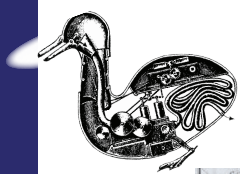


ARTIFICIAL INTELLIGENCE & DIGESTIVE DISEASES

L'Institut des Maladies de l'Appareil Digestif

FUJIFILM
Value from Innovation

ARTIFICIAL INTELLIGENCE & DIGESTIVE DISEASES

Coordinated by Emmanuel Coron & Geoffroy Vanbiervliet

IMAD

With the support of the SANTEDIGE FOUNDATION & of FUJIFILM Europe

Under the aegis of the French Society for digestive endoscopy

IMAD

SUMMARY

IMAD

The IMAD in Nantes and the French Society for Endoscopy have come together for an original initiative: the publication of a book devoted to artificial intelligence in digestive diseases. On the eve of the introduction of these technical innovations in medical practice, the book responds to the urgent need to summarise the current knowledge. Let's make no mistake about it: the rise of artificial intelligence marks a crucial moment in the history of our disciplines. It comes after 15 years of effort to invent new strategies in order to achieve precise, decision-making and proactive endoscopy. At the same time, it opens a new chapter and deeply challenges the role of physicians.

The medical community should not only be a spectator of this technical revolution. Our institutions are present at the interface between techno-industrial actors, patients, political and administrative authorities. They will play a key role, not only in the evaluation of the tools, but also in the definition of transversal solutions. Because they are the only ones able to understand the patient in his singularity, our institutions will have a role to play in the revolution that is coming. They will play it.

EMMANUEL CORON
Director of the Institute of Digestive Diseases (IMAD),
Nantes University Hospital

IMAD

AI and health: history and models

This pictorial representation based on the author's assessment has but an indicative value

1956	Conference of Dartmouth
1957	Creation of the Perceptron
1958	Invention of the Lisp language
1972	Start of the MYCIN program
	First AI winter
1979	Beginning of the Neocognitron
1997	Deep Blue defeats Kasparov
2011	Watson at Jeopardy
2011	Watson in the Emergency Department (UK)
2016	AlphaGo champion
2019	Turing price awarded to Y. Lecun

by Romain Galmiche
Independant publisher

To define the applications of artificial intelligence in health, we firstly need to define what artificial intelligence (AI) is. A first definition is that of a set of techniques allowing the aggregation and processing of heterogeneous and/or complex data. Health systems are in fact used for 1) the automatic interpretation of information, 2) the fusion of data, 3) the construction of representations from information. This first delineation of AI offers the advantage of being operational, but it is nevertheless frustrating. Its main merit is the ability to group a number of devices for image interpretation, voice recognition, statistical analyses, and even tools that help with decision making. It however omits the diversity of the processes. The «black box» effect of AI is well known by practitioners. It may be a trap as well as a facility, considering that it leads to absolute confidence in the machine, without considering its limitations.

Beyond the diversity of techniques, a second definition of AI is as a disciplinary field, at the intersection of formal logic, cognitive and computer science. It aims to reproduce or simulate the elemental operations that characterise human intelligence. This definition is certainly more satisfactory than the first one since it better accounts for the originality of the methods used and the starting hypotheses that are made. In particular, it takes into account the complementarity of the so-called cognitive and connectionist approaches, referring to the two major conceptions of the human mind, i.e. either that of a discrete logical/modular analysis system vs that of an auto-adjusted system using a strategy of weighting [1].

This definition, despite being satisfactory and rigorous, does not reflect yet another « technoscientific » dimension, i.e. the expectations and projections raised in the society. The concept of high intelligence, as opposed to the current low AI, has been raised by the philosopher Searle. He concluded that it was impossible for AI to access the true semantic understanding by simple manipulation of symbols. Since then, the term of high AI was reinvested in a positive sense by essayists that imagined the advent of AI able to supersede human capacity. This fantasy dimension, accessory as it might seem on first instance, cannot be ignored when it comes to health. The topic of transhumanism is another illustration and a good example of this dimension [2].

A brief historical overview of AI is warranted in order to present its various facets, its uses, the fundamental and experimental models available, and the expectations it raised.

AI or the history of an emulation between the man and the machine

"That good old artificial intelligence"

Artificial intelligence as a discipline has an official date of birth: the conference given by John MacCarthy at Dartmouth in 1956, a conference that Marvin Minsky and Herbert Simon, among others, attended [3-5]. An ambitious and coherent program was discussed there: assigning machines with a number of cognitive tasks that are normally attributed to humans. This precise date comes after an already old historical past. One could cite for example Alan Turing and his machine, a «proto-computer» that was given the task of denying or confirming a mathematical theorem. Another proposal by Turing: reproducing mental processes of animals or young children, that are easier to isolate and understand than those of adults, has a promising future ...

From its inception, AI was quickly in competition with the concept of cybernetics: this field of science, founded by Norbert Wiener, is defined as the study of the regulation of complex systems and is at the origin of robotics. Even today, the contours of the two disciplines, AI and robotics, can seem blurry. An autonomous car includes an AI device. Likewise, the contemporary rise of the «Internet of things» allows the AI-dependent interoperability of robots, particularly in the domain of health: pacemakers, surgical robots and others, being able to send and integrate complex data in real time. On a more theoretical ground, medical cybernetics, focused on the idea of internal regulation, has contributed to the emergence of «mathematical physiology», heralding the concept of biomarkers, the latter being finely studied by AI applied to health [6].

Research on AI firstly aims to reproduce certain mental traits, such as those that allow humans to think and act in a limited amount of time, based on limited resources and imperfect information [3,4]. The first AI thus responds to a heuristic problem: optimising the use of algorithms. One could quote the Alpha-beta pruning algorithm of MacCarthy, a method that is applicable for the exploration of a decision tree by eliminating suboptimal branches, and it has long been used in chess programs. Similarly, the A* algorithm, applied to determine the shortest path is widely used in practical applications, such as GPS and for medical imaging with MRI. Quoting Marvin Minsky, the idea that was central to the period pre-1962 was to minimise the magnitude of trials and errors. The aim of reducing computational operations is also a formal requirement. One of the first symbolic languages, Lisp, quickly became the language of choice of AI. It is based on lambda calculus, i.e. a mathematical theory dedicated to solving problems with limited resources and to categorise problems according to the level of difficulty. In fact, one of the most delicate tasks is to model non-monotonic logic, that is to say capable of questioning one of its entry hypotheses...

This formal, logic-driven approach is complemented by the use of statistics. Bayesian networks allow for the representation of a set of variables in a graph [7]. They form the basis of a system that can aid in the decision making, since they allow an agent to move along the logic knots where the plausibility of a hypothesis is calculated based on probability. This system, based on induction, allows the calculation of inferences, i.e. to consider propositions as true based on prior propositions. Bayesian networks are a powerful help for diagnosis in medicine, but are also used in imaging, since they permit modelling of tumor volume after CT or PET scan based on the algorithm FLAB (Fuzzy Locally Adaptive Bayesian Segmentation). A Bayesian probabilistic model permits the delineation of homogeneous regions of the tumor and fuzzy transition areas [8,9].

Neural networks

The logical study and the efforts that are summarised led some authors to talk about the GOF AI « Good Old Fashioned Artificial Intelligence ». At the turn of the 60s, a new field of study, aiming to copy neuronal circuits, emerged. The artificial neuron is inspired by the work of the neurologists Warren MacCulloch and Walter Pitts. It consists of an input and an output: if the input signal reaches a threshold value, an output signal is produced and modulated depending on a pre-specified synaptic weight. The networking of these artificial neurons constitutes one of the main branches of machine learning. According to what is known as the Hebb rule, a neural network learns by strengthening stimulated synaptic connections, according to the principle that: «cells that fire together, wire together». The first network of neurons was conceived by the psychologist Rosenblatt in 1957 and it was called the Perceptron. It consists of a single layer of neurons, and in its initial design, weighed more than a ton. It was funded by the United States of America Department of Defense Advanced Research Projects Agency (ARPA) and operated as a binary classifier (*figure 1*). It was used to perform supervised learning, since it required an initial training in order to adjust the weights until the data used for learning fit in the right class. The strategy in general is promising, especially for image recognition, since it can be applied in a widespread manner, provided that the cases analysed are not too different from the data used for learning.

The application of AI that led to the most optimistic projections in the 1960s: speech recognition, instant language translation, chess invincibility were already mentioned. However, these prospects faced a «winter of artificial intelligence» at the turning of the 1970s, as revealed by a sharp fall in the number of grants. The unrealistic nature of many projects, as well as a number of limitations explain this. The rise of microcomputers and compiled languages have put the initial successes of AI in perspective, considering also the cost and the heavy hardware requirements of AI (the Lisp machine). Moreover, the book by Minsky and Papert «Perceptron: an introduction to computational geometry» revealed some of the inherent limitations of the Perceptron. A neural network consisting of a single layer of neurons cannot handle non-linear functions, i.e. those that fall into Boolean logic (OR) (*figure 1C*) [3,4

This major limitation is however overcome by the rise of multi-layer neural networks. Indeed, these are not limited by non-linear functions and they offer new prospects in terms of form recognition. Fukushima's Neocognitron was build in the 80s and took its inspiration from neuroscience studies of the cat retina, by combining a layer of simple neurons with several layers of complex neurons. The latter permitting the maintenance of the representation of an object, even if it moves in the visual field. In the Neocognitron, the

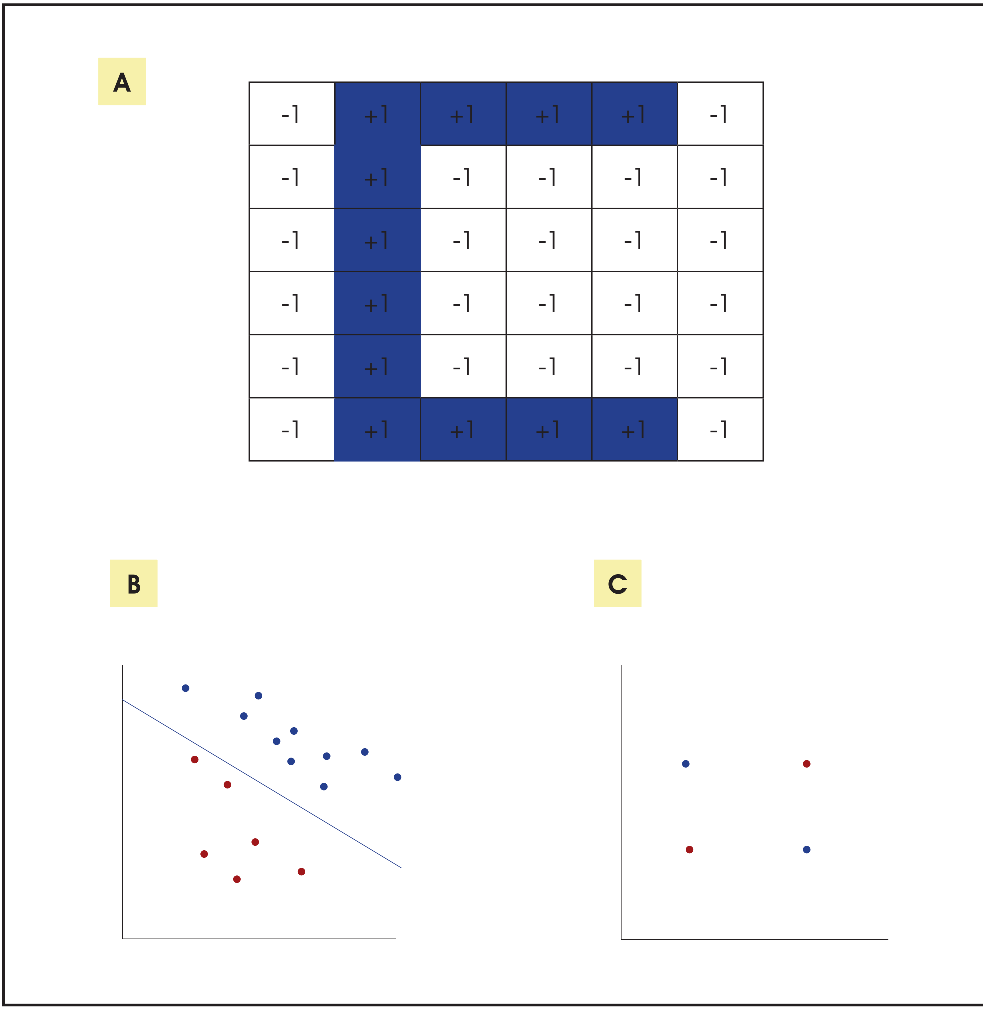


Figure 1. Operation and limitations of the perceptron. A) Pattern intended to determine the synaptic weight in a neural network in a two-dimensional space. Here, the visual recognition pattern of a «C»; B) Binary classification according to a linear function; C) Non-linear classification using exclusive Boolean logic.

first layer of neurons is trained as for a Perceptron, but the underlying layers are trained under blind conditions: in this case, learning is said to be non-supervised. The neural multi-layer networks that were developed in the following decades used procedures with more or less autonomous learning, called reinforcement learning. A pattern is presented to the machine and a reward is programmed in case of positive recognition. At this point however, one large problem that remains is to extract the key characteristics: data extraction and labelling remain an essential limitation of machine learning.

The problem of multi-layer neural network learning can be formulated as the need to decrease the cost function, i.e. the difference between the observed and the desired behavior of the model. While the possibility to use backpropagation of the cost function, from the output to the input, is not specific to deep learning, it is an essential characteristic. Thus, the system can discover the interdependencies between the input variables without being trained for a particular task. This method of learning of multilayer networks has yielded spectacular results since the years 2000, especially in the field of image recognition. It has greatly benefited from the increasing availability of image databases (including images.net, created by American academics) and the development of computer graphic cards with increasing calculation power. The so-called convolutional networks allow the recognition of a pattern whatever its position in the structure of the network.

Deep neural networks, after some initial skepticism, have encountered popularity, owing to some spectacular successes (AlphaGo, 2016). They cannot however be presented as a panacea. A large number of systems are composite in nature and use complex formal AI ontologies (*figure 2*) together with Bayesian models and deep learning. Moreover, an important limitation of deep learning is a lack of explanability: the observer has no access to the learning algorithm. Finally, deep neural networks are a powerful tool for detecting correlations, but they cannot translate them into causality. ”*The current AI has no common sense, and common sense is vital. It conditions our connection to the world. It fills in the blanks and fills in the implicit*” (Yann Le Cun) [3].

Artificial Intelligence and health, a long-established relationship

The introduction of AI techniques in health cannot be simply summarised with a linear representation: devices reaching maturity -> application in health. From the beginning, AI research was inclined toward medicine, perhaps because it corresponds to the notion of «bounded rationality» (Herbert Simon), a domain where an agent makes a decision based on limited information. A conceptual shift can then occur from rationality that is perceived as being ideal to procedural rationality, in which what

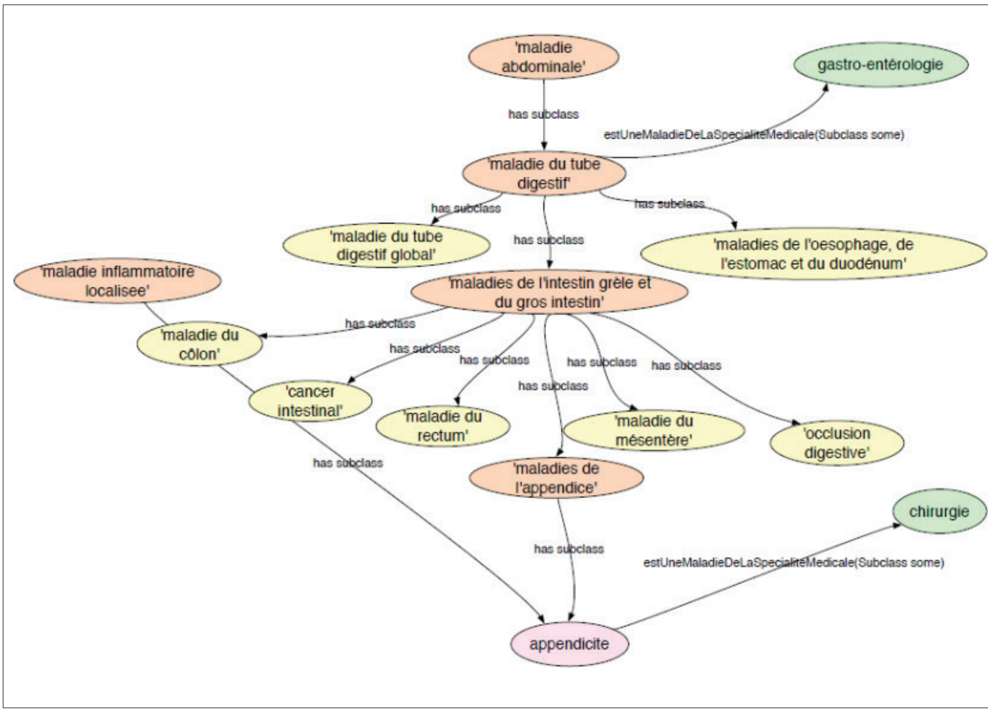


FIGURE 2. Ontology converging towards a diagnosis of appendicitis within the framework of the French Lerudi project. In the context of symbolic AI, knowledge about a medical field is organised into an ontology, that is, a structured set of concepts and their relationships, which describe a field while respecting the principles of formal language. An ontology provides a data model for reasoning within the domain in question.

matters is the compliance of a cognitive process with a number of rules. This shift is at work in the first health application of AI, MYCIN.

MYCIN and Watson: the seminal models

MYCIN is one of the first expert systems, developed at Stanford at the beginning of the 70s [7]. It consisted of a machine with causal inferences with almost 600 rules allowing the examination of the likely causes of an infection and suggest a treatment. It was designed with two perspectives: - to stimulate AI research, including graph theory, by emulating medical reasoning – to offer an effective support for solid data-driven diagnosis to physicians. At that time, the development of logical tools for health purposes was considered a key step toward more reproducible and less subjective medical practice. The first medical databases were also created around this time, with the appearance of Pubmed…MYCIN remains an ambivalent project, designed to imitate and at the same time help the physician. It has pioneered some practical aspects that are still relevant today regarding application design, such as the partnership between engineers and specialists in the field. However, this application has failed to settle into clincal practice. The reason for this however is not a lack of effectiveness or clinical relevance, but rather the underestimation of the time needed to update data. The main utility of MYCIN for physicians is for learning.

The application of AI to health has not escaped the winter of AI, and it took until the year 2010 and the introduction of Watson (IBM) [5] for this field to actually make a come-back with a different system. Watson is not a system dedicated to the medical sphere, but it is a versatile program suited for natural language exploration (Deep QA) that was first implemented in a game show, the Jeopardy, where candidates answer questions on general knowledge. Its computing power enables it to directly explore the web to find the right answer. This capacity can be transposed to the medical field to provide an aid for diagnosis: Watson uses the context of a case to generate a list of diagnoses. While MYCIN proceeds by induction, Watson is able to work by abduction: it formulates hypotheses based on correlations before verifying them on specific cases. This ability of the machine is in fact close to human «serendipity», i.e. the ability to discover general patterns from non-structured data [8].

AI for what medicine?

This new ability of deep learning to find correlations resonates with the need of a 4P medical model: predictive, personalised, preventive and participative. An application for participative medicine for example allows a patient to detect the possible malignancy of a dermatological lesion on his own, with a reliability comparable to that of an expert. Some applications of AI in health can be implemented in frequently used consumer devices (phone, connected watch, etc.) and their qualification as medical devices is a complex issue: one of the first uses of Watson in the medical sphere was to help manage the reception of patients at British hospital emergency centers with a «bot», i.e. a vocal assistant.

AI is also a useful adjunct for the personalisation of medical care. It permits the aggregation of complex data, including images, and allows the caregiver to have access to detailed and evolving characterisation of pathologies. The notion of a "digital patient" (N. Ayache) has implications at two levels: 1) at the clinical level, modeling with AI allows for individualised treatment based on a predictive model of the patient's reaction; 2) after anonymisation and data compression, the establishment of a large population-wide database is of major interest for medical research.

Is the future of research written ***in silico*** ?

Setting up such large population databases offers new research perspectives, either clinical research or in terms of public health. AI determines unsuspected correlations between phenotypes and genotypes, or between populations and their environment. It is well-suited to find weak signals, that might be otherwise overlooked, and encourages a proactive and cost-effective management of diseases. In pharmacological evaluation, deep learning and in silico data analyses offer alternative solutions to the reference methods (Cox regression) that rely on the comparison of a cohort with a control group. With the SCCS (Self Controlled Case Series) algorithms, the assessment of secondary effects is possible at the scale of the individual, by comparing the patient at the time an adverse effect occurs with the time period that preceded it. The work needed to homogenise the cohort is also reduced to a minimum. Such an approach proved useful in France for the withdrawal of the antidiabetic drug pioglitazone.

