastly different, determined by side effect burden rather than the predicted treatment response on which the model was trained. To that end, it is important to appreciate that while in many ways AI could revolutionize healthcare, the means by which it did so would need to be through AI-assisted practice, not AI-driven.

Stewardship with artificial intelligence in healthcare

As described, supervised or unsupervised learning represent approaches in AI, depending on whether the ground truth is known, or indeed provided. Further to this is the concept of feature extraction and transformation, wherein if an array of different data/variables is provided for a patient, then the model will train, automatically extract and weight the most important or predictive features. While this approach is hugely enticing, perhaps even sounding somewhat "hands-off", this is not the case, and in-fact, the machine can be misled and is vulnerable to garbage or misleading data.

One could train a model to predict a diagnosis of IBS in a cohort of individuals, compared to healthy controls, using demographics, symptom scoring, or psychophysiological data. If for instance, all the IBS patients used in training the model by happenstance were female, the model would *likely* learn this feature, and erroneously only then label female patients as having IBS, namely that a patient being male would be negatively weighted in the model.

References

[1] Baruch J. Steer driverless cars towards full automation. Nature 2016;536(7615):127.

[2] Research N. Netflix research articles. 2018. https://research.netflix.com/articles.

[3] Cellan-Jones R, Hawking S. Warns artificial intelligence could end mankind. BBC News; 2014.

[4] Panetta K. 5 trends emerge in the gartner hype cycle for emerging technologies. 2018. https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018/.

[5] Ruffle JK, Farmer AD, Aziz Q. Artificial intelligence assisted gastroenterology − promises and pitfalls. Am. J. Gastroenterol 2019 Mar;114(3):422−8. https://doi.org/10.1038/s41395-018-0268-4. ePub ahead of print.

[6] Russell SJ, Norvig P. Artificial inteliigence, a modern approach. Pearson Education; 2009.

[7] Kinar Y, Kalkstein N, Akiva P, Levin B, Half EE, Goldshtein I, et al. Development and validation of a predictive model for detection of colorectal cancer in primary care by analysis of complete blood counts: a binational retrospective study. J. Am. Med. Inform. Assoc 2016;23(5):879−90.

[8] Dey A. Machine learning algorithms: a review. Int. J. Comput. Sci. Inf. Technol 2016;7(3):1174−9.

[9] MathWorks. MathWorks. Classification learner. 2018. https://uk.mathworks.com/help/stats/classificationlearner-app.html.

[10] Schmidhuber J. Deep learning in neural networks: an overview. Neural Network 2015;61:85−117.

[11] Tensorflow. An open source machine learning framework for everyone. 2018. https://www.tensorflow.org/.

[12] Nigam VP, Graupe D. A neural-network-based detection of epilepsy. Neurol. Res. 2004;26(1):55−60.

[13] Waxman JA, Graupe D, Carley DW. Automated prediction of apnea and hypopnea, using a LAMSTAR artificial neural network. Am. J. Respir. Crit. Care Med 2010;181(7):727−33.

[14] Hill AB. The environment and disease: association or causation? Proc. Roy. Soc. Med 1965;58:295−300.

[15] Fedak KM, Bernal A, Capshaw ZA, Gross S. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. Emerg. Themes Epidemiol 2015;12:14.

[16] Carey WD. How should a patient with an isolated GGT elevation be evaluated? Clevel. Clin. J. Med 2000;67(5):315−6.

[17] Limdi JK, Hyde GM. Evaluation of abnormal liver function tests. Postgrad. Med. J 2003;79(932):307.

[18] Panetta K. Top trends in the gartner hype cycle for emerging technologies. Gartner; 2017. p. 2017.

[19] Huang A, Li J, Summers RM, Petrick N, Hara AK. Improving polyp detection algorithms for CT colonography: pareto front approach. Pattern Recogn. Lett. 2010;31(11):1461−9.

[20] RSNA AIMI. 2017: Rads who use AI will replace rads who don't Stanford University: artificial intelligence in medicine & imaging. 2017. https://aimi.stanford.edu/about/news/rsna-2017-rads-who-use-ai-will-replace-rads-who-don-t.

[21] Sreedharan A, Martin J, Leontiadis GI, Dorward S, Howden CW, Forman D, et al. Proton pump inhibitor treatment initiated prior to endoscopic diagnosis in upper gastrointestinal bleeding. Cochrane Database Syst. Rev 2010;7.

[22] Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. Br. Med. Bull 2010;96:5−21.

[23] Farmer AD, Ruffle JK. Irritable bowel syndrome. Hamdan Med. J 2015;8(3).

[24] Farmer A, Ruffle J. Irritable bowel syndrome: enigmatic for doctors, problematic for patients. Trends Urol. Men's Health 2017;8(2):13−6.

[25] Drossman DA, Hasler WL. Rome IV-functional GI disorders: disorders of gut-brain interaction. Gastroenterology 2016;150(6):1257−61.

[26] Misawa M, Kudo SE, Mori Y, Cho T, Kataoka S, Yamauchi A, et al. Artificial intelligence-assisted polyp detection for colonoscopy: initial experience. Gastroenterology 2018;154(8):2027−2029 e3.

[27] Alagappan M, Brown JRG, Mori Y, Berzin TM. Artificial intelligence in gastrointestinal endoscopy: the future is almost here. World J. Gastrointest. Endosc 2018;10(10):239−49.

[28] Leenhardt R, Vasseur P, Li C, Saurin JC, Rahmi G, Cholet F, et al. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. Gastrointest. Endosc 2019;89(1):189−94.

[29] Hirasawa T, Aoyama K, Tanimoto T, Ishihara S, Shichijo S, Ozawa T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. Gastric Cancer 2018;21(4):653−60.

[30] East JE, Rees CJ. Making optical biopsy a clinical reality in colonoscopy. Lancet Gastroenterol Hepatol 2018;3(1):10−2.

[31] Yuan Y, Meng MQ. Deep learning for polyp recognition in wireless capsule endoscopy images. Med. Phys 2017;44(4):1379−89.

[32] Zhou T, Han G, Li BN, Lin Z, Ciaccio EJ, Green PH, et al. Quantitative analysis of patients with celiac disease by video capsule endoscopy: a deep learning method. Comput. Biol. Med 2017;85:1−6.

[33] Byrne MF, Chapados N, Soudan F, Oertel C, Linares Perez M, Kelly R, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. Gut 2017.

[34] Holzberger K. AI marketplace is open for business! developers create deep learning algorithms for radiology. Nuance. 2018. https://whatsnext.nuance.com/healthcare/ai-marketplace-is-open-for-business-developers-create-deep-learning-algorithms-for-radiology/.