In silico studies are not only expected to evaluate solutions that analyse data but could also help in their development. The GAN (Generative Adversarial Network) makes it possible to complete a system of incomplete data: a system generates the missing data, whose relevance is assessed by a "rival" network. This kind of functioning of networks is approaching human creativity...

Today, genetic analysis of oncological markers offers a striking example of the complementarity between supervised and unsupervised learning approaches, their strengths and their limitations. Genome sequencing of cancers requires huge power for data analysis and storage: it takes between 10 gigabytes and several terabytes to store a single genome. Data modeling with hidden Markov chains allows for genome annotation and, for example, the determination of coding regions. It constitutes an approach of unsupervised learning that is flexible and well-suited to reconstruct the history of a cancer in terms of genomic alterations. It is also possible to track some of the processes that are involved, such as smoking, sun exposure, etc. thanks to the technology of matrix factorisation comparable to those used by video platforms to adjust their content in a personalised manner. This unsupervised learning approach does not prevent the use of supervised learning, for example with the aim of testing whether a profile predicts tumor recurrence. This classification approach nevertheless suffers from the discrepancy between the large volume of data per individual and the limited number of individuals characterised: its performance is poor in comparison to models that solely rely on clinical data.

AI and health in France

The Villani report was made public in 2018 and it is an important date for AI in health in France. Its writing occured in a context of mistrust toward robotics in general, and responded to the concern of ensuring a harmonious economic and academic development of the discipline. Akin to the launch of CNIL (French National Commission on Informatics and Liberty) in the 70s, an emphasis on the ethical use of data and transparency of algorithms as well as the use of «open source» softwares is advocated in order to overcome the public technophobia and converge with the European framework (GDRP, General Data Protection Regulation). Two requisits are essential:

- The need to boost the economic and industrial aspect in order to facilitate the implementation of relevant and sustainable initiatives in the long-term;
- The necessity to group public health data that are already huge into a single centralised system (*table 1*).

The top-down logic for health data grouping and the integration of the private sector are notably based on the Israeli model, i.e. that of a rich and dynamic ecosystem that values data. The choice of centralisation is anchored in the French practices and history, in contrast to the choices made at the same time by Germany, and interoperability of the two neighboring countries is not currently envisioned.

The launch of the PRAIRIE institute (PaRis Artificial Intelligence Research InstitutE), supported by a number of prestigious universities and institutions, corresponds to the need to develop some of the skills that might not always be technical, but that also correspond to the «soft skills» needed in AI engineering. This approach is highly relevant to AI in health where human interoperability, application design and user experience are essential.

TABLE 1. The main French databases and organisations that constitute the national health data system

DATABASES	TYPES OF DATA	ORGANISATION
SNIIRAM (national system of information)	▰ 15 thematic databases with aggregated data (datamarts) oriented toward a particular purpose: monitoring of medical expenses (Damir), analysis of the private care supply, biology, pharmacy, medical devices, private institutions; ▰ general sample of beneficiaries (EGB), with 1/97th of the population: the EGB enables the performance of longitudinal studies and analysis of health trajectories of approx. 660,000 beneficiaries regarding public hospitals as well as private care; ▰ database of individual data of beneficiaries (DCIR) for studies regarding healthcare expenditures	Cnam
Cépi-DC	National statistics on medical causes of death	Inserm/Insee
PMSI (Program of medicalisation of information systems)	Systematic collection of minimal administrative and medical data that are used primarily to finance healthcare facilities (activity based pricing) and for the organisation of healthcare (planning)	Hospitals

The most spectacular announcement was the creation of a National Health Data Hub, under a shared public/private governance and proposed to be quickly operative (from 2019). However, political and legal constraints have proven insurmountable: the strict standards of the CNIL (French National Commission on Informatics and Liberty) regarding data anonymisation required the use of a foreign service provider, Microsoft, rather than an French partner that would be new to health data management. This use has been challenged by a decree of the Council of State ... Because of these hesitations, the top-down logic was substituted with a bottom-up approach and the launching of regional database initiatives, including: the West Data hub: the first health database in France that emanated from the Health cooperation group (GSC) HUGO (including the University Hospital of Nantes, Angers and Ancenis); the «Entrepot des Données de Santé» (EDS) from Assistance Publique-Hopitaux de Paris (AP-HP).

Finally, AI and its integration into medical practive are also a concern to the National High Authority of Health (HAS). This authority issued a standard for the evaluation of medical devices with artificial intelligence in 2020, an initiative aimed to give a more precise and relevant legal framework for manufacturers of health devices in France. The initial step of validation of this line of products will remain dependent on the European (CE) marking, but the French HAS intends to maintain its evaluation of this family of medical devices, in order to improve knowledge and transparency of this technology.

The author would like to thank Harold Mouchere for his precious comments.

REFERENCES

1. Smolensky P.. IA connexionnisteIA cognitiviste et cerveau. In *: Introduction aux sciences cognitives,* ed. Daniel Andler. Paris : Gallimard, 1992
2. Ganascia JY. *L'intelligence artificielle.* Dominos. Paris : Flammarion 1993
3. Le Cun Y, Dehaene, Girardon J. *La Plus Belle Histoire de l'intelligence*. Paris, Odile Jacob, 2018.
4. Le Cun Y. *Quand la machine apprend. La révolution des neurones artificiels et de l'apprentissage profond*. Paris : Odile Jacob, 2019.
5. Jean A. Une brève introduction à l'intelligence artificielle. *Médecine sciences* ; 36 : 1059-67
6. Haiech J. Parcourir l'histoire de l'intelligence artificielle, pour mieux la définir et la comprendre. *Med Sciences* 2020 ; 36 : 919–23
7. Matuchansky. Intelligence clinique et intelligence artificielle : Une question de nuance. *Médecine/sciences* 2019, 35 : 797-803. Doi : 10.1051/medsci/2019158.
8. Korb KB, Nicholson A. *Bayesian artificial intelligence.* London : Chapman & Hall/CRC Press UK, 2004.
9. Negnevitsky M. *Artificial Intelligence: A Guide to Intelligent Systems,* Addison-Wesley, 2011.

10. Levas A, Bagchi S. Gondek D, Mueller ET. Watson: Beyond Jeopardy! *Artificial Intelligence* 2013 ;199-200 : 93-105

11. Raikomar A, Dean J, Kohane I. Machine Learning in Medicine. *NJOM* 2021 ; 380 : 1347-58

12. Nordlinger B, Villani C. *Santé et intelligence artificielle.* Paris : CNRS, 2018.

13. Ayache N. *Le Patient numérique.* Leçon inaugurale. Paris : Collège de France, 2013.

14. Cuggia M , Combes S. The French Health Data Hub and the German Medical Informatics Initiatives: Two National Projects to Promote Data Sharing in Healthcare. *Yearb Med Inform* 2019 ; 28: 195–202. doi: 10.1055/s-0039-1677917

Principles and applications of AI

This pictorial representation based on the author's assessment has but an indicative value

$$v(x)=\sum^{n} w_i x_i + w_{n+}$$

$$v(x)=w_1 x_1 + w_2 x_2 + w_3 = 0$$

$$v(x)=\sum^{n} w_i x_i + w_{n+}$$

$$J(W)=\sum_k J^k(X_k, Y_k, W)$$

$$J(W)=\sum_k J^k(X_k, Y_k, W)$$

$$J^k(X_k, Y_k, W) = -\sum_c y_c \log(\hat{y}_c$$

$$\hat{Y}=[\hat{y}_0, \ldots, \hat{y}_C$$

by Harold Mouchère
University of Nantes,
Laboratoire des Sciences du Numérique de Nantes

Artificial intelligence (AI) covers many concepts. This chapter is focused on one part of this large family of algorithms and approaches: deep neural networks. Any interested readers can refer to [1] for further reading

Basic Principles of deep neural networks

There are many architectures for neural networks, but they all have two things in common: their basic building block is the formal neuron and the gradient descent algorithm is used for learning. We will explain these fundamental notions and we will define modern architectures: convolutional networks and recurrent networks.

From the isolated neuron to the multi-layered perceptron

Since the 1950s, researchers working in the field of cybernetics have tried to reproduce the behavior of a biological neuron. The modern artificial neuron still looks like the one proposed by Rosenblatt in 1957, as shown in *figure 1*.

The formal neuron is made up of n inputs numbered from 1 to n. These inputs may be the extracted characteristics (automatically or not) from the shape that needs to be recognised. Each entry is associated with a weight wi. The «body» of the neuron performs a linear combination of these inputs with their respective weights. An additional supplementary input, always set at 1, allows the addition of a bias to the sum called the potential function $v(x)=\sum_1^n w_i x_i + w_{n+1}$. This potential then goes through an activation function that makes the decision. Different functions can be applied depending on the use of the formal neuron.

The formal neuron can therefore be summed up with a simple equation: y=f(v(x)). *Figure 2* provides a representation of this function for a two-dimensional problem: (x1,x2) are the two characteristics and the equation v|x|=w1x1 + w2x2 + w3 =0 represents the decision boundary. The main difficulty consists in finding values for the coefficients (w1, w2, w3) that minimise the number of errors (the points on the wrong side of the line). For this, we use learning algorithms that we explain in the next section.

A single neuron will only allow to solve problems that are linearly separable and in two classes. For more complex problems, it is possible to perform iterative projections of the data in different spaces with the aim of transforming the problem until it becomes linearly separable. Each projection is done using a layer of neurons within a multi-layer perceptron (MLP). As shown in *figure 3*, an MLP is composed of an input layer (the X vector of entries), several hidden layers performing space permutations, and finally the output layer known as the decision layer. These different layers are completely connected, i.e. each neuron is connected to all the neurons of the former layer with a weight. The final decision layer can be seen as a set of linear separators with two classes: each neuron tries to separate its class from all the others. If the activation function used in the neurons of the hidden layers is not linear, then the new space of representation, called latent space, allows the next layer to solve the problem linearly. This change in space of representation constitutes a strength of neural networks.

This architecture is limited by its complexity. Indeed, the number of connections (and therefore parameters to solve) increases with a polynomial order depending on the number of features analysed, the number of neurons in the hidden layers and the number of outputs.

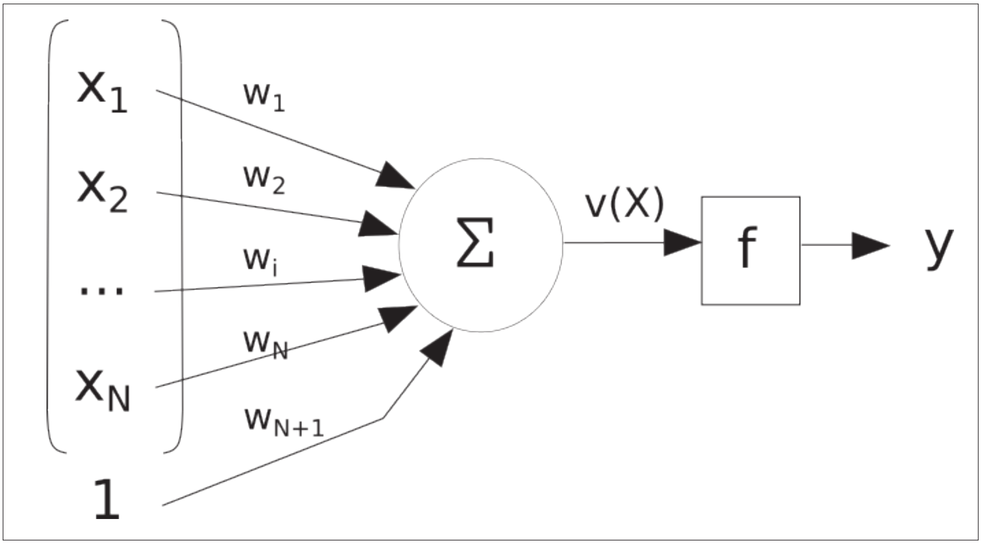


FIGURE 1. The formal neuron: the inputs xi are weighted by wi, this potential v(x) is passed to the activation function leading to the output y.

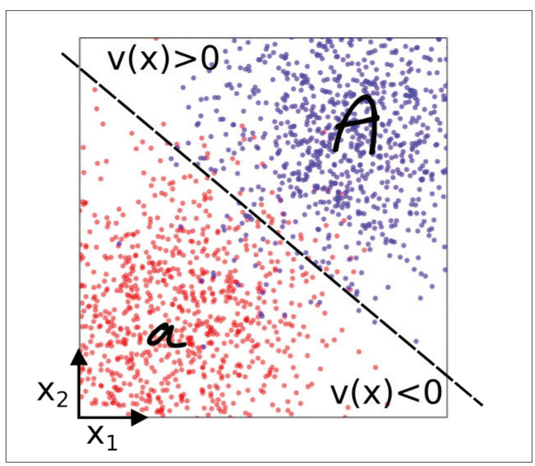


FIGURE 2. A single neuron can be regarded as a linear separator: 2D example of a problem that is not lineairly separable.

Machine learning

The algorithms that allow to find the best weights from well-labelled examples qualify as supervised learning. The expected input and output data pairs, labelled (Xk, Yk) constitute the basis for learning. By randomly using the examples one by one, it is possible to gradually estimate the best weights, following three steps:

- Calculate the output of the neural network for the C classes: $\hat{Y}=[\hat{y}_0,\ldots,\hat{y}_C]$,
- Estimate the error of the current configuration of weights *W : J(W)*,
- Modify the weights W to reduce this error

The algorithm used to carry out this learning is called stochastic gradient descent. In order to use this algorithm, it is necessary to first define a cost function J(W) that measures the amount of error committed by the network with the whole of the data:

$$J(W)=\sum_k J^k(X_k,Y_k,W)$$

The cost function that is used must be adapted to the problem that we want to solve. In the case of regression, a simple mean of quadratic errors is usually used. One can also use an approach based on probability density using the Kullback-Leiber cost function. When the classifications analysed correspond to a number of outputs ycs with probabilities adding up to 1, a cross entropy approach is generaly used:

$$J^k(X_k,Y_k,W)=-\sum_c y_c \log(\hat{y}_c$$

A key step is to find the set of weights W that minimise the value of the cost function chosen from the learning dataset. For this, the value of the weights can be modified in the oposite direction of the cost gradient:

$\hat{y}_c$

$$\Delta W_k=-\eta\frac{\partial J^k}{\partial W}$$

This calculation depends on the cost function used, the choice of the activation function of each neuron, the connections between the neurons and a parameter called the learning rate. A few years ago, these calculations and their implementation were complex, but they are now completely transparent for users of the latest software libraries. One potential pitfall of this learning strategy is called gradient vanishing. At each level, the quantity of gradient is dispersed in the preceding neurons. Using the Rectified Linear Unit (ReLu) reduces this effect.

Convolutional Networks and recurrent networks

Because of the complexity of the fully connected layers, it is impossible to directly apply this to images that constitute an input that is too large. A solution popularised by LeCun in 1998 [2] is to use a convolutional network that enormously reduces the number of parameters. As shown in *figure 4*, a convolution kernel, which can be seen as a small perceptron, is applied on a small portion of the image. The number of weights necessary for this local calculation is reduced because the neuron is not connected to the whole image. This kernel is applied to all parts of the image by sliding it along the lines and columns (the convolution step). With a single kernel and a reduced set of weights, it becomes possible to extract local information on the whole image. One can in parallel apply several tens of kernels, each one extracting a local characteristic from the image. Multiple convolution layers are then stacked, each producing new features used as input for the following layer. The corresponding maps can be reduced by local aggregations (Sum-pooling or max-pooling). Thus, the successive layers extract higher level information, with less complexity. A fully connected final layer allows these characteristics to be used for the final decision. Learning of these convolutional networks is done exactly as explained above, but they take advantage of sharing weights within the convolutional layers to quickly learn local structures. Their implementation is also highly parallelisable, and using GPU cards facilitates their use.

Convolutional networks were created to handle 2D inputs (images), but their use has been extended to volumes (3D convolutions) and sequences (2D + t or 3D + t). Nevertheless, analysing a sequence sometimes requires taking into account a long period. Recurrent networks allow for propagation of information along the sequence.

A recurrent network also uses the principle of convolution because the same neuron is applied along the input signal. But in addition to only seeing a portion of this signal, the activation of the neuron is calculated by using the output of the neuron(s) at the previous step. Several formulations of recurrent neurons have been proposed: GRU (Gate Recurrent Unit) and LSTM (Long Short Term Memory).

Recurrent networks can be used for different tasks, classifying sequences being the simplest: the network is a single extractor of characteristics that feeds a fully connected final decision layer. The recurrent networks also enable making a sequence of decisions, for example suited for sequential tasks, such as automatic language translation.

Example of application in gastroenterology

Deep networks are used for multiple applications in the medical field [3]. We have selected an application in gastroenterology [4] for the detection and classification of pathological images obtained from capsule endoscopy in patients with Crohn’s disease.

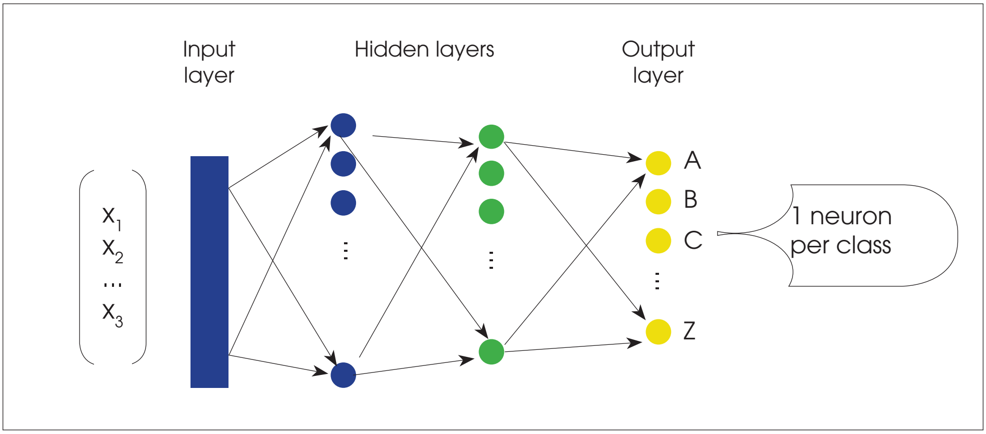


FIGURE 3. A multi-layer percepton (MLP). Each neuron is connected to all neurons of the former layer.

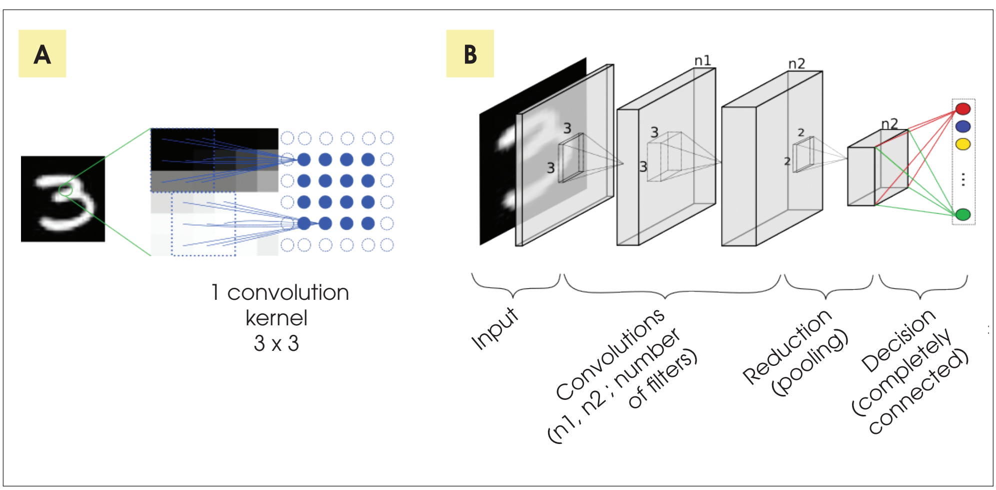


FIGURE 4. Convolution network. A.) a 3x3 convolution applied to a 2D image. B.) A complete simplified network: a 2D image as input, 2 convolution layers, 1 pooling layer and a completely connected decision layer.

It was first necessary to collect and annotate the images (63 patients, 3498 images) for different kinds of lesions (6 types of lesions) and non-pathological images. The first difficulty comes from labelling. Several lesions may appear on an image (1.2 per image on average). The data have been labelled by 3 experts, and in more than 15% of the images, one of the experts did not agree (consensus by voting). The proportion of the different classes is also a challenge because learning is stochastic and works best when the classes are balanced (which is rarely the case). After agreement, 60% of images are not pathological and not all lesions are detected at the same frequency (from 5.8% for stenoses to 14.6% for ulcers<10mm). The database may therefore be used for detection (2 levels) or identification of a lesion (7 classes).

Several architectures have been tested among the «on shelf» solutions: ResNet and VGG, which are the networks most often used in the image analysis community. For example VGG19 [5] uses a 224 x 224 image as input and it is composed of 16 layers of convolution (from 64 to 512 kernels of 3 x 3 pixels), interspersed with five layers of max-pooling for local aggregation of the extracted information. The last layer is passed through three fully connected layers, with 4096 neurons in the first two, the last layer doing the classification (the 2 or 7 classes of our problem). There are approximately 140 million parameters to learn.

The evaluation was made by cross-validation so as not to be dependent on data labelling during the learning-validation-test. Considering the problem of detection of pathological images, the best system (ResNet34) obtains an average performance of 94.56% on the 5 validations and an area under the curve ROC of 98.23%.

Tomorrow's challenges

To date, the approaches that proved useful in their various application fields are supervised systems, that is to say those that use data labelled by human experts for a dedicated task. The constitution of large databases (public or private) therefore makes it possible to tackle different applications. Ongoing reflections regarding data property and confidentiality, especially medical data, are therefore crucial points.

In parallel, AI researchers aim to reduce the amount of labelled data necessary for learning. Two solutions have emerged in recent years. The first solution is unsupervised learning, or at least weakly supervised learning, and consists of providing raw unlabelled data to the network, to let it discover a latent space of reduced dimension that could be useful for other tasks. The transfer of knowledge is pursuing the same goal: building a representation space using a similar problem for which a large amount of data is available.

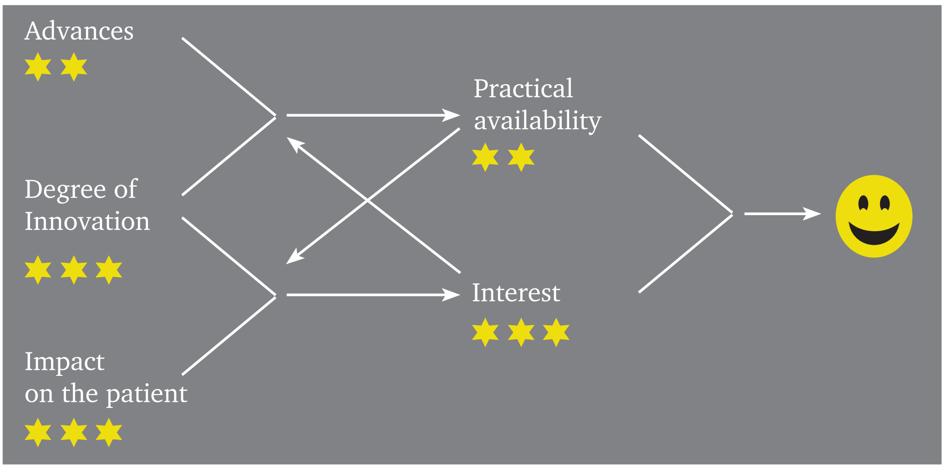
With the popularisation of its use, the user acceptability of decisions made by AI is an essential problem, especially in the medical field. This problem is often addressed by the creation of explainable AI, that would be able to convince the user, expert or not, of the correctness of the decision. One of the paths explored for deep networks, which are regarded as black boxes, is to use the attention process, in order to be able to delineate the portion or structure that allowed to make the decision from the input signal.

REFERENCES

1. Goodfellow I, Bengio Y, Courville A. *Deep Learning*, MIT Press, 2016.
2. Lecun Y, Bottou L, Y. Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 1998 ; 86 : 2278-24, 1998, doi: 10.1109/5.726791.
3. Esteva A, Chou K, Yeung S *et al.* Deep learning-enabled medical computer vision. *npj Digit Med* 2021 ; 4, 5. doi : 10.1038/s41746-020-00376-2
4. Vallée R, De Maissin A, Coutrot A, Mouchère H, Bourreille A, Normand N. CrohnIPI: *An endoscopic image database for the evaluation of automatic Crohn's disease lesions recognition algorithms*. Proc. SPIE 11317, Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging, 113171Q (28 February 2020); doi : 10.1117/12.2543584
5. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv* 2014. arXiv:1409-1556.

6. Klare P, Sander C, Prinzen M, *et al.* Automated polyp detection in the colorectum: a prospective study (with videos). *Gastrointest Endosc* 2019 ; 89: 576-82.e1. doi : 10.1016/j.gie.2018.09.042.
7. Repici A, Badalamenti M, Maselli R, *et al.* Efficacy of Real-Time Computer-Aided Detection of Colorectal Neoplasia in a Randomized Trial. *Gastroenterology* 2020 ; 159 : 512-20.e7. doi : 10.1053/j.gastro.2020.04.062.
8. Lui TKL, Guo C-G, Leung WK. Accuracy of artificial intelligence on histology prediction and detection of colorectal polyps: a systematic review and meta-analysis. *Gastrointest Endosc* 2020 ; 92 : 11-22.e6. doi : 10.1016/j.gie.2020.02.033.
9. ASGE Technology Committee, Abu Dayyeh BK, Thosani N, *et al.* ASGE Technology Committee systematic review and meta-analysis assessing the ASGE PIVI thresholds for adopting real-time endoscopic assessment of the histology of diminutive colorectal polyps. *Gastrointest Endosc* 2015;81:502.e1-502.e16. doi : 10.1016/j.gie.2014.12.022.
10. Byrne MF, Chapados N, Soudan F, *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* 2019;68:94-100. doi : 10.1136/gutjnl-2017-314547.
11. Mori Y, Kudo S-E, Wakamura K, *et al.* Novel computer-aided diagnostic system for colorectal lesions by using endocytoscopy (with videos). *Gastrointest Endosc* 2015 ; 81 : 621-9. doi : 10.1016/j.gie.2014.09.008.
12. Kudo S-E, Misawa M, Mori Y, *et al.* Artificial Intelligence-assisted System Improves Endoscopic Identification of Colorectal Neoplasms. *Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc* 2020 ; 18 : 1874-81.e2. doi : 10.1016/j.cgh.2019.09.009.
13. Gong D, Wu L, Zhang J, *et al.* Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. *Lancet Gastroenterol Hepatol* 2020 ; 5 : 352-61. doi : 10.1016/S2468-1253(19)30413-3.
14. Zhou J, Wu L, Wan X, *et al.* A novel artificial intelligence system for the assessment of bowel preparation (with video). *Gastrointest Endosc* 2020 ; 91 : 428-35.e2. doi : 10.1016/j.gie.2019.11.026.

AI and upper digestive tract : Revolution is coming!



This pictorial representation based on the author's assessment has but an indicative value

by Emmanuel Coron [1] & Gabriel Rahmi [2]
[1] *Institute of diseases of the digestive system, CHU Nantes; University of Nantes*
[2] *Georges Pompidou Hospital, AP-HP ; University Paris-Descartes*

On first instance, AI seems more advanced for the small intestine and the colon, for different reasons. The interest of automatic recognition of lesions in capsule endoscopy is obvious and led to precursor studies and applications. Likewise, better detection of colonic polyps leads to a reduction in the risk of colorectal cancer and better differentiation between adenomas and hyperplastic polyps, permitting the adoption of a resect-and-discard strategy that has a major benefit in terms of cost-effectiveness. Thanks to these economic opportunities, small intestinal and colonic lesions were the first to motivate the interest of academic and industrial researchers and start the application of AI in endoscopy. The relatively homogeneous nature of the lesions to recognise has also probably facilitated the obtention of the initial impressive results that were reported in the first published studies.

Nevertheless, the upper digestive tract has witnessed several developments in AI. There are also huge expected benefits from its application. Upper digestive cancers are common and have dire prognosis, which requires their detection at an early stage in oder to increase the chances of applying curative approaches, or at least obtaining prolonged survival in these patients. When detected early, the survival rates in patients with upper digestive cancers are at more than 90% [2].

Integrating AI in the context of rapid progress of endoscopic techniques and in real life?

Endoscopic techniques are quickly evolving and the successive generation of optical lenses or digital chromoscopy (Blue Light Imaging [BLI], Linked-Colour Imaging [LCI, Narrow Band Imaging [NBI]) unveil the process of carcinogenesis at its earliest stages. One of the best examples is the visualisation of Intrapapillary capillary loops [IPCL] in the oesophageal squamous mucosa, whose appearance changes through the different stages of dysplasia and mucosal invasion [3]. The challenge is not so much to track these anomalies, but rather to interpret them with sufficient expertise, as was recently reminded by the European Society of Digestive Endoscopy (ESGE) [4]. Similarly, various precancerous lesions are identifiable in Barrett's mucosa using optical lenses and optical or digital dyes. Recognising the different subtypes of lesions is difficult. It then becomes obvious that the main difficulty is not so much to master a technique such as submucosal dissection, but rather to be able to select the lesions that are the best candidates for this treatment and to properly delineate the margins and the risk of deep mucosal infiltration before this interventional procedure is proposed. There is a glaring difference between the Western world and countries in Asia, where the careful use of innovative optical diagnostic tools has enabled a reduction in the mortality from upper gastrointestinal cancers [5-7].

The complex learning of endoscopic semiology with these new tools should not lead us to forget that, in the majority of centers in the world («in real life»), endoscopists are not experts and they use a simple approach with white light and endoscopes that are not equipped with optical zoom. Moreover, the variety of «basal» lesions that can be found in the oesophagus or the stomach complicates automatic recognition with AI programs. In fact, an important issue is to simultaneously avoid missing risky lesions identified during screening and to facilitate the characterisation of complex lesions that may require expert intervention. AI has already demonstrated its ability to identify microstructures and quantify architectural patterns at the pixel level within an image, i.e. at a level that is undetectable by the human eye. This superiority is further accentuated by the robustness of computers, that are insensitive to fatigue, stress or boredom caused by repetitive procedures.

Applications of AI in oesophageal pathology

Barrett's oesophagus

Barrett's oesophagus (BE) presents a particular difficulty. Dysplasia may be present as a thickened or flat area, with architectural irregularity of the mucosa or its microvessels, a fact that complicates the precise delineation of the lesions. BE is rare in Asian countries, and therefore no Chinese or Japanese groups working on AI have worked on it yet. However, as is often the case in endoscopic work on BE, the Dutch have played a pioneering role [9,10]. The GastroNet portal, emanating from the consortium ARGOS, is a large international database that has helped to group and sort 494,364 endoscopic images, 1,544 of which are from dysplastic and non-dyplastic BE. The AI system that was then set-up was validated with two different datasets each including 160 patients. The performance of the algorithm was compared with that of a panel of 53 international endoscopists, all of whom were beaten by AI, as shown in a study based on a retrospective analysis of the evaluated images [9]. This algorithm was then incorporated into a real-time video endoscopy system in a pilot study of 20 patients (n=10 dysplastic BE vs n=10 non-dysplastic BE). The comparison of AI and the human eye was only based on white light images, according to a standardised protocol used for the exploration of BE. From the distal portion to the proximal portion of the BE, the endoscope was stopped every 2 cm, and three images were registered by the endoscopist at each level and analysed by the AI system. This strategy permitted the evaluation of the reproducibility of the predictions made by AI for a given segment, and to see whether an analysis combining the different levels would improve the system's performance. In addition to identifying areas of possible dysplasia, AI allowed for the contouring of the area and thus helped the endoscopist to simultaneously perform targeted biopsies or an endoscopic resection. The system reached a concordance of 75% between the 3 images at each level. Image analysis reached a diagnostic performance of 84%, sensitivity of 76% and specificity of 86%. When the system identified an image with a high confidence index, the corresponding values reached 91%. Overall, the system detected 9 patients out of 10 that the endoscopist had identified as «dysplatic BE». The mucosectomy specimen obtained from the one patient «missed» by AI did however only include non-dysplastic Barrett's mucosa, raising the problem of the gold standard analysis. In this pilot study, the AI system was very fast, with a delay of 0.2s and 0.3s for identifying and contouring potentially dysplastic lesions. This system is therefore extremely promising and it is currently being evaluated in a prospective international study *(figure 1)*.

Squamous cell carcinoma and AI

The detection of oesophageal squamous cell carcinoma (OSCC) is another important issue, due to the high prevalence of this cancer in some parts of the world and its dire prognosis. The study by Guo *et al*. [11] examined the possibility of detecting superficial OSCC by AI combined with Narrow band imaging (NBI) (*figure 2*). This Chinese AI system was trained to recognise the corresponding images from 6,473 images of OSCC and non-dysplastic oesophageal mucosa. The first dataset used for validation included OSCC images (n=1480) vs diverse images (n=5191) including normal mucosa, oesophagitis, heterotrophy of gastric mucosa, submucosal tumor and even oesophageal varices. The majority of lesions were flat or slightly depressed, but relatively large in size (34 mm on average). In this dataset, the sensitivity reached 98% and specificity 95%. Of course, a potential bias is related to the choice of images analysed by the AI system. For this reason, a second dataset based on videos (n=27 without zoom and n=20 with zoom) was also tested. Irrespective of the use of zoom, the lesion was detected with full sensitivity of 100% (using one or several images). The analysis by image showed that the performance were significantly higher using images with zoom (sensitivity 96.1%) compared to those made without zoom (sensitivity 60.8%). According to the authors, an increased risk of artifacts was produced by movement in the absence of zoom. In a last validation set comprised of 33 cases, the entire video (non-filtered, 80 seconds in length on average) was analysed by the system. In this last set, the image and patient specificity reached 99.9% and 90.9%, respectively. This study is iconic because, despite many potential biases (single center study, the use of selected images and videos), it addressed highly heterogeneous lesions, either OSCC or controls. In addition, it is one of the first to integrate the approach of chromoscopy combined with zoom, in this case for the analysis of IPCL. This study therefore demonstrates that AI is able to identify hard-to-interpret images, at least for European endoscopists.

Applications of AI in gastric pathology

In the stomach, most studies to date have examined the possibility of detecting early stages of cancer with white light. The Chinese multicenter study published by Luo et al. [12] aimed to set up a system called GRAIDS (Gastrointestinal Artificial Intelligence Diagnostic System). In this study, 1,036,496 images obtained from 8,424 patients were used to develop and test GRAIDS. The initial phase consisted of

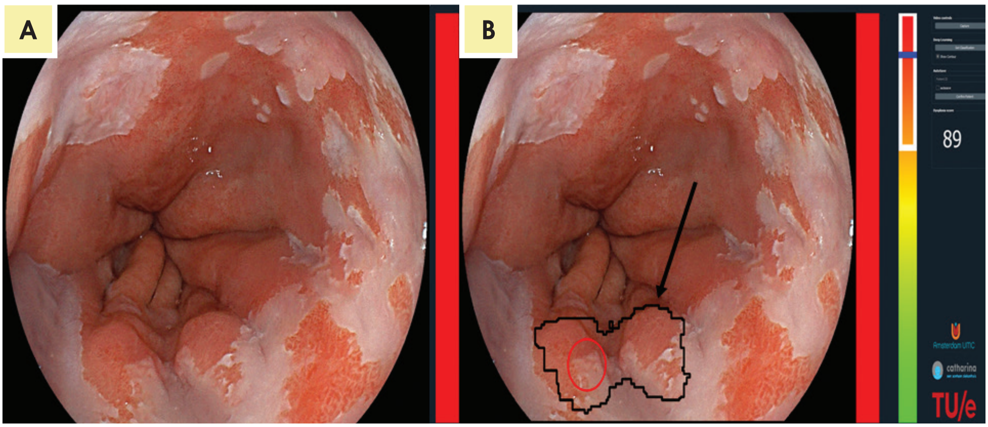


FIGURE 1. Barrett's Oesophagus (BO) examined in white light without AI (A) and with the contribution of AI (B). The AI system shows the area suspected of malignant degeneration (black) and a preferential location to target for biopsy (red). The lateral bar provides an estimation of the confidence index for the prediction of neoplasia (in this case, 89%).

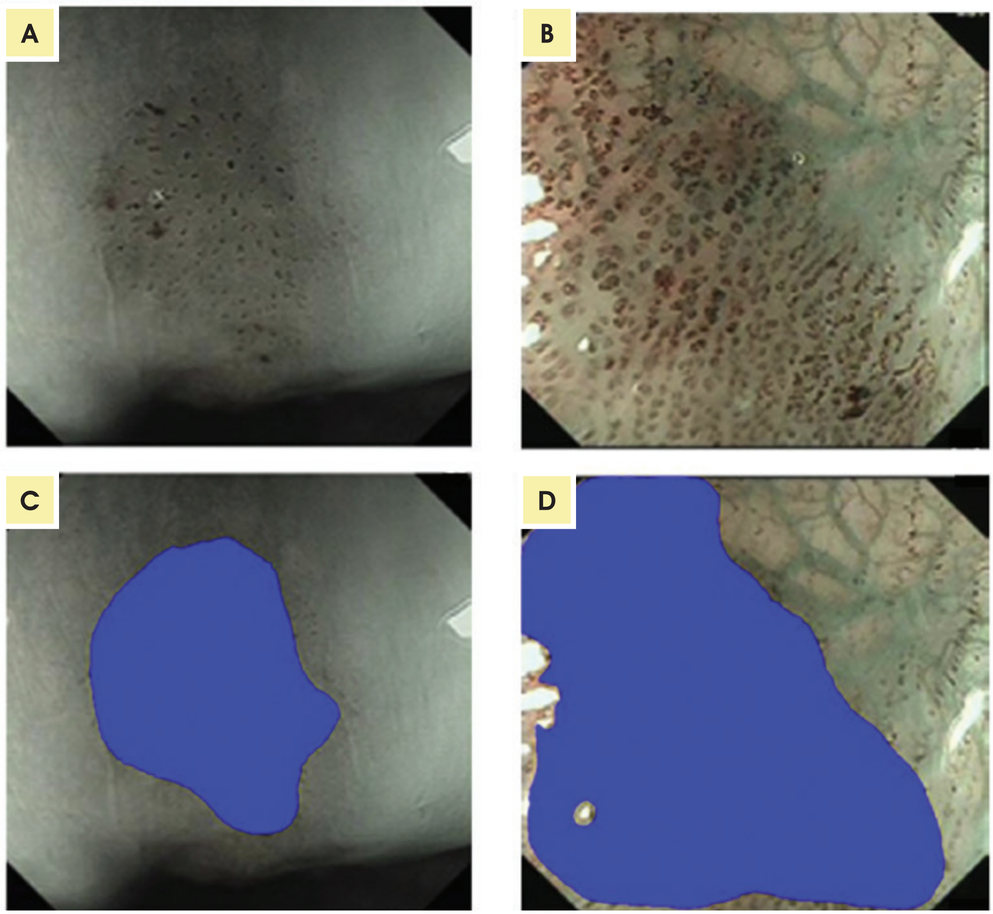


FIGURE 2. Oesophageal Squamous Cell Carcinoma (OSCC) examined in white light without AI (A, B) and with the contribution of AI (C,D). The AI system indicates the area at risk and suggests where to perform a biopsy or resection (blue). Evaluation with the zoom mode (D) further increases the system's performance, and might in the near future help to automatically estimate the depth of infiltration.

training the system, before a stage of intrinsic verification and internal validation. The second phase allowed for a first prospective assessment in a national hospital center, followed by a second prospective evaluation in five general hospitals. The performance of GRAIDS was compared to that of endoscopists with three different levels of expertise (expert, competent, in training). The diagnostic performance of GRAIDS was 93% in the setting of a national hospital center, similar to that measured in the five general hospitals (92-98 %). Importantly, this AI system had a sensitivity comparable to that of an expert endoscopist (94%), higher than that of a competent endoscopist (86%) or an endoscopist in training (72%). Rather than competing with the human eye, combining GRAIDS with human interpretation improved the sensitivity of endoscopists for the detection of cancers, modestly for experts (98%) but in a greater manner for competent endoscopists (98 %) and endoscopists in training (96%). False positives (mucus, abnormalities of gastric surface related to gastric contraction) were easily identified with the human eye and therefore did not lead to excessive biopsies. The system is also quite fast (25 images analysed per second in video in real time), therefore allowing its use in clinical practice. It is important to note that this system also allowed for the identification of oesophageal cancer using white light, whose incidence in Europe is similar to that in China, making this approach an attractive alternative in a European cohort

Conclusions and perspectives

The literature on artificial intelligence (AI) on the upper digestive tract is quickly evolving. The first meta-analyses, such as that by D'Arribas et al. that has brought together 19 studies with 1,116 patients (with almost 24,000 images) suffering from OSCC, cancer ocurring in Barrett's mucosa or gastric cancer, are encouraging [1]. Taking together these different cancers, AI had a sensitivity of 90%, specificity of 89%, positive predictive value of 87% and negative predictive value of 91% for the detection of lesions. This is suggestive of great future utility in clinical practice, keeping in mind that the studies at this stage are essentially pilot studies, performed using AI systems that will improve with time.

In practice, the commercialisation of systems dedicated to the detection and soon, to the characterisation of lesions of the upper digestive tract, is imminent. It is however important to keep in mind that AI can only be a companion for the endoscopist, who remains the only judge that can thoroughly analyse the gastroduodenal mucosa, perform targeted biopsies or resections. AI will never identify atypical lesions or diagnostic traps as well as an experienced endoscopist. It will also be necessary to overcome a number of ethical, regulatory and economic obstacles to the use of AI in routine practice, and to continue to assess the benefits of this technology using methodologically sound studies

REFERENCES

1. Arribas J, Antonelli G, Frazzoni L, Fuccio L, *et al.* Standalone performance of artificial intelligence for upper GI neoplasia: a meta-analysis. *Gut* 2020 : epub. doi : 10.1136/gutjnl-2020-321922.
2. Veitch AM, Uedo N, Yao K, East JE. Optimizing early upper gastrointestinal cancer detection at endoscopy. *Nat Rev Gastroenterol Hepatol* 2015 ; 12 : 660-7. doi : 10.1038/nrgastro.2015.128.
3. Inoue H, Kaga M, Ikeda H, *et al.* Magnification endoscopy in esophageal squamous cell carcinoma: a review of the intrapapil-lary capillary loop classification. *Ann Gastroenterol* 2015 ; 28 : 41-8.
4. Dekker E, Houwen BBSL, Puig I, *et al*. Curriculum for optical diagnosis training in Europe: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. *Endoscopy* 2020 ; 52 : 899-923. doi : 10.1055/a-1231-5123
5. Chiu PWY, Uedo N, Singh R, et al. An Asian consensus on standards of diagnostic upper endoscopy for neoplasia. *Gut* 2019 ; 68 : 186-197. doi : 10.1136/gutjnl-2018-317111.

6. Hamashima C. Systematic Review Group and Guideline Development Group for Gastric Cancer Screening Guidelines. Update version of the Japanese Guidelines for Gastric Cancer Screening. *Jpn J Clin Oncol* 2018 ; 48 : 673-83. doi : 10.1093/jjco/hyy077

7. Jun JK, Choi KS, Lee HY, *et al*. Effectiveness of the Korean National Cancer Screening Program in Reducing Gastric Cancer Mortality. *Gastroenterology* 2017 ; 152 : 1319-328.e7. doi : 10.1053/j.gastro.2017.01.029.

8. de Groof AJ, Struyvenberg MR, van der Putten J, *et al*. Deep-Learning System Detects Neoplasia in Patients With Barrett's Esophagus With Higher Accuracy Than Endoscopists in a Multistep Training and Validation Study With Benchmarking. *Gastroenterology* 2020 ; 158 : 915-29.e4. doi : 10.1053/j.gastro.2019.11.030

9. de Groof J, van der Sommen F, van der Putten J, *et al*. The Argos project: The development of a computer-aided detection system to improve detection of Barrett's neoplasia on white light endoscopy. *United European Gastroenterol J* 2019 ; 7 : 538-47. doi : 10.1177/2050640619837443.

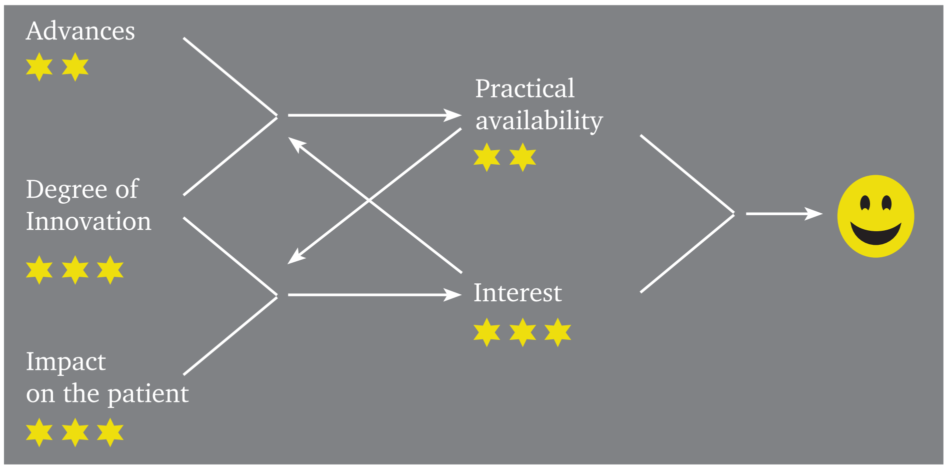
10. de Groof AJ, Struyvenberg MR, Fockens KN, *et al*. Deep learning algorithm detection of Barrett's neoplasia with high accuracy during live endoscopic procedures: a pilot study (with video). *Gastrointest Endosc* 2020 ; 91 : 1242-50. doi : 10.1016/j.gie.2019.12.048.

11. Guo L, Xiao X, Wu C, *et al.* Real-time automated diagnosis of precancerous lesions and early esophageal squamous cell carcinoma using a deep learning model (with videos). *Gastroin-test Endosc* 2020 ; 91 : 41-51. doi : 10.1016/j.gie.2019.08.018

12. Luo H, Xu G, Li C, *et al*. Real-time artificial intelligence for detection of upper gastrointesti-nal cancer by endoscopy: a multicentre, case-control, diagnostic study. *Lancet Oncol* 2019 ; 20 : 1645-54. doi : 10.1016/S1470-2045(19)30637-0

AI applied to videocapsule endoscopy of the small bowel: State of the art and perspectives

This pictorial representation based on the author's assessment has but an indicative value



by Xavier Dray [1,2], Aymeric Histace [2] et Romain Leenhardt [1,2]
[1] *Sorbonne University, Digestive Endoscopy Center ,*
Saint-Antoine Hospital, AP-HP, Paris
[2] *ETIS UMR 8051 (CY Paris Cergy University, ENSEA, CNRS), Cergy*

The indications for videocapsule endoscopy (VCE) of the small bowel (SB) are part of International recommendations for the exploration of unexplained digestive bleeding (UDB) and the suspicion of Crohn's disease (CD). The SB-VCE is also indicated in some cases of intestinal polyposis and refractory celiac disease. The gastroenterologists spend 30 to 120 minutes on examining and interpreting complete SB-VCE records consisting of tens of thousands of images. This step is tedious, often monotonous and demanding. It requires dedicated time schedules and sustained concentration. Recently, Beg et al. have shown that the performance of an observer declined after the examination of just one SB-VCE, eventhough experienced users reported an average of 3.4 examinations per session [1].

It is becoming clear that artificial intelligence (AI) solutions will have a major impact in medicine, especially for imaging and digestive endoscopy, and in particular for SB-VCE given its time-consuming and tedious interpretation [2]. These solutions have already entered the market, and some are already available for colonoscopy and SB-VCE [2].

The present review examines the current data and the future perspectives of AI in SB-VCE, as well as potential obstacles to its full implementation. .

The specificities of AI applied to VCE

In conventional endoscopy, the operator maneuvers the endoscope according to what is observed on the screen. In this context, AI solutions help in the detection and characterisation of lesions, synchronously to human observation. The main technological challenge of AI in this context is to display the results with accuracy and synchronously, i.e. as soon as the images are captured (24-60 frames per second).

AI applied to SB-VCE faces the opposite challenge. Delayed examination of SB-VCE images with respect to their capture is the rule, either when it is carried out by a human or by an AI solution. The time constraint on image processing is therefore not the main problem. On the other hand, the transit of the capsule along the digestive tract is out of the control of the operator; in the near future, AI will select a limited number of images to be analysed by the endoscopist (say 2%), while the 98% remaining images will not be read by the operator. In total, the main objective of AI in SB-VCE is to achieve high sensitivity for the detection of abnormalities while significantly reducing the time spent on reading (in connection to the issue of specificity and the number of "false positive" images that the operator will have to invalidate). This quest for high sensitivity primarily relies on machine learning (oriented by the choice of human experts), while the help with characterisation relies mostly on deep learning approaches (without the intervention of human experts)

Detection of lesions and abnormalities

The detection of lesions and abnormalities is the topic of *Table 1* and *Figure 1*.

Blood

Before the introduction of AI, the performances of a suspected blood indicator (SBI) were very sensitive (96%) but poorly specific (17% to 65%) [3].

Since then, Xing et al. proposed a method based on deep learning (DL) that is highly supervised with sensitivity, specificity and accuracy superior to 98% [4]. More recently, Aoki et al. have developed a less supervised algorithm, based on neural network

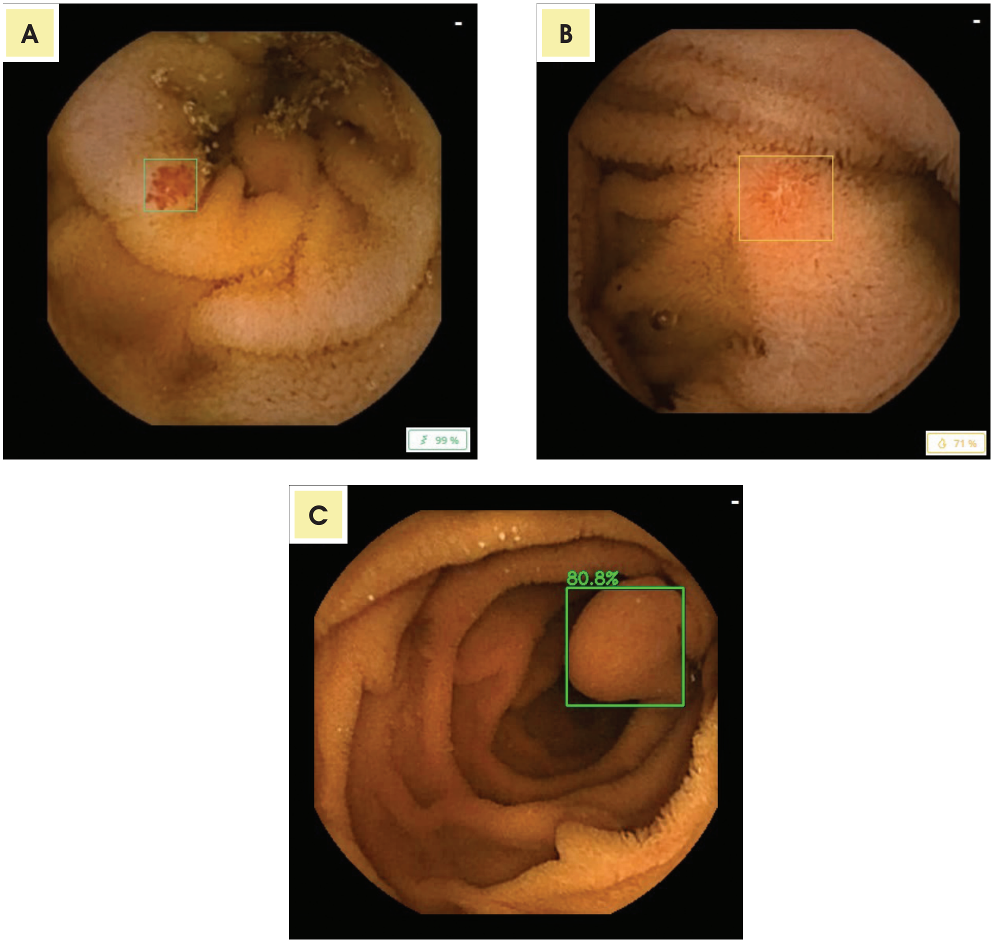


FIGURE 1. Examples of lesions detected by the artificial intelligence system Axaro® (Augmented Endoscopy®, France) devoted to endoscopic videocapsules. The detected lesions are symbolised by a pictogram (vascular, inflammatory, tumoral...) with a degree of confidence for the diagnosis (percentage) displayed in the thumbnail. A) Angiectasia; B) aphthoid erosion; C) submucosal tumor.

TABLE I. An overview of the main recent studies on AI applied for the detection and/or characterisation of lesions and abnormalities in videocapsule endoscopy.

First author	Target	Number of patients/images	Study design
Xing X, 2018	Sang	30 patients 1000 images	Retrospective
Aoki T, 2020	Sang	41 patients 49.191 images	Retrospective
Noya F, 2017	Angiectasias	36 patients 1.648 images	Retrospective
Leenhardt R,2019	Angiectasias	408 patients 1.200 images	Retrospective
Tsuboi A,2020	Angiectasias	48 patients 10.488 images	Retrospective
Fan S, 2018	Ulcerations	144 patients 21.160 images	Retrospective
Aoki T, 2019	Ulcerations	180 patients 15.800 images 16 patients	Retrospective
Wang S, 2019	Ulcerations	1504 patients 47.202 images	Retrospective
Klang E, 2020	Ulcerations	49 patients 17.640 images	Retrospective
Saito H, 2020	Protruding lesions	385 patients 48.091 images	Retrospective
Iakovidis D, 2014	Multiclass	251 patients 1.370 images	Retrospective
Ding Z, 2019	Multiclass	6,970 patients 113 millions images	Retrospective

*Average interpretation time.

Training/ Validation	Images/video	Performance
Yes	Still images	Sensitivity: 98,5% Spécificité : 99,5%
Yes	Still images	Sensitivity: 96,6% Specificity: 99,9%
Yes	Still images	Sensitivity: 89,5% Specificity: 96,8%
Yes	Still images	Sensitivity: 100% Specificity: 96%
Yes	Still images	Sensitivity: 98,8% Specificity: 98,4%
Yes	Still images	Sensitivity: 96,8% Specificity: 94,8%
Yes Non	Still images Videos	Sensitivity: 88,2% Specificity: 90,9% 3 minutes reading *
Yes	Still images	Sensitivity: 89,7% Specificity: 90,5%
Yes	Still images	Sensitivity: 96,8% Specificity: 96,6%
Yes	Still images	Sensitivity: 90,7% Specificity: 79,8%
Yes	Still images	Sensitivity: 94,0% Specificity: 95,4%
Yes	Videos	Sensitivity: 99,9% Specificity: 97,0% 6 minutes reading *

(NN) analysis of SB-VCE images with sensitivity of 96.6%, specificity of 99.9% and accuracy of 99.9% [5]. Although these results seem promising, they have all been obtained using still images, and video analyses are still lacking, both for detection and characterisation (fresh blood, clots or melena, traces or massive hemorrhage).

Angiectasias

Angiectasias or angiodysplasias are the most common injuries seen in SB-VCE. As proof of concept, our team used an AI approach based on NN in our study, reaching sensitivity of 100% and specificity of 96% for their detection from still images [6]. In 2020, Tsuboi et al. have developed an AI system based on a database of 189 videos of SB-VCE, with sensitivity of 98.8% and specificity of 98.4%. These promising results require validation with large prospective clinical trials [7].

Erosions and ulcers

The diversity of these lesions in terms of depth, size, shape and etiology constitutes a challenge for AI in SB-VCE. The AI solutions based on NN give excellent diagnostic performance after assessment on still images [8]. In a pilot series on videos, including 16 SB-VCE with 37 images with loss of substance and 4 normal videos, the time required to interpret videos by experts with/without AI decreased from 12 to 3 minutes while the detection rate was unchanged (87% and 84%, respectively) [9]. Taken together, improvement in the detection of ulcerated lesions is still needed for better assessement of videos.

Protruding lesions

The team of Saito et al. collected 30,584 images of polyps/masses/tumors from 292 patients, in order to develop an AI solution based on NN. The system was then tested on an independent batch of images, and showed a sensitivity of 90.7% and specificity of 79.8% [10]. While the proof of concept is imminent, there are currently no studies available on videos regarding the critical issue of the automated detection of protruding lesions. Furthermore, it will likely be difficult to characterise some of the protruding lesions that appear as variations of the norme (mucosal folds, phlebectasias, lymphangiectasia, chylous cysts, nodular lymphoid hyperplasia, Brunner's glands...) from the various diseases, benign or malignant, of epithelial or submucosal origins. .

Multiclass detection

In 2014, Iakovidis et al. proposed a method able to detect several types of lesions by classifying the images as abnormal. Instead of learning how to recognise different kinds of lesions, the method aims to distinguish an abnormal image from a normal one. In other words, "what is not normal is, by definition, abnormal" (eventhough a subsequent risk is to then not be able to correctly categorise the anomaly) (Figure 2) [11]. A few years later, Ding et al. managed to reach the performance of a gastroenterologist in terms of detection of lesions and variations of normal mucosa (lymphangiectasias, red dots ...) by training a DL model with more than 100 million images of SB-VCE collected from 77 centers in China [12]. The false positive rate was only 3%. Noteworthily, the average time spent on reading a video was 97 min with conventional analysis (surprisingly ...) but only 6 min with the aid of AI [12].

Characterisation

Only few studies in the literature have reported using AI solutions for lesion characterisation in SB-VCE. Our team has shown that, following the detection of angiectasias, their segmentation (the determination of the anatomical margins of the lesion) can be carried out by AI at the pixel scale, thus opening the way to the extraction of characteristics, such as the lesion size, type, depth [6]. The main limitation of this approach is that it relies on databases with annotations of very high quality and on heavy calculations. .

Combined Multiclass detection and characterisation

By combining a "weakly supervised" approach for detection with different sub-neural networks for classification, Otani et al. have recently reported a solution with detection performance of 0.928 for ulcerations, 0.884 for vascular abnormalities and 0.902 for protruding lesions, using still images [13].

Assessment of the quality of bowel preparation

While recommended, the evaluation of the quality of bowel preparation by gastroenterologists with VCE remains poorly reproducible [14]. Two AI algorithms have shown that AI can produce a cleanliness score that is robust and reliable [14, 15]. A recent study from our team provides a proof of concept that a DL approach has a sensitivity of about 90% for the ranking of videos compared to a consensus of experts, for the recognition of a badly prepared bowel (*figure 3*) [15].

Industrialisation

The first step forward was made in 2019 by the firm Ankon® (China), following the confirmation that an algorithm based on NN was able to detect abnormalities in SB-VCE with performance equivalent to gastroenterologists in a large multicenter study (cited earlier), while considerably saving time spent on analysis (to about 6 minutes) [12]. The company Jinshan (distributing the capsule Omom®) has also marketed its solution in Asia.

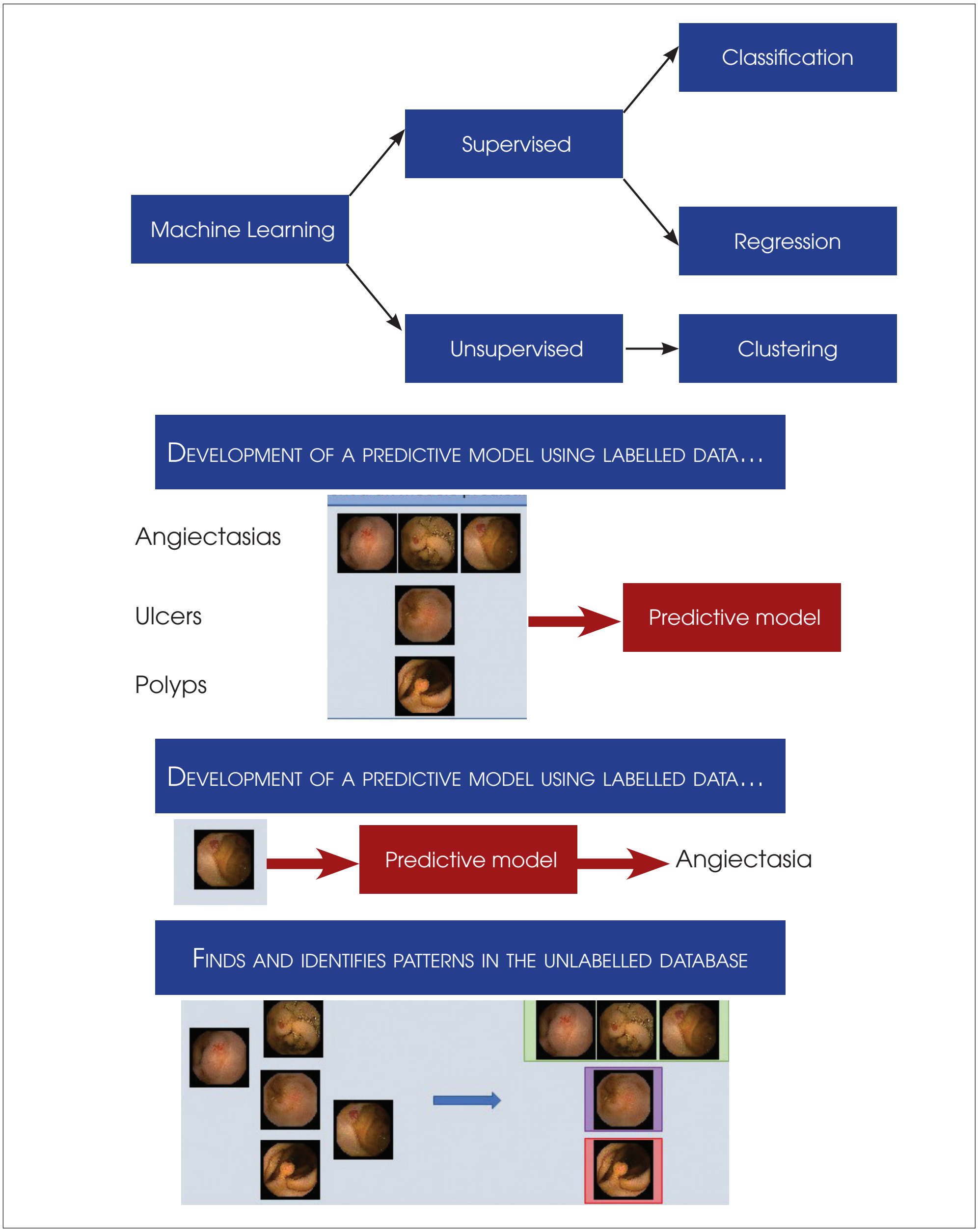


Figure 2. Principles of supervised and unsupervised machine learning.

Current limitations

Several important limitations of current studies of AI in SB-VCE should be recognised. Firstly, these studies are retrospective. In addition, most datasets used for training are based on selected images, leading to an inherent risk of overfitting. Studies with external validation are rare (or do not exist), most likely because the research teams that use these devices use different architectures of networks and images (different contrasts, resolutions and labels). Finally, some aspects are still barely addressed by AI researchers, such as the recognition of anatomical marks (pylorus, ileocecal valve) or the assessment of the size of the lesions.

Priority needs

Highly supervised NN should one day enable the current technologies to address the challenge of lesion classification in terms of diagnosis and etiology. However, this development is hampered by the poor diagnostic agreement between users of SB-VCE, around 60% [16], probably due to inter-individual variations in the vocabulary used, the characterisation of lesions and their interpretation according to the clinical context. In the absence of a consensus for the reproducible interpretation of SB-VCE, machine learning is illusory. A European group has addressed this challenge by proposing an expert consensus on the nomenclature, the description and the etiology of vascular, ulcerative and protruding lesions, depending on the indications of SB-VCE [17]. This a necessary step for the development of solid databases and the establishment of «ground truth» necessary for machine learning.

Once these principles have been acquired, databases of better quality will be constructed allowing the application of active learning. This will not only allow to prospectively enrich these databases, but also to point out the errors of AI.

Potential obstacles

A number of potential obstacles to the development of AI solutions that we have listed in the text can be briefly mentioned, keeping in mind the ethical, economic and regulatory aspects of AI applied to endoscopy. First of all, knowledge is still lacking regarding the criteria that are gained by AI to perform any given technical task (detection, characterisation...). This lack of knowledge will not facilitate AI use in Medicine in general, and endoscopy and SB-VCE in particular. The responsibility of interpreting SB-VCE will continue to rest on Medical doctors, eventhough they will only see a small proportion of the captured images. These images will be archived and therefore it will be possible to re-examine them by experts. Until health authorities establish re-

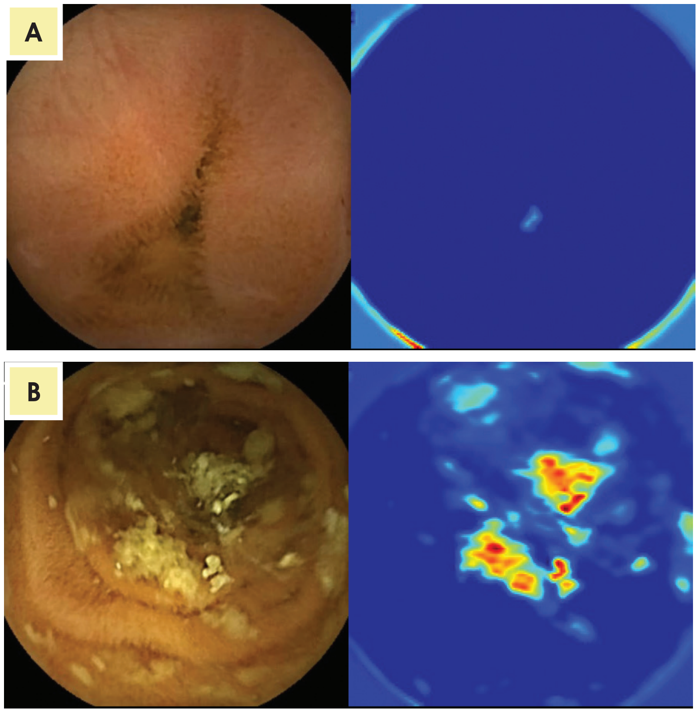


FIGURE 3. Representative images of endoscopic videocapsule and their use for analysing the quality of visualisation with the Axaro® AI system (Augmented Endoscopy, France): (a) adequate (b) inadequate. The original image is on the left and a heatmap is shown on the right, with visible areas in cold colours (blue) and poorly visible area in hot colours (red). The results are not given in the form of a heatmap (whose only interest here is to support the discussion) but as a proportion of dirty images out of the whole video for each quartile of the sequence. A percentage of lower than 21% is considered to reflect adequate preparation at the video scale [15].

gulatory principles for the use of AI solutions in Medicine (which are neither drugs nor devices), the doctor and the institution are in an unclear situation from a legal point of view, a situation that is likely to delay the adoption (and therefore the development and improvement) of these techniques. However, these improvements have far-reaching implications, considering the current situation of medical demography, particularly in our specialty, where medical time is priceless and should be spent on « noble » tasks. One can wonder how confident patients (and the society) will be (and therefore whether they will authorise) a system that does not include a safety net with proofreading, especially in its early years (and with its first errors). Our group has started performing audits on doctors and patients regarding these issues. These issues will open up essential ethical considerations. Finally, the transfer of major technological developments made in Asia is not necessarily possible immediately in Western countries. Indeed, it is accompanied with complex issues of digital and economic sovereignty.

Perspectives

The perspectives of AI development in SB-VCE include application of its use in the stomach, the colon, and panenteric systems (gastrointestinal or intestinal and colic capsules). If their interpretation becomes less cumbersome and more automated, it is likely that devices with multiple/panoramic heads will become the norm regarding the inspection of the small bowel. Finally, these developments are expected to boost the market of videoendoscopic capsules, with more prescribed and interpreted examinations. These examinations could eventually be performed at home, and one cannot exclude the possibility that in the future patients will download the videos taken of their digestive tract for online interpretation via secure platforms.

Conclusion

Twenty years after its revolutionary inauguration in gastrointestinal endoscopy, the VCE is coming to a golden age thanks to the development of AI. Solutions for automatic detection are now readily available on the market. Prospective controlled trials performed by independent teams are still necessary before medical doctors can entirely rely on these solutions and significantly reduce their interpretation time while still being sure of excellent detection performance. When this first critical step is reached, the automatic characterisation (in terms of diagnosis and relevance) will be the next step. The process promises however to take much longer, because the regulatory and ethical stakes are great, and in parallel VCE technology is following its own path, opening new fields of application (panenteric endoscopy, virtual chromoendoscopy, active locomotion, treatment options) in which AI will play a role.

TABLE 2. Main recent studies on artifical intelligence for the evaluation of bowel cleanliness during capsule endoscopy.

First author	Ground truth / Judgement Criterion	Device	Training	Test & Validation	Performance
Noorda R, 2020	Non-validated scale, with 2 to 4 points, evaluated by two independent experts (without adjudication)	Pillcam® SB3, Medtronic, USA	563 images (increased to 55,293) from 35 videos	854 images from 30 videos	Still images
Leenhardt R, 2020	Validated scale, with 2 to 10 points, evaluated by three independent experts (with adjudication)	Pillcam® SB3, Medtronic, USA	600 images (increased to 3,000) from 30 videos	156 videos	Still images

REFERENCES

1. Beg S, Card T, Sidhu R, Wronska E, Ragunath K. How many capsule endoscopy cases can be read before accuracy is affected? *Gut* 2018; 67 (S1) :A1–30.
2. Dray X, Iakovidis D, Houdeville C, Jover R, Diamantis D, Histace A, Koulaouzidis A. Artificial intelligence in small bowel capsule endoscopy: Current status, challenges, and future promise. *Journal of Gastroenterology and Hepatology* : 2021-01-15 , doi : 10.1111/jgh.15341
3. Yung DE, Sykes C, Koulaouzidis A. The validity of suspected blood indicator software in capsule endoscopy: a systematic review and meta-analysis. Expert Rev Gastroenterol. *Hepatol* 2017;11:43–51
4. Xing X, Jia X, Meng MQ. Bleeding Detection in Wireless Capsule Endoscopy Image Video Using Superpixel-Color Histogram and a Subspace KNN Classifier. *Annu Int Conf IEEE Eng Med Biol Soc* 2018 : 1-4.
5. Aoki T, Yamada A, Kato Y, et al. Automatic detection of blood content in capsule endoscopy images based on a deep convolutional neural network. *J Gastroenterol Hepatol* 2020 ; 35 : 1196-200
6. Leenhardt R, Vasseur P, Li C, et al. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. *Gastrointest Endosc* 2019 ; 89 : 189-94
7. Tsuboi A, Oka S, Aoyama K et al. Artificial intelligence using a convolutional neural network for automatic detection of small-bowel angioectasia in capsule endoscopy images. *Dig Endosc* 2020 ; 32 : 382-90.
8. Klang E, Barash Y, Margalit RY, et al. Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy. *Gastrointest Endosc* 2020 ; 91 :606-13
9. Aoki T, Yamada A, Aoyama K, et al. Clinical usefulness of a deep learning-based system as the first screening on small-bowel capsule endoscopy reading. *Dig Endosc* 2020;32 : 585-91
10. Saito H, Aoki T, Aoyama K, et al. Automatic detection and classification of protruding lesions in wireless capsule endoscopy images based on a deep convolutional neural network. *Gastrointest Endosc* 2020;92 : 144-51

11. Iakovidis DK, Koulaouzidis A Automatic lesion detection in capsule endoscopy based on color saliency: closer to an essential adjunct for reviewing software. *Gastrointest Endosc* 2014 ; 80 : 877-8

12. Ding Z, Shi H, Zhang H, Meng L, Fan M, Han C, et al. Gastroenterologist-Level Identification of Small-Bowel Diseases and Normal Variants by Capsule Endoscopy Using a Deep-Learning Model. *Gastroenterology* 2019 ; 157 : 1044-1054

13. Otani K, Nakada A, Kurose Y et al. Automatic detection of different types of small-bowel lesions on capsule endoscopy images using a newly developed deep convolutional neural network. *Endoscopy* 2020 ; 52 : 786-91. doi : 10.1055/a-1167-8157

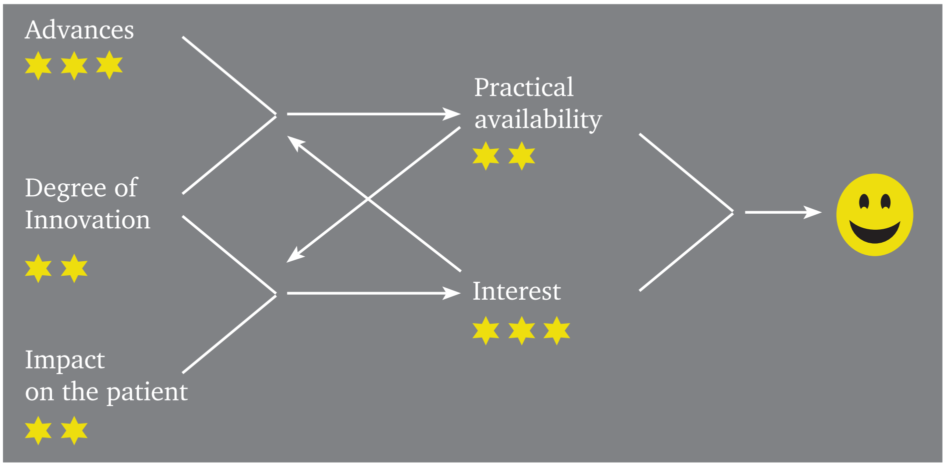
14. Noorda R, Nevárez A, Colomer A, Beltrán VP, Naranjo V. Automatic evaluation of degree of cleanliness in capsule endoscopy based on a novel CNN architecture. *Sci Rep* 2020 ; 10 : 17706

15. Leenhardt R, Souchaud M, Houist G, et al. A Neural Network-based Algorithm for Assessing the Cleanliness of Small Bowel during Capsule Endoscopy. *Endoscopy* 2020 ; epub. doi: 10.1055/a-1301-3841.

16. Jang BI, Lee SH, Moon JS, Cheung DY, Lee IS, Kim JO, et al. Inter-observer agreement on the interpretation of capsule endoscopy findings based on capsule endoscopy structured terminology: a multicenter study by the Korean Gut Image Study Group. *Scand J Gastroenterol* 2010;45(3):370-4

17. Leenhardt R, Koulaouzidis A, McNamara D. A guide for assessing the clinical relevance of findings in small bowel capsule endoscopy: Analysis of 8064 answers of international experts to an illustrated script questionnaire. *Clin Research Hepatol Gastroenterol* (sous presse).

AI and pathology of the lower digestive tract



This pictorial representation based on the author's assessment has but an indicative value

by Aymeric Becq
[1] *Center for digestive endoscopy, Saint-Antoine Hospital, APHP, Paris*
[2] *Institute of Intelligent Systems and Robotics, Paris Jussieu*

Colonoscopy is the procedure of reference for the detection of precancerous colic polyps [1]. This screening procedure is effective but not perfect. The adenoma detection rate (ADR) is highly variable (7-53%) and the rate of missed polyps is 22% [2,3]. This poses the problem of interval cancers and their mortality, a problem directly related to ADR [3]. Facing these limitations, a challenge is to reduce inter-operator variability and optimise the prevention of colorectal cancer by colonoscopy. In this context, Artificial Intelligence (AI) represents a possible solution. In fact, there are two major points for which AI may be helpful:

- Helping to detect polyps (Computer-Aided Detection, CADe). The objective is to increase the ADR;
- Helping to characterise polyps (Computer-Aided Diagnosis, CADx). The objective is to reduce the unnecessary resection of hyperplastic polyps (HP).

The European recommendations issued by the ESGE in 2019 carefully mention AI as a potential tool for detection and characterisation, provided that high precision and reproducibility can be demonstrated in good quality multicenter trials. The difficulties and the risks presented by a wide-scale implementation of colonoscopy are also mentioned: a decrease in the number of endoscopists, greater confidence in AI, poor representation of ground truth in the databases that are used to train AI algorithms, hacking [4] ...

Polyp Detection

Regarding polyp detection in the colon, AI must be sensitive, produce few false positives, and operate in real time. In order to limit the risk of not detecting a polyp, that is to say in order to be sensitive enough, the detection threshold used by the algorithm is decreased at the expense of specificity, with more potential false positives. This is desirable, as long as the gain in sensitivity does not greatly increase the number of false positives.

The use of deep learning approaches, such as Convolutional Neural Networks (CNN) became possible thanks to the increased power of processors that are required to analyse the large datasets necessary to set up AI algorithms. They are trained to identify polyps vs non-polyp lesions in a large database without researcher guidance.

Over the last few years, several AI algorithms were conceived, allowing the evaluation of images and videos, followed by real time videos. Among the existing systems, it is possible to mention the GI-Genius (Medtronic), Wision AI (Shanghai Wision AI Co., Ltd.) and CAD-EYE (Fujifilm). The use is intended to be simple and pragmatic. Following the «plug and play» principle, the device works directly once it is connected and switched on. Two signals warn the endoscopist when a polyp has been identified: a visual signal on the screen, and an (optional) sound signal. During a colonoscopy, 25 to 60 images/second scroll on the screen. Thus, in order to work in real time, AI must recognise and display the signal in a time frame of 1/25 to 1/60th of second (40 ms), requiring fast and powerful processors.

Table 1 summarises the main studies published on this subject to date. The first in vivo study, published in 2019, was a prospective study based on 55 colonoscopies and showed an ADR of 29.1% with AI vs 30.9% with an endoscopist. The ADRs were therefore considered comparable, suggesting that the AI algorithm could be improved, but also showing that real time detection is feasible [5].The first prospective randomised study, published in 2019 by Wang *et al.*, compared the performance of an AI system (Wision-AI) in 1,130 patients with an indication of colonoscopy. ADR was greater with

TABLE 1. An overview of the main studies evaluating AI for the detection of colonic polyps.

Study	Type of study	Number of images of polyps	Image mode	Real time video	Evaluation of diminutive polyps	Résults
Urban et al. (2018)	RS	4 088	WL	+	+	
Figueiredo et al (2018)	RS	1 680	WL	-	-	Sensitivity 99.7% Specificity 84.9% Accuracy 84.9%
Wang et al (2018)	ECR	6749	WL	+	+	9% increase in ADR
Misawa et al (2019)	RS	87	WL	-	+	Sensitivity 90% Specificity 63.3% Accuracy 76.5%
Matsui et al (2019)	MA	606	WL	+	+	Sensitivity 92.3% Specificity 89.8% Accuracy 96%
Lui et al (2019)	RS	12 654	NBI	+	+	Increase in ADR 39.2% vs 23.9%
Repici et al (2020)	ECR	X	WL	+	+	Increase in ADR 54.8% vs 40.4%

RS: retrospective study, RCT: randomised controlled trial, MA: meta-analysis, WL: white light, NBI: Narrow Band Imaging .

AI (29% vs 20%; $p < 0.001$). The rate was notably higher for diminutive polyps (70.6% vs 63.5%; $p<0.001$). The limitations of the study were the lack of generalisability (lack of selection of patients, an average age of 50 years, a study conducted in one country: China, where the prevalence of colorectal cancer is lower). The polyps that were most often detected by AI were diminutive and/or hyperplastic, raising questions regarding the utility of AI. The first study carried out with an AI system with a CE marking appeared in 2020. It was a randomised controlled multicenter clinical study that included 685 patients with a screening colonoscopy. The ADR was 54.8% with AI vs 40.4% for control (RR=1.3 [CI: 1.14-1.45]). AI enabled the detection of more adenomas by colonoscopy, and more adenomas of <5 mm and between 6 and 9 mm in size [6]. A meta-analysis published in 2020 brought together six studies on AI-assisted detection and showed a global sensitivity of 95.0% [CI95 : 91.0-97.0%] and specificity of 88.0% [CI95: 58.0-99.0 %] [7].

Altogether, in 2021, the problem of detection of polyps by AI can be considered solved. The next steps are, on one hand, the clinical validation with other randomised trials to examine the impact of AI on the ADR, and on the other hand, the administrative steps for the marketing of the different systems.

Characterisation of polyps

Characterisation of polyps is the second aim of AI, but it is different from the first and does not target the same endoscopists. Indeed, it will generate greater interest among expert endoscopists than its use for detection. The interest is also greater depending on the country and continent, Japanese being for example more interested in aspects related to characterisation.

The ASGE (American Society for Gastrointestinal Endoscopy) Technology Committee has issued and validated thresholds considered acceptable for the use of new technologies allowing real time histology of diminutive polyps during colonoscopy: the PIVI criteria (Preservation and Incorporation of Valuable endoscopic Innovation). Before a technology can be used as a guide in the decision of whether or not to resect rectosigmoid polyps that are suspected of being hyperplastic and <5 mm, the negative predictive value for adenomatous histology should be of at least 90% [8]. This threshold should be considered as a reference for the assessement of the performance of AI in predicting the histology of colonic polyps.

There are two types of systems. Systems that scan images obtained from classical colonoscopy, and systems that analyse zoomed images (endocytoscopy). The existing systems include the AI4GI, CAD-EYE (Fujifilm) and EndoBRAIN® (endocytoscopy) systems. *Table 2* summarises the main studies published on this subject to date.

A 2019 study published by Bryne et al. constitutes a reference regarding colonoscopy images. It is based on deep learning analysis of more than 60,000 images of normal mucosa/colonic polyps selected from 223 videos, using Narrow Band Imaging (NBI), in near vision and close/very close mucosa (Near focus technology, Olympus, Tokyo, Japan). The system displayed a level of confidence for the diagnosis of normal mucosa, hyperplastic polyps (NICE 1) or adenoma (NICE2). In an independent examination of 125 videos of polyps, lasting 10-20 seconds, the performance of the system was compared to that of pathological analysis, and was found to go beyond the PIVI criteria with a sensitivity of 98%, specificity of 83%, and diagnostic accuracy of 94% [9].

Endocytoscopy allows the in vivo examination of the morphology of the superficial cell layer of the tissue and their nuclei using high-magnification (X520) after methylene blue (1%) and crystal violet (0.05%) staining. The combination with NBI allows the added inspection of microvessels. A system called EndoBRAIN® has been evaluated in a pilot study published in 2015 on images of polyps of <10 mm, showing a sensitivity of 92.0%, specificity of 79.5% and accuracy of 89.2% for the differentiation between neoplastic and non-neoplastic polyps. The time required for analysis was 0.3 seconds and the performance was highly reproducible, similar to those by expert endoscopists and significantly better than endoscopists in training [10]. This system, based on deep learning of almost 70,000 images was further studied in a retrospective multicenter validation study that was published in 2020, before marketing in Japan [11]. The performance of AI was as follows. Images with staining: sensitivity 96.9%, specificity 100% and accuracy 98.0%. Images taken with NBI: sensitivity 96.9%, specificity 94.3% and accuracy 96.0%. The performance was superior to endoscopists in training, as well as expert endoscopists, regarding sensitivity. The negative predictive value for rectosigmoidal diminutive polyps was 100%, an interesting observation suggesting that the polyps suspected of being hyperplastic could be left untouched according to the PIVI thresholds.

This study however presented a number of limitatons, such as the toxicity of methylene blue, the staining duration at diagnosis (73 seconds). The authors stated that virtual colouring endocystoscopy was acceptable, but that methylene blue staining may help to predict the degree of submucosal invasion. This remains however to be evaluated [11].

Finally, the meta-analysis published in 2020 that we mentioned earlier also brought together 18 studies (three of which were prospective studies) on AI-assisted histologi-

Table 2. Main studies evaluating AI for the characterisation of colonic polyps.

Study	Type of study	Number of images of polyps	Type of image	Real time video	Evaluation of diminutive polyps	Results
Mori et al. (2015)	RS	176	Endocytoscopy	-	-	Sensibility 92 % Specificity 79,5 % Accuracy 89,2 %
Misawa et al. (2016)	RS	100	Endocytoscopy withNBI	-	-	Sensibility 84,5 % Specificity 97,6 % Accuracy 90,0 %
Kominami et al. (2016)	P	118	NBI + Zoom	+	+	Sensibility 95,9 % Specificity 93, 3 % Accuracy 94, 9%
Takeda et al. (2017)	RS	200	NBI + Zoom	-	-	Sensibility 89.4% Specificity 98.9% Accuracy 94.1%
Mori et al. (2018)	P	466	Endocytoscopy with NBI and staining	+	+	Prediction Rate 98,1 %
Chen et al. (2019)	RS	284	NBI + Zoom	-	+	Sensibility 96,3 % Specificity 78,1 % Accuracy 90,1 %
Repici et al (2020)	RS	125	NBI	+	+	Sensibility 98 % Specificity 83 % Accuracy 94 %

RS: retrospective P : prospective, NBI : Narrow Band Imaging.

cal prediction. Altogether, 7,680 images of polyps were examined and showed a total sensitivity of 92.3% [CI95: 88.8-94.9%] and specificity of 89.8% [CI95: 83.5 -93.0%] [7]. Interestingly, among these studies, five used deep learning algorithms, with a global sensitivity of 94.7% and specificity of 91.7%. Also, the global sensitivity for the characterisation of diminutive polyps was 93.5% and the specificity was 90.8%. Finally, AI was in general superior to non-expert endoscopists [7].

Combined tasks

Some AI softwares can perform multiple tasks such as detection and characterisations. This is for example the case with the CAD EYE (Fujifilm®) system. This system is based on deep learning, with an algorithm that processes 60 images per second during real time analysis. A CE marking has been obtained in 2020. Detection is performed with LCI (Linked Colour Imaging) chromoendoscopy or in white light. Characterisation is performed with BLI (Blue Light imaging) chromoendoscopy, without image freezing or using zoom. The algorithm does not show a high level of confidence or precision. For neoplastic lesions, a yellow circle is displayed (adenomas and cancers). For hyperplastic lesions, a green circle appears (hyperplastic polyps and sessile serrated adenomas/polyps). Validation studies are in progress.

Another interesting system is the ENDOANGEL system, developed by deep learning (neural networks). It aids in the detection and characterisation of polyps and performs quality monitoring in real time: recognition of the cecum, monitoring of withdrawal speed and assessment of the quality of preparation (Boston's score). In a randomised study that included 704 patients, the ADR was higher with AI than without (odds ratio=2.30 [CI95 1.40-3.77, $p<0.001$]). Withdrawal speed was significantly greater with AI (6.38+2.48 min vs 4.76+2.54 min; $p<0.001$) [12]. Regarding the evaluation of the cleanliness of the colon, another study based on image analysis showed an accuracy of 93.33%, higher than that of the endoscopist, with a precision of 80% on images with bubbles [13].

Conclusions

Today, AI is able to detect and characterise colonic polyps with high precision. The technology is ready: the equipment (computing power) and software (artificial intelligence, neural networks) performances made clear progress over the past years. Validated algorithms for detection and for characterisation are now available in Europe and in Japan. That being said, the majority of polyps detected are diminutive, raising issues regarding the real clinical interest, and experience is lacking. In addition, the algorithms suffer from a number of limitations: the majority of the published studies have been carried out on images (instead of real-time videos) and used a limited number of images of polyps for training and validation of the AI. In addition, human data were used to train AI, a situation that might have transferred some bias and could result in databases that are poorly representative of the general population. The resulting algorithm could then be biased. Similarly, the use of colonoscopies with an average ADR is unlikely to improve the performance of AI, and it is likely that a greater benefit will be gained from training driven by images of polyps missed by endoscopists.

Lastly, more studies are needed to assess the effect of AI on ADR and, in the long term, the incidence of colorectal cancer, and to evaluate the ability of AI to detect large, flat polyps. The answer is expected in the near future ...

REFERENCES

1. Burt RW, Cannon JA, David DS, *et al.* Colorectal cancer screening. *J Natl Compr Cancer Netw JNCCN* 2013 ; 11 : 1538-75. doi : 10.6004/jnccn.2013.0180.
2. Corley DA, Jensen CD, Marks AR, *et al.* Adenoma detection rate and risk of colorectal cancer and death. *N Engl J Med* 2014 ; 370 : 1298-306. doi : 10.1056/NEJMoa1309086.
3. van Rijn JC, Reitsma JB, Stoker J, Bossuyt PM, van Deventer SJ, Dekker E. Polyp miss rate determined by tandem colonoscopy: a systematic review. *Am J Gastroenterol* 2006 ; 101 : 343-50. doi : 10.1111/j.1572-0241.2006.00390.x.
4. Bisschops R, East JE, Hassan C, *et al.* Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2019. *Endoscopy* 2019 ; 51 : 1155-79. doi : 10.1055/a-1031-7657.

5. Klare P, Sander C, Prinzen M, *et al.* Automated polyp detection in the colorectum: a prospective study (with videos). *Gastrointest Endosc* 2019 ; 89: 576-82.e1. doi : 10.1016/j.gie.2018.09.042.

6. Repici A, Badalamenti M, Maselli R, *et al.* Efficacy of Real-Time Computer-Aided Detection of Colorectal Neoplasia in a Randomized Trial. *Gastroenterology* 2020 ; 159 : 512-20.e7. doi : 10.1053/j.gastro.2020.04.062.

7. Lui TKL, Guo C-G, Leung WK. Accuracy of artificial intelligence on histology prediction and detection of colorectal polyps: a systematic review and meta-analysis. *Gastrointest Endosc* 2020 ; 92 : 11-22.e6. doi : 10.1016/j.gie.2020.02.033.

8. ASGE Technology Committee, Abu Dayyeh BK, Thosani N, *et al.* ASGE Technology Committee systematic review and meta-analysis assessing the ASGE PIVI thresholds for adopting real-time endoscopic assessment of the histology of diminutive colorectal polyps. *Gastrointest Endosc* 2015;81:502.e1-502.e16. doi : 10.1016/j.gie.2014.12.022.

9. Byrne MF, Chapados N, Soudan F, *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* 2019;68:94-100. doi : 10.1136/gutjnl-2017-314547.

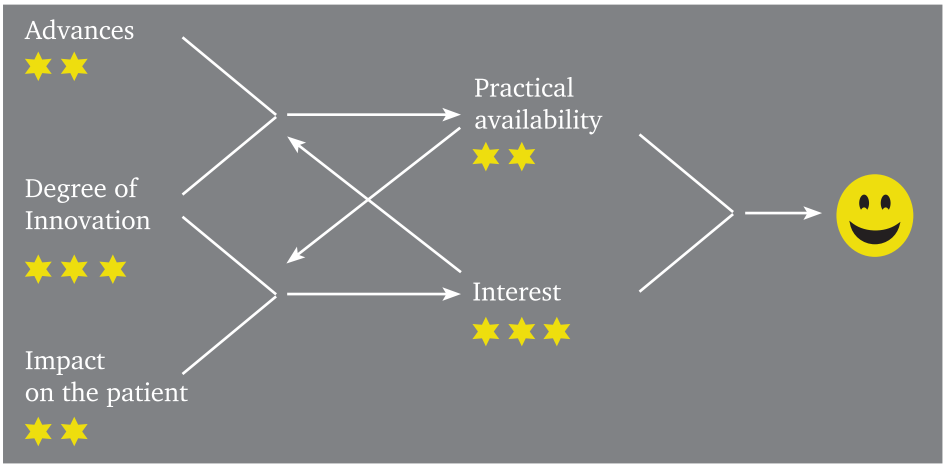
10. Mori Y, Kudo S-E, Wakamura K, *et al.* Novel computer-aided diagnostic system for colorectal lesions by using endocytoscopy (with videos). *Gastrointest Endosc* 2015 ; 81 : 621-9. doi : 10.1016/j.gie.2014.09.008.

11. Kudo S-E, Misawa M, Mori Y, *et al.* Artificial Intelligence-assisted System Improves Endoscopic Identification of Colorectal Neoplasms. *Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc* 2020 ; 18 : 1874-81.e2. doi : 10.1016/j.cgh.2019.09.009.

12. Gong D, Wu L, Zhang J, *et al.* Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. *Lancet Gastroenterol Hepatol* 2020 ; 5 : 352-61. doi : 10.1016/S2468-1253(19)30413-3.

13. Zhou J, Wu L, Wan X, *et al.* A novel artificial intelligence system for the assessment of bowel preparation (with video). *Gastrointest Endosc* 2020 ; 91 : 428-35.e2. doi : 10.1016/j.gie.2019.11.026.

AI and intestinal inflammatory bowel diseases (IBD)



This pictorial representation based on the author's assessment has but an indicative value

by Catherine Le Berre
Institute of Diseases of the Digestive System,
Nantes University Hospital - Hôtel Dieu, Nantes

Chronic intestinal inflammatory bowel diseases (IBD) are chronic diseases associated with altered regulation of immunity in the digestive tract. In the case of ulcerative colitis (UC), the rectum and colon are exclusively altered, while Crohn's disease (CD) can affect the entire digestive tract. There is currently no «gold standard» to diagnose IBD. Updated European recommendations issued in 2019 indicate that the confirmation of the diagnosis is based on the convergence of clinical, biological, endoscopic, histological, and even radiological signs [1]. Several medical examinations are therefore necessary for the initial diagnosis of IBD. Colonoscopy with multiple staged biopsies remains the examination performed on first intention. An upper gastrointestinal endoscopy is very often carried out at the same time.

Exploration of the small intestine by entero-MRI or videocapsule endoscopy (VCE) is recommended for patients for which there is a strong clinical suspicion of CD despite having normal results for conventional endoscopic explorations, and it should be systematically performed in patients that are newly diagnosed with CD [1].

However, a significant number of gastroenterologists (36.9%) have no easy access to VCE, as shown in a recent Spanish survey [2], and it is well known that reading an entero-MRI is radiologist-dependent with great heterogeneity among centers in the quality of interpretation [3]. Regarding conventional endoscopic examinations, a significant inter-observer variability was observed regarding the assessment of the severity of inflammation [4]. The medical management of IBD can therefore considerably vary from one patient to another.

In the last ten years, big data retrieved from large clinical trials and electronic health records, from biological collections, Omics databases (genomic, transcriptomic, metabolomic, proteomic) were increasingly used in order to identify patient profiles of clinical relevance. By combining multiple types of information, Artificial Intelligence (AI) and «Machine Learning» could increase the diagnostic performance and help to predict the evolution of IBD, whose complex pathophysiology involves immunity, microbiota, genetics and environmental factors. It might thus help in the standardisation of the medical management of patients with these diseases.

Applications of AI in the diagnosis and assessment of the severity of IBD

Diagnosis

As discussed above, assessing the small-intestine is often difficult in newly diagnosed patients or patients suspected of CD. Yet the damage of the proximal small bowel is associated with a significant risk of stenoses and subsequent multiple surgical interventions, with a potentially great impact on the prognosis and therapeutic management. The time required to analyse a VCE (25-60 minutes) is an important limitation to the broad performance of this examination, combined with the fact that it requires sustained attention for the whole time the images are displayed (15-20 images per second). Many teams have therefore developed programs for the automatic detection of bowel lesions of CD by VCE, most of which have reached diagnostic precision of above 90% [5-7].

With the emergence of «deep learning» and the powerful neural networks, pre-processing of data is almost not used anymore. Instead, most teams have applied weakly-supervised learning, where data are given to the algorithm without preprocessing but with an annotation regarding their pathological nature. The main interest of these neural networks is their simplicity of use, once they are developed, and their ability to self-improve. The team at the Nantes University Hospital, in collaboration with LS2N, has recently developed its own neural network trained with a personal database named Crohn-IPI, composed of 1,628 normal images and 1,590 images of CD lesions, achieving an accuracy of 89.3%, sensitivity of 87.8% and specificity of 90.7% [8].

Machine Learning techniques have also been used for the semiautomatic interpretation of entero-scanner or entero-MRI, based on parietal thickness, the dilation of the loops of the small intestine, the luminal diameter, and the existence of stenoses, and they have proven effective for the identification of intestinal lesions with an accuracy comparable to that of expert radiologists [9,10].

Endoscopic and morphological examinations sometimes fail to distinguish between the two IBD phenotypes. It is however essential to accurately diagnose CD from UC, since this is of great importance for the choice of medical treatment, and especially surgical treatment, in particular when a total colectomy is required with an ileo-anal anastomosis. The genetic characteristics of patients could aid in the phenotypic distinction. Currently, Genome-wide Assocation studies (GWAS) have identified several hundred loci associated with the risk of developing either CD or UC [11]. A recent technique of prioritisation based on machine learning recently identified 67 further gene candidates from a list of 16,390 genes [12]. In 2018, an american team has developed an algorithm called probabilistic pathway score that is no longer solely based on known genes but also on gene interactions at the individual level, in order to efficiently distinguish patients with CD from those with UC [13].

Severity assessment

Considering the relative subjectivity of endoscopic severity scores in UC (Mayo endoscopic sub-score, UCEIS), many teams have worked on developing algorithms based on deep learning in order to achieve an automatic stratification of the endoscopic lesions in these patients.

For example, a Japanese team has used a neural network to identify patients in endoscopic remission, as defined by a Mayo score of 0 or 0-1 with an area under the curve (AUROC) of 0.86 and 0.98, respectively, with a better performance in the rectum than in the rest of the colon [14]. An American team has also developed its own neural network enabling to distinguish the Mayo 0-1 patients (endoscopic remission) from Mayo 2-3 patients (moderate to severe activity) with an AUROC of 0.966 [15]. Beyond distinguishing patients with remission from those with active disease, two teams have each developed an algorithm that defines the exact Mayo endoscopic subscore, with concordance between automated interpretation and human analysis of around 70% in both cases [16,17].

Eventhough it is still a matter of debate, histological healing is considered to be an important goal of medical treatment in patients with UC, necessitating the performance of biopsies during control endoscopy. Several teams have used deep learning in order to skip this step of pathological analysis based solely on endoscopy. The neural network that has been developed by Takenaka et al. identified in time real patients in endoscopic remission (defined by a UCEIS score of 0) with a precision of 90.1%, and those in histological remission (defined by a Geboes score of <3) with an accuracy of 92.9% [18]. A Belgian team developed an algorithm that defines a « red density » score from endoscopic images of patients with minimally-active to moderately-active UC, a score that turned out to be significantly correlated with the Robarts histology index (r=0.74), the Mayo endoscopic subscore (r=0.76) and the UCEIS score (r=0.74) [19]. Another Japanese team has used endocytoscopy to develop a real time system helping to diagnose persistant histological inflammation (defined by a Geboes score higher than or equal to 3.1) with a sensitivity of 74%, specificity of 97%, accuracy of 91%, and above all perfect reproducibility (100%) [20].

All the studies that we previously cited have described neural networks that were trained with still images captured during endoscopy, but an American team has recently managed to develop an algorithm trained on whole videos obtained during endoscopy performed in a Phase 2 trial evaluating mirikizumab, an anti-IL-23, in patients with moderate to severe UC, enabling the calculation of the Mayo endoscopic subscore with concordance of 0.844 and UCEIS score of 0.855 [21].

Studies addressing the automatic evaluation of lesion severity in patients with CD are more rare. The team of Kopylov has recently trained a neural network with more than 17,640 images of VCE taken from 49 patients with CD; 7,391 images showed at least one mucosal ulceration and 10,249 images were normal. The ulcers were

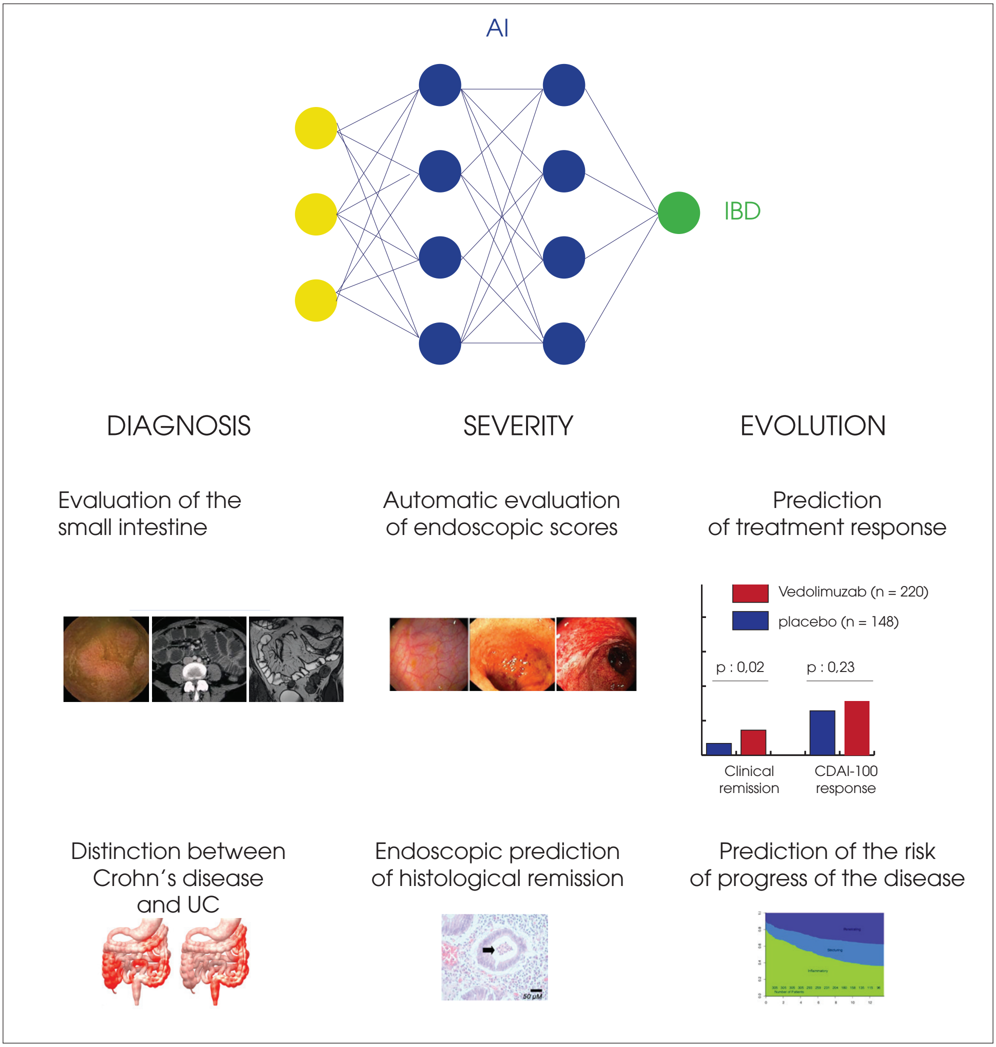


Figure 1. The utility of Artificial Intelligence in the management of IBD.

classified into three grades, from minimal to severe, by two experts, with an overall inter-observer agreement of only 31%, that still reached 76% for the distinction of grades 1 and 3. The neural network was able to distinguish grade 1 ulcers from grade 3 ulcers with an accuracy of 0.91, grade 2 ulcers from grade 3 ulcers with an accuracy of 0.78, and grade 1 ulcers from grade 2 ulcers with an accuracy of 0.62, i.e. opening up perspectives for its use in the diagnosis and monitoring of patients with CD of the small intestine [22].

Applications of AI for the prediction of the response to treatment and relapse

Over the past decades, a better understanding of the pathophysiology of IBD has led to great advances, with more therapies available and revised therapeutic objectives. Indeed, the advent of biological therapies (anti-TNF, anti-IL12/IL23, anti-integrin) and small molecules (JAK inhibitors) has revolutionised the management of these diseases. The overall response to each of these treatments is however less than 50%, and it is still impossible to predict which IBD patients will respond to these treatments.

AI-based systems have therefore been developed in order to predict the responses to each treatment available against IBD. Waljee et al. have for example developed an algorithm that uses the age and certain common biological parameters to distinguish responders from non-responders to a thiopurine treatment with an AUROCs of 0.856, versus 0.594 for 6-TGN concentration determination [23]. The same team used data from the GEMINI clinical trial, evaluating the efficacy of Vedolizumab in patients with UC and CD, to develop a machine learning-based model that could predict in the 6th week which patients would stay in clinical remission without corticosteroids in week 52, with an AUROCs of 0.73 for UC, and 0.75 for CD [24,25], allowing a selection of the patients in which Vedolizumab should be maintained in the absence of apparent benefits at week 6. More recently, a multi-omics analysis was conducted to identify transcriptomic biomarkers and genomic predictors of endoscopic response to ustekinumab treatment with diagnostic accuracy reaching 98% [26]. In the framework of acute severe colitis, predicting the response to treatment is even more important, considering that approximately 15% of patients still require a colectomy in 2020... Inefficient medical treatment may delay the surgical treatment, raising morbidity and mortality in this context. A French study developed an algorithm based on microRNA expression profiles determined from colonic biopsies obtained from UC patients presenting as severe acute colitis, enabling the identification of corticosteroid responders with an accuracy of 93%, infliximab responders with an accuracy of 84% and ciclosporin responders with an accuracy of 80% [27]. More recently, an Indian team used neural network based on simple clinical and biological parameters, such as the blood levels of hemoglobin and serum albumin, to efficiently predict the response of severe acute colitis to drug therapy [28].

AI can also be used to predict the course of IBD. A neural network analysing biopsies obtained from CD patients at an early stage was able to identify those at risk of stenotic and/or fistulising progression and those at risk of surgery with an accuracy of over 80% [29]. The team of Waljee developed a different model based on the analysis of electronic medical records to predict hospitalisations and the need to use corticosteroids in the following 6 months with AUROCs of 0.87 [30].

Conclusion and perspectives

AI is a growing discipline with the potential to revolutionise clinical decision-making for IBD. Machine Learning offers the possibility of aggregating large amounts of data to improve our performance in the diagnosis of IBD, to homogenise our practice through the automated evaluation of endoscopic severity of injuries, and to predict the response to treatment and the evolution of the disease for more personalised Medicine. AI should also improve the endoscopic detection of non-conventional dysplastic lesions that are specific to IBD and that are currently difficult to detect by standard endoscopy. Some obstacles remain before AI can be implemented in common practice, but deep learning will for sure be our companion in the near future for daily clinical practice and the management of patients with IBD.

REFERENCES

1. Maaser C, Sturm A, Vavricka SR, *et al.* ECCO-ESGAR Guideline for Diagnostic Assessment in IBD Part 1: Initial diagnosis, monitoring of known IBD, detection of complications. *J Crohns Colitis* 2019 ; 13 : 144-64K.
2. Marín Jiménez I. Survey on the use of capsule endoscopy for patients with known or suspected Crohn's disease in Spain. *Rev Esp Enferm Dig* 2015 ; 107. http://online.reed.es/fichaArticulo.aspx?iarf=221687760-749233419277.
3. Taylor SA, Mallett S, Bhatnagar G, *et al.* Magnetic resonance enterography compared with ultrasonography in newly diagnosed and relapsing Crohn's disease patients: the METRIC diagnostic accuracy study. *Health Technol Assess* 2019 ; 23 : 1–162.
4. Daperno M, Comberlato M, Bossa F,*et al*. Inter-observer agreement in endoscopic scoring systems: preliminary report of an ongoing study from the Italian Group for Inflammatory Bowel Disease (IG-IBD). *Dig Liver Dis* 2014 ; 46 : 969–973.
5. Girgis HZ, Mitchell BR, Dassopoulos T, *et al.* An intelligent system to detect Crohn's disease inflammation in Wireless Capsule Endoscopy videos 2010; 1373–76.
6. Kumar R, Qian Zhao, Seshamani S, *et al.* Assessment of Crohn's Disease Lesions in Wireless Capsule Endoscopy Images. *IEEE Transactions on Biomedical Engineering* 2012 ; 59 : 355–62.
7. Klang E, Barash Y, Margalit RY,*et al.* Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy. *Gastrointest Endosc* 2020 ; 91 : 606-13.e2.
8. Vallée R, Coutrot A, Normand N, *et al*. *Accurate small bowel lesions detection in wireless capsule endoscopy images using deep recurrent attention neural network. IEEE 21st International Workshop on Multimedia Signal Processing (MMSP 2019)*. Kuala Lumpur (Malaisie), 2019. https://hal.archives-ouvertes.fr/hal-02296282
9. Mahapatra D, Vos FM, Buhmann JM. Active learning based segmentation of Crohns disease from abdominal MRI. *Comput Methods Programs Biomed*. 2016;128:75–85.
10. Stidham RW, Enchakalody B, Waljee AK, *et al.* Assessing Small Bowel Stricturing and Morphology in Crohn's Disease Using Semi-automated Image Analysis. *Inflamm Bowel Dis* 2020 ; 26 : 734–42.

11. Cleynen I, Boucher G, Jostins L, *et al.* Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *Lancet*. 2016 ; 387 : 156–67.

12. Isakov O, Dotan I, Ben-Shachar S. Machine Learning-Based Gene Prioritization Identifies Novel Candidate Risk Genes for Inflammatory Bowel Disease. *Inflamm Bowel Dis* 2017 ; 23:1516–23.

13. Han L, Maciejewski M, Brockel C, *et al.* A probabilistic pathway score (PROPS) for classification with applications to inflammatory bowel disease. *Bioinformatics* 2018 ; 34 : 985–93.

14. Ozawa T, Ishihara S, Fujishiro M, *et al.* Novel computer-assisted diagnosis system for endoscopic disease activity in patients with ulcerative colitis. *Gastrointest Endosc* 2019 ; 89 : 416-21.e1.

15. Stidham RW, Liu W, Bishu S, *et al*. Performance of a Deep Learning Model vs Human Reviewers in Grading Endoscopic Disease Severity of Patients With Ulcerative Colitis. *JAMA Netw Open* 2019 ; 2 : e193963.

16. Yao H, Najarian K, Gryak J, *et al.* Fully automated endoscopic disease activity assessment in ulcerative colitis. *Gastrointest Endosc* 2020; epub. doi: 10.1016/j.gie.2020.08.011

17. Bhambhvani HP, Zamora A. Deep learning enabled classification of Mayo endoscopic subscore in patients with ulcerative colitis. *Eur J Gastroenterol Hepato*l 2020; epub. doi: 10.1097/MEG.0000000000001952

18. Takenaka K, Ohtsuka K, Fujii T, *et al.* Development and Validation of a Deep Neural Network for Accurate Evaluation of Endoscopic Images From Patients With Ulcerative Colitis. *Gastroenterology* 2020 ; 158 : 2150–57.

19. Bossuyt P, Nakase H, Vermeire S, *et al.* Automatic, computer-aided determination of endoscopic and histological inflammation in patients with mild to moderate ulcerative colitis based on red density. *Gut* 2020 ; 69 : 1778–86.

20. Maeda Y, Kudo S-E, Mori Y, *et al.* Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video). *Gastrointest Endosc*. 2019 ; 89 : 408–415.

21. Gottlieb K, Requa J, Karnes W, *et al.* Central Reading of Ulcerative Colitis Clinical Trial Videos Using Neural Networks. *Gastroenterology* 2020; epub.

22. Barash Y, Azaria L, Soffer S, *et al.* Ulcer severity grading in video capsule images of patients with Crohn's disease: an ordinal neural network solution. *Gastrointest Endosc* 2020; epub. doi : 10.1016/j.gie.2020.05.066.

23. Waljee AK, Joyce JC, Wang S, *et al.* Algorithms outperform metabolite tests in predicting response of patients with inflammatory bowel disease to thiopurines. *Clin Gastroenterol Hepatol* 2010 ; 8 : 143–50.

24. Waljee AK, Liu B, Sauder K, *et al.* Predicting corticosteroid-free endoscopic remission with vedolizumab in ulcerative colitis. *Aliment Pharmacol Ther* 2018 ; 47 : 763–72.

25. Waljee AK, Liu B, Sauder K, *et al.* Predicting Corticosteroid-Free Biologic Remission with Vedolizumab in Crohn's Disease. *Inflamm Bowel Dis* 2018 ; 24: 1185–92.

26. Verstockt B, Sudahakar P, Creyns B, *et al.* DOP70 An integrated multi-omics biomarker predicting endoscopic response in ustekinumab treated patients with Crohn's disease. *Journal of Crohn's and Colitis* 2019 ; 13 : S072-73.

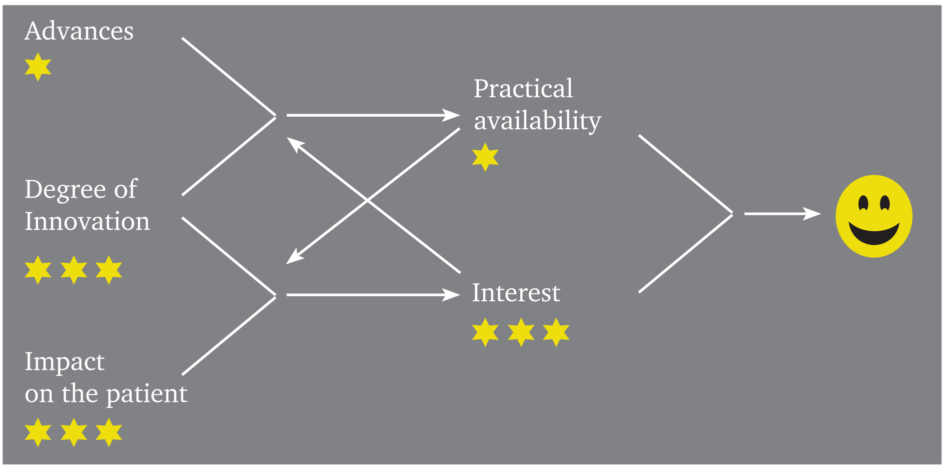
27. Morilla I, Uzzan M, Laharie D, *et al.* Colonic MicroRNA Profiles, Identified by a Deep Learning Algorithm, That Predict Responses to Therapy of Patients With Acute Severe Ulcerative Colitis. *Clin Gastroenterol Hepatol* 2019 ; 17 : 905–13.

28. Ghoshal UC, Rai S, Kulkarni A, *et al.* Prediction of outcome of treatment of acute severe ulcerative colitis using principal component analysis and artificial intelligence. *JGH Open* 2020 ; 4 : 889–97.

29. Klein A, Mazor Y, Karban A, *et al.* Early histological findings may predict the clinical phenotype in Crohn's colitis. *United European Gastroenterol J* 2017; 5: 694–701.

30. Waljee AK, Lipson R, Wiitala WL, *et al.* Predicting Hospitalization and Outpatient Corticosteroid Use in Inflammatory Bowel Disease Patients Using Machine Learning. *Inflamm Bowel Dis*. 2017;24:45–53.

AI and pancreas



This pictorial representation based on the author's assessment has but an indicative value

by Geoffreoy Vanbiervliet [1]
[1]*CHu de Nice*

Medical applications of artificial intelligence (AI) are raising considerable interest. This is relatively easy to understand, considering that AI is an easy (contrary to what is usually thought) and effective means of achieving greater precision, earlier and more reproducible diagnoses and saving time or rather accord more time to each patient. Pancreatic diseases, because of their dire prognosis (such as for pancreatic ductular adenocarcinoma), the therapeutic and surgical consequences they imply, require a diagnostic agility that is still insufficient today. AI in pancreatic diseases is still in its infancy, but one can guess that it will bring so much dexterity to the practitioner in the coming years that it will make it an irreplaceable tool. The aim of this chapter is to present the first results of AI in pancreatology, with a brief mention of the use of echo-endoscopy, which is essential in the field.

Deep learning and pancreatic imaging

The assessment of AI for pancreatic imaging is still in its infancy. The use of deep learning for the analysis of CT images of the pancreas has only recently been described. A study from Taiwan reported a diagnostic precision of AI of 0.989 with an almost perfect area under the curve (ROC) of 0.999 (95% CI: 0.998 to 1.000) for the diagnosis of pancreatic adenocarcinoma [1]. The system was also significantly more sensitive than the radiologist (0.983 vs 0.929, i.e. a difference of p=0.054 [95% CI: 0.011 to 0.098]; p=0.014). The same is true with MRI for which preliminary studies report for the moment lower area under the curve, of between 0.7189 for inflammatory pancreatic disease and 0.93019 for the diagnosis of neuro-endocrine tumors [2]. Additionally, in this multicenter retrospective study, the correlation with the human reader was excellent (□ coefficient of 0.8862 (CI 95%: 0.7759 to 9738)). To date, no study has been published on the testing of AI for percutaneous ultrasound of the pancreas. In summary, the literature regarding non-invasive pancreatic cross-sectional imaging is promising but preliminary.

AI and endoscopic ultrasound imaging of the pancreas

Diagnosis of adenocarcinoma

One of the problems in pancreatology is the inability to distinguish tumor tissue from inflammatory or even normal tissue (Table 1). Despite the modern imaging techniques, an inappropriate diagnosis of cancer still accounts for almost 7% of cephalic duodeno-pancreatectomies [3]. Surprisingly, the first studies proposing AI as a tool to highlight the components and characteristics of pancreatic tissue can be tracked back to the year 2000. Three studies between 2000 and 2008 have been published that offered only weak level of evidence, but paved the way for the use of AI to aid in the early diagnosis of pancreatic ductular adenocarcinoma.

The seminal publication of 2001 by a team from the Mayo Clinic demonstrated that a computer software can digitize pixel columns with varying levels of gray on an image of a pancreatic mass scanned with a mechanical radial echo-endoscopy probe [4].

Histological analysis was available for each lesion analysed. While the software only interpreted a single image generated by a probe that is today considered obsolete, without any additional clinical or iconographical support, it already had a performance comparable to that of a human expert, perfectly sensitive (100%) but still not specific (50%).

The use of neural networks was reported in 2008 for the analysis of pancreatic images obtained using endoscopic ultrasound (EUS). After training the system with images of normal tissue areas, with typical signs of chronic pancreatitis (lobularity, calcifications and hyperechoic appearance) and adenocarcinoma, Das et al. [5] showed that the software obtained an excellent area under the curve of 0.93 (AUROC) in terms of diagnostic precision in a validation cohort.

The same mathematical techniques based on neural networks have been used for the analysis of elastography histograms and contrast-enhanced harmonic EUS (SonoVue®) in the context of exploration of syndromes of pancreatic mass by EUS [6-7]. With these two methods widely used in current medical practice, they led to an improvement in the area under the curve in diagnostic performance: the software increased the ability of tumor elastography to distinguish benign vs malignant lesions from 0.847 (95% CI : 0.721 to 0.972) to 0.957, and pseudo-tumoral pancreatitis vs adenocarcinoma to 0.965 [6]. A gain of sensitivity was also noted when the software was coupled to quantitative evaluation of the contrast-enhancement (84.8% for fine needle puncture vs 87.5% for contrast-enhanced EUS vs 94.6 % for contrast-enhanced EUS coupled to AI) [7]. The common limitations of these initial studies were the small size of the validation cohorts, an internal cross-validation, a lack of centralised reinterpretation or comparison with the performance of human experts, and the absence of histological proof for all tissues analysed, especially in cases of chronic pancreatitis.

More recently, three studies gave a solid ground to the concept of Endoscopic Ultrasound computer-assisted diagnosis (EUS-CAD). Using either neural networks of support vector machines, the texture of echoendoscopic images was analysed as a way to perform a differential diagnosis of pancreatic cancer vs chronic pancreatitis [8-10]. The diagnostic precision of these mathematical models was high, ranging from 87 to 98%, eventhough only one internal cross-validation method was adopted.

Finally, «Deep Learning» was only applied in pancreatic endoscopy in 2020 in a preliminary prospective Japanese study on 139 patients (76 with adenocarcinoma, 34 with chronic pancreatitis and 29 with normal pancreas) [11]. 920 images were used to train and validate a convolutional neural network, of which only 470 were used as a testing cohort (47 cases separately and independently tested). The system performance was compared either to histology (in case of adenocarcinoma) or echo-endoscopic diagnosis (Rosemont Classification for EUS evaluation in case of chronic pancreatitis). The software reached a sensitivity of 92.4%, specificity of 84.1%, positive predictive value of 86.8% and negative predictive value of 90.7%. These results are obviously excellent and virtually identical to the performance achieved with the best fine needles used for pancreatic biopsy under EUS.

AI seems highly comparable and perhaps even better than human experts in the field. A recent retrospective study on a test cohort of 123 patients firstly confirmed that the inter-observer correlation was moderate (kappa coefficient of 0.458) among experts (seven in this study). It then showed that, compared to human experts, the machine had significantly higher diagnostic performance for the main pancreatic diseases, including adenocarcinoma (75.6% vs 61.6%; p=0.026) [12].

Diagnosis of pancreatic cystic lesions

Optimising the diagnosis of pancreatic mucinous cystic lesions and the estimation of their malignant potential is the holy grail of EUS and it has inspired multiple attempts using EUS-CAD. Only one retrospective study has been published to date, but it is guiding the way [13]. This study was based on a cohort of 206 patients with a diagnosis of intra-pancreatic mucinous papillary tumor (IPMPT), established by pathological examination of the resected specimen. Among these patients, a test cohort of 50 patients (including 23 with cancer) and 3,970 endoscopic ultrasound images of IPMPT was analysed. The area under the ROC regarding AI performance for the diagnosis IPMPT was 0.98, significantly higher than in the presence of a nodule (0.74, p=0.001). The diagnostic performance of the machine (94%) was not only superior to that of humans (preoperative diagnosis made by the operator) (56%), but also to that of the selected indications of surgery, relative (40%) and absolute (68 %), as defined by international recommendations.

Diagnosis of Autoimmune pancreatitis

Autoimmune pancreatitis (AIP) of type I is typically found in the context of jaundice and pseudo-obstructive syndrome in people aged >60 years, making the diagnosis difficult. In this context, the use of AI to distinguish between inflammatory tissue lesions and adenocarcinoma makes sense. The team of Michael Levy at the Mayo clinic has recently reported its retrospective experience based on 4,945 images and 1,852 EUS videos obtained from 583 patients who presented AIP, chronic pancreatitis, adenocarcinomas as well as normal pancreas [12].

The convolutional neural network was built with a well-known technique and trained and validated with randomly-assigned cases (460 patients, of which 118 (25.7%) had AIP) and a test cohort (123 patients including 28 (22.8%) with AIP). The performance speed of the system was remarkable since it reached 955 images per second, well above the 30 required to perform real time analysis of EUS with its current standards.

The diagnostic performance obtained was excellent, especially when only the videos were analysed. The sensitivity and specificity of the system for its ability to distinguish

TABLE 1. Diagnostic performance of AI and echo-endoscopic diagnosis of pancreatic adenocarcinomas.

AUTHOR, YEAR	N	TECHNIQUE
Norton ID, 2001	35 (14 PC et 21 AC)	Mechanical radial echo-endoscopy
Das A, 2008	56 (22 PN, 12 PC et 22 AC)	Mechanical radial echo-endoscopy
Saftoiu A, 2008	68 (22 PN, 11 PC, 32 AC et 3 TNE)	Linear electronic endoscopy and elastography (histogram)
Zhang MM, 2010	216 (153 AC, 43 PC, 20 PN)	Écho-endoscopie linéaire électronique
Zhu M, 2013	388 (262 AC et 126 PC)	Linear electronic endoscopy
Saftoiu A, 2015	167 (112 AC et 55 PC)	Linear electronic endoscopy and elastography and contrast enhancement (SonoVue®) (Time-intensity curve)
Ozkan M, 2016	323 (202 AC et 130 PN)	Linear electronic endoscopy
Tonozuka R, 2020	139 (76 AC, 34 PC, et 29 pancréas normaux)	Linear electronic endoscopy
Marya NB, 2020	123 (60 AC, 28 AIP, 16 PC et 19 pancréas normaux)	Linear electronic endoscopy

Diagnostic accuracy (AUROC)	Sensitivity (%)	Specificity (%)
80	100	50
0,93	93 (IC 95% (89-97))	92 (IC 95% (88-96))
0,965	-	-
0,98	94,3	99,4
0,94	96,2	93,4
	94,6 (IC 95 % (88,2-97,8))	94,4 (IC 95% (83,9-98,6))
0,875	83,3	93,3
0,94	92,4	84,1
0.976	95 (IC 95% (91-98))	Images fixes

AIP from chronic pancreatitis were 94% and 71 %, respectively with an area under the curve (ROC) equal to 0.892 (CI 95%: 0.829 to 0.946); for the ability to distinguish AIP from adenocarcinoma, 90% and 93%, respectively with an area under the curve (ROC) of 0.963 (CI 95%: 0.941 to 0.981). Finally, the ability of the system to distinguish AIP from adenocarcinoma was compared with the performance of human experts. Specificity was nearly identical, with 88.2% (CI 95%: 72.6% to 96.7%) for AI vs 82.4% (76.9% to 87.0%) for human experts, but the machine performed better in terms of sensitivity (88.2% (CI 95%: 63.6% to 98.5%) vs 53.8% (CI 95%: 44.4% to 62.9%).

Training of technical aspects of pancreatic endoscopy

Training and learning recently became a field of application in AI. We know that training in the field of EUS takes a long time and that it is hard to master, making its end performance operator-dependent in clinical practice. A Chinese team recently proposed an AI-based platform for the training and quality control of AI [14]. Six pancreatic areas were identified and anotated beforehand in 19,486 images. Internal and external validation was performed and test stations were built based on 396 video clips. A comparison between the prediction model and EUS experts was made, and the learning speed was analysed in a cross over study with eight students.

The results that were obtained were excellent. During the video test, the model achieved a precision of 86.2% for the classification of the areas, with an excellent correlation coefficient between experts and the machine of more than 0.8. Regarding the recognition of pancreatic areas, the system increased the learning speed of students, with an increase in the diagnostic precision from 0.672 to 0.784 (i.e. a difference of 0.112 (CI 95%: 0.058 to 1.663), $p=0.002$).

In a world where one should never practice for the first time on a human subject, more codified and democratic use of medical simulation for the training of students is becoming the rule. We argue that such models will be advantageous in the assistance during the first steps in humans. Accompaniment however remains crucial in a discipline where the eye and verbalised human experience have demonstrated their full value.

Future, hope and limitations

This chapter aims to simply and clearly present the immense potential of AI in the field of pancreatology and in particular its exploration by endoscopic ultrasound. Because it is fast and easy to use, we expect that, in the close future, its diagnostic performance will increase, with an already excellent sensitivity that ranks this technology at the same level or even above those achieved by ultrasound-guided needle biopsy. Who would not want to avoid the unnecessary comorbidity associated with an unnecessary puncture? Who would not want to avoid the inevitable variability of interpretation associated with the sometimes difficult to analyse samples? How could one refuse not to be tributary of a technique that necessitates, no matter how much we value it, a high degree of expertise, echo-anatomical, scientific and endoscopic knowledge.

However, several limitations still stand in the way: in order to supply an AI system, it is necessary to have large databases with many annotated images, something that is not easy in EUS. The delay in this field in 2021, in France and elsewhere, is partially explained by complex legal and ethical aspects. This obstacle could be lifted with the organisation of anonymous databases containing encrypted basic medical data, providing a resource for research and development, as is for example done in the North American model of the WestHealth.org institute. Finally, we should get out of the outdated methodological debate based on the dichotomy between retrospective/ insufficient vs prospective/ideal. The quality of studies should inevitably increase. A remaining need will be to establish a gold standard for quality in histology, a criticism regularly made about studies that evaluate the diagnostic exploration of the pancreas by EUS.

To address these three issues, it will be necessay to perform national, as well as international, multicenter studies before the technology can finally be included in our diagnostic arsenal.

REFERENCES

1. Liu K-L, Wu T, Chen P-T, et al. Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation. Lancet Digit Health 2020; 2: e303–e313. doi:10.1016/S2589-7500(20)30078-9
2. Gao X, Wang X. Performance of deep learning for differentiating pancreatic diseases on contrast-enhanced magnetic resonance imaging: A preliminary study. Diagn Interv Imaging 2020; 101: 91–100. doi:10.1016/j.diii.2019.07.002
3. Gerritsen A, Molenaar IQ, Bollen TL, et al. Preoperative characteristics of patients with presumed pancreatic cancer but ultimately benign disease: a multicenter series of 344 pancreatoduodenectomies. Ann Surg Oncol 2014; 21: 3999–4006. doi:10.1245/s10434-014-3810-7
4. Norton ID, Zheng Y, Wiersema MS, et al. Neural network analysis of EUS images to differentiate between pancreatic malignancy and pancreatitis. Gastrointest Endosc 2001; 54: 625–629. doi:10.1067/mge.2001.118644
5. Das A, Nguyen CC, Li F, et al. Digital image analysis of EUS images accurately differentiates pancreatic cancer from chronic pancreatitis and normal tissue. Gastrointest Endosc 2008; 67: 861–867. doi:10.1016/j.gie.2007.08.036
6. Sâftoiu A, Vilmann P, Gorunescu F, et al. Neural network analysis of dynamic sequences of EUS elastography used for the differential diagnosis of chronic pancreatitis and pancreatic cancer. Gastrointest Endosc 2008; 68: 1086–1094. doi:10.1016/j.gie.2008.04.031
7. Sâftoiu A, Vilmann P, Dietrich CF, et al. Quantitative contrast-enhanced harmonic EUS in differential diagnosis of focal pancreatic masses (with videos). Gastrointest Endosc 2015; 82: 59–69. doi:10.1016/j.gie.2014.11.040
8. Zhu M, Xu C, Yu J, et al. Differentiation of pancreatic cancer and chronic pancreatitis using computer-aided diagnosis of endoscopic ultrasound (EUS) images: a diagnostic test. PloS One 2013; 8: e63820. doi:10.1371/journal.pone.0063820

9. Zhang M-M, Yang H, Jin Z-D, et al. Differential diagnosis of pancreatic cancer from normal tissue with digital imaging processing and pattern recognition based on a support vector machine of EUS images. Gastrointest Endosc 2010; 72: 978–985. doi:10.1016/j.gie.2010.06.042

10. Ozkan M, Cakiroglu M, Kocaman O, et al. Age-based computer-aided diagnosis approach for pancreatic cancer on endoscopic ultrasound images. Endosc Ultrasound 2016; 5: 101–107. doi:10.4103/2303-9027.180473

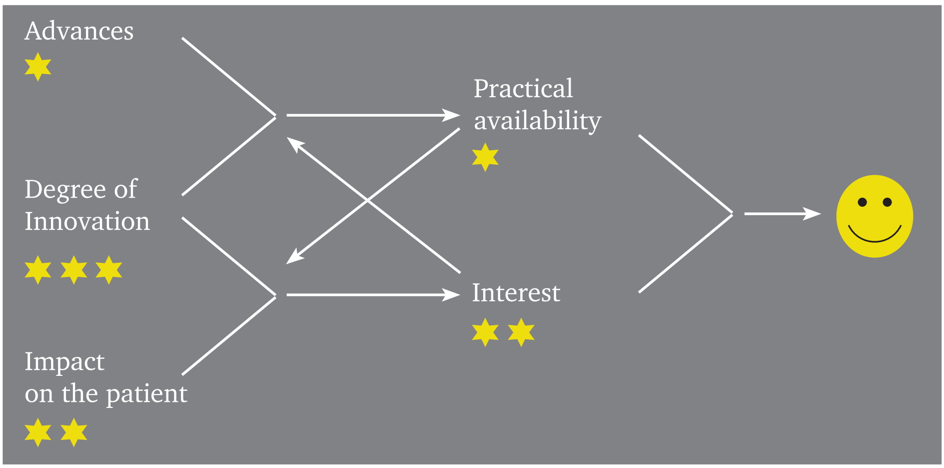
11. Tonozuka R, Itoi T, Nagata N, et al. Deep learning analysis for the detection of pancreatic cancer on endosonographic images: a pilot study. J Hepato-Biliary-Pancreat Sci 2020. doi:10.1002/jhbp.825

12. Marya NB, Powers PD, Chari ST, et al. Utilisation of artificial intelligence for the development of an EUS-convolutional neural network model trained to enhance the diagnosis of autoimmune pancreatitis. Gut 2020. doi:10.1136/gutjnl-2020-322821

13. Kuwahara T, Hara K, Mizuno N, et al. Usefulness of Deep Learning Analysis for the Diagnosis of Malignancy in Intraductal Papillary Mucinous Neoplasms of the Pancreas. Clin Transl Gastroenterol 2019; 10: 1–8. doi:10.14309/ctg.0000000000000045

14. Zhang J, Zhu L, Yao L, et al. Deep learning-based pancreas segmentation and station recognition system in EUS: development and validation of a useful training tool (with video). Gastrointest Endosc 2020; 92: 874-885.e3. doi:10.1016/j.gie.2020.04.071

At the border of other disciplines: New technologies



This pictorial representation based on the author's assessment has but an indicative value

by Lucille Quénéhervé
[1] *Department of Gastroenterology, CHU Brest La Cavale Blanche*

The progress made with Artificial Intelligence (AI) in digestive endoscopy is now so obvious that it tends to overshadow other technological advances and developments. Nevertheless, new imaging modalities that are not based on video-endoscopy, either using white light or virtual chromoendoscopy, also benefit from new developments in AI.

Some authors consider that two main branches of AI could be discerned in the medical field: a virtual branch and a physical branch [1]. The Virtual branch includes medical imaging, diagnostic aid, prognostic evaluation aid, etc., while the different types of robots belong to the physical branch. The applications of AI in endoscopic innovations are therefore multiple and in constant development.

Innovative Medical imaging technologies and AI

Confocal endomicroscopy

Confocal endomicroscopy (CEM) is a technology based on the detection of fluorescence emitted by a tissue and passing through a pinhole upon illumination by a laser. In general, it requires intravenous injection of a fluorescent contrast product with a defined wavelength. In digestive endoscopy, CEM can no longer be considered an innovative technology, but the prospect of performing targeted optical biopsies in real time thanks to the automatic interpretation of the images is raising interest. Indeed, the development of CEM is currently limited by the necessary expertise required to interpret the images that are very different from those which are obtained by endoscopy using white light. A few studies have been published on AI applied to CEM.

Upper Digestive tract

Barrett's mucosa, as a preneoplastic condition of the oesophageal mucosa, constitutes a perfect model system for the development of imaging techniques. In order to build and train a reliable computer-assisted system, a large amount of manually-labelled data is often necessary. Indeed, such pre-analysed data allow for the selection of the optimal subsets of features and they form the basis for a robust classification by supervised learning. While it is relatively easy to collect a large number of unlabelled images during each session of CEM, the collection of a large set of labelled CEM images is a long and expensive process. An American team has therefore proposed to improve the classification of images of Barrett's mucosa obtained by CEM by using unlabelled images using a strategy of semi-supervised learning [2]. A semi-supervised deep neural network based on convolutional auto-encoding has therefore been developed to improve the classification of Barrett's mucosa. The team postulated that if this technology works for analysing CEM images, it could also likely be able to analyse Optical coherence tomography (OCT) images, for example.

Lower Digestive tract

Improving screening and clinical decision making for colorectal cancer is also a field of active research for new endoscopic technologies. In a Romanian study, CEM images of healthy colonic mucosa and colic adenocarcinoma have been analysed by an AI system [3]. The computer-assisted diagnosis was applied to the images and the model was trained as a two layer neural network in order to perform an automatic diagnosis of malignancy according to seven parameters.

Optical coherence tomography (OCT)

Optical coherence tomography (OCT) is based on interferometry. A laser emits a beam of light that is divided into two arms, one directed towards the biological tissue (sample) and the other (reference) directed toward a mirror. When the sample light beam and the reference beam are reflected towards the sensor, they are combined to create interference. The corresponding profile is then analysed to determine the composition of the traversed media, according to the principle of time delay of the "echo". A conventional OCT device scans through the surface of a tissue at different depths, forming a longitudinal scan called scan amplitude or A-scan, just like with ultrasound. By assembling the different A-scans measured during beam scan, a 2D reconstruction called tomogram or B-scan is obtained. By stacking the B-scans, a 3D image is obtained.

Upper digestive tract

Volumetric Laser Endomicroscopy (VLE) is a technology derived from OCT, in which a laser is emitted at the end of a catheter and automatically and longitudinally withdrawn by rotation. The catheter is included in a balloon that is passed through the operator channel of the endoscope and inflated at the oesophagogastric junction, in order to perform a circumferential scan of the layers of the oesophageal wall with near micrometric resolution. Recent VLE devices also allow for the marking of areas of interest with the laser, eventually allowing their removal during the same endoscopic session. Such devices could improve the detection of Barrett's associated neoplasia, but the interpretation of VLE images is complex and takes a long time, due to the large amount of visual information with levels of gray.

A Dutch team showed in a first study that an algorithm trained with clinical information allowed for computer-assisted detection of dysplastic lesions within Barrett's mucosa [4]. This study was based on the analysis of VLE images taken ex vivo after scanning resected specimens and included a correlation with the findings of histological analyses. However, computer-assisted detection that analyses neighbouring sections would improve the detection of neoplasia in Barrett's oesophagus compared to single image analyses. The same team has therefore examined the feasibility of automatically extracting the data from multiple adjacent frames for computer-assisted detection of neoplasia in Barrett's mucosa [5]. However, these studies only relied on ex vivo-obtained images.

Another study was carried out [6], but this time from VLE images collected in vivo, demonstrating the effectiveness of encoding by main dimension (dividing a problem into many separate areas, in this case by separately analysing each column (A-line) of the image) followed by machine learning analysis. These encouraging data call for clinical validation in real time.

Meanwhile, a US team has developed a technology to improve the VLE images in real time and identify areas of interest thanks to AI [7] (Figure 1). This scanning allows the endoscopist to mark lesions with a laser and perform targeted resection during the same endoscopic session. A prospective study is underway in the USA to validate this technique.

Lower digestive tract

An American team has developed a convolutional neural network in order to recognise the pattern of human colon mucosa from images obtained with OCT scanning. The network was trained with images obtained ex vivo from surgical specimens and could successfully distinguish between normal colorectal mucosa and neoplastic tissues with an area under the curve (AUCROC) of 0.998 [8]. The authors suggest that the OCT coupled with pattern recognition could give a precise computer-assisted real time diagnosis during a colonoscopy.

Hyperspectral imaging

Hyperspectral imaging is based on the collection and analysis of narrow and continuous strips with wavelengths that are not limited to the visible spectrum, but instead the whole electromagnetic spectrum. This tool could be used to determine tumor margins during a surgical tumor resection [9]. Within oesophageal resection samples, different areas were classified as cancerous areas or not (Figure 2). The objective of the researchers was to develop a perioperative modality that would guide the surgical procedure. In the future, we predict that such an approach could for example help to determine the margins of a lesion before performing a submucosal dissection.

Robotic endoscopes and AI

Several teams are currently developing robots for flexible endoscopy in order to: 1) facilitate the progression in the digestive tract while limiting undesirable effects of the examination; 2) assist the operator during a technical procedure, such as a resection for example [10-12]. Their practical implementation implies that these devices are able to complete the operator procedure, and not only be equal to the operator. Currently, the models that aim to help with «active» endoscopy do not work in complete

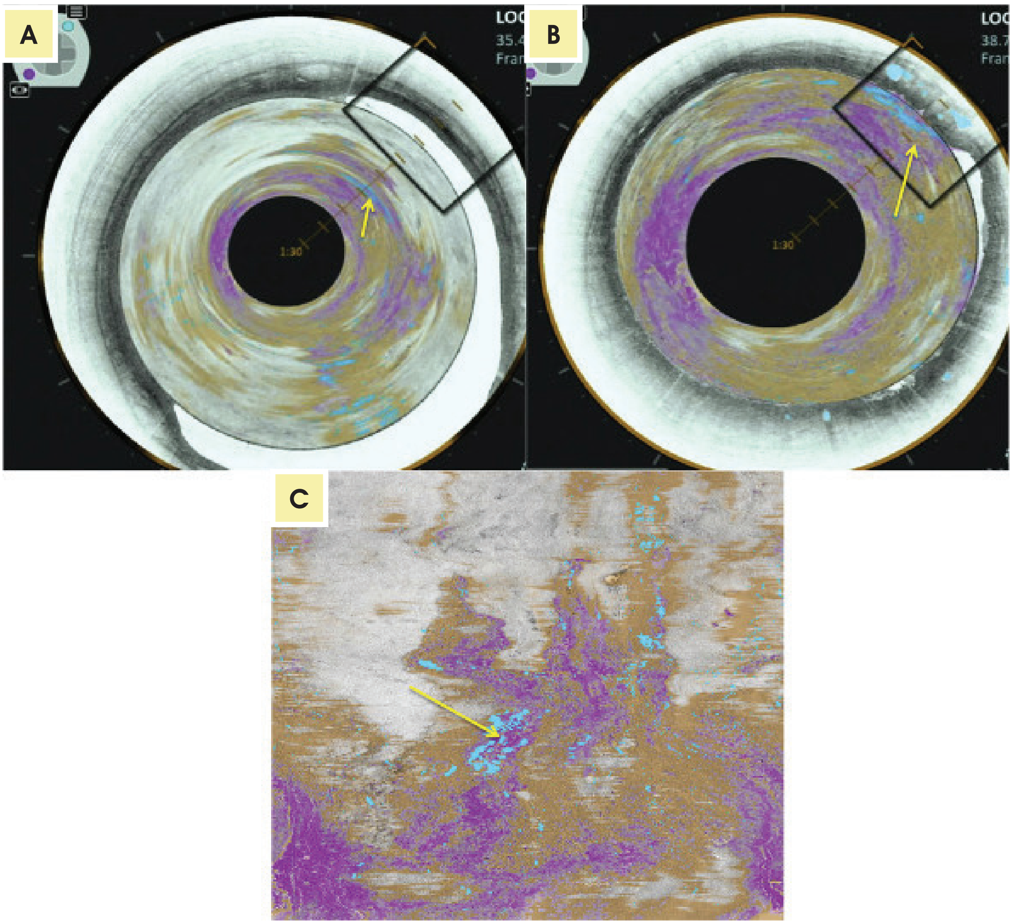


FIGURE 1. Volumetric Laser Endomicroscopy of Barrett's mucosa *in vivo* with image processing by artificial intelligence. A luminal view facing an area of overlap (yellow arrow) with the three characteristics of dysplasia (orange: lack of stratification, blue: glandular structures, pink: hyper-reflective surface). (A) View of the proximal oesophagus. (B) A closer view of the area suspected of dysplasia. The en face view is also represented (C). Taken from [7].

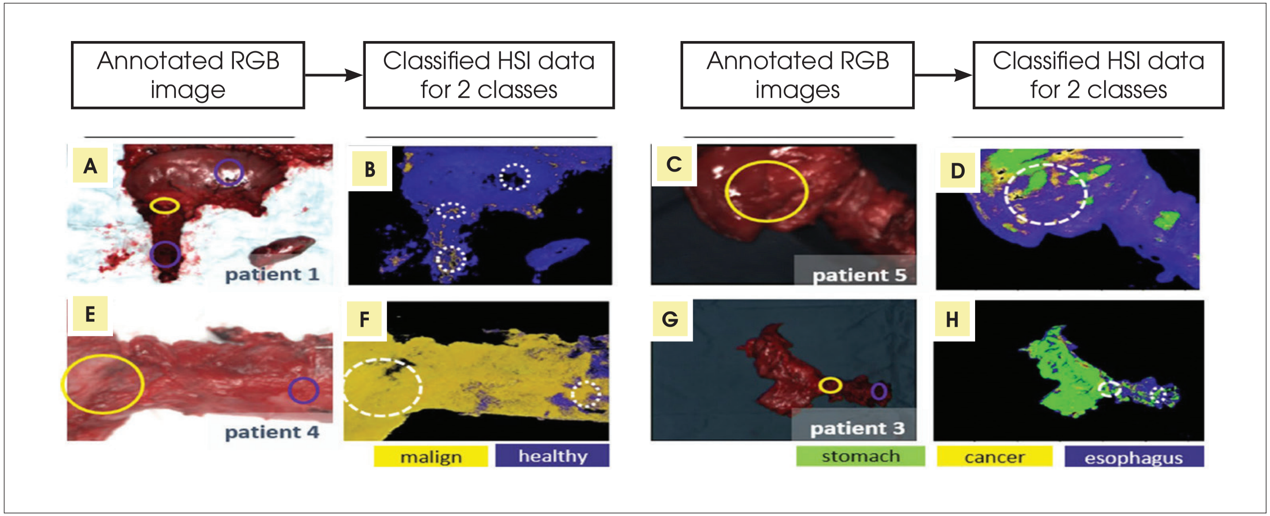


FIGURE 2. Hyperspectral Imaging assisted with artificial intelligence applied for the determination of tumor margins in surgically-resected oesophaeal cancer samples. A two-class approach with annotated microphotographs (A,E) and hyperspectral images (HSI) (B,F) and, on the right, a three-class approach applied to annotated microphotographs (C,G) and hyperspectral images (D,H). According to [9].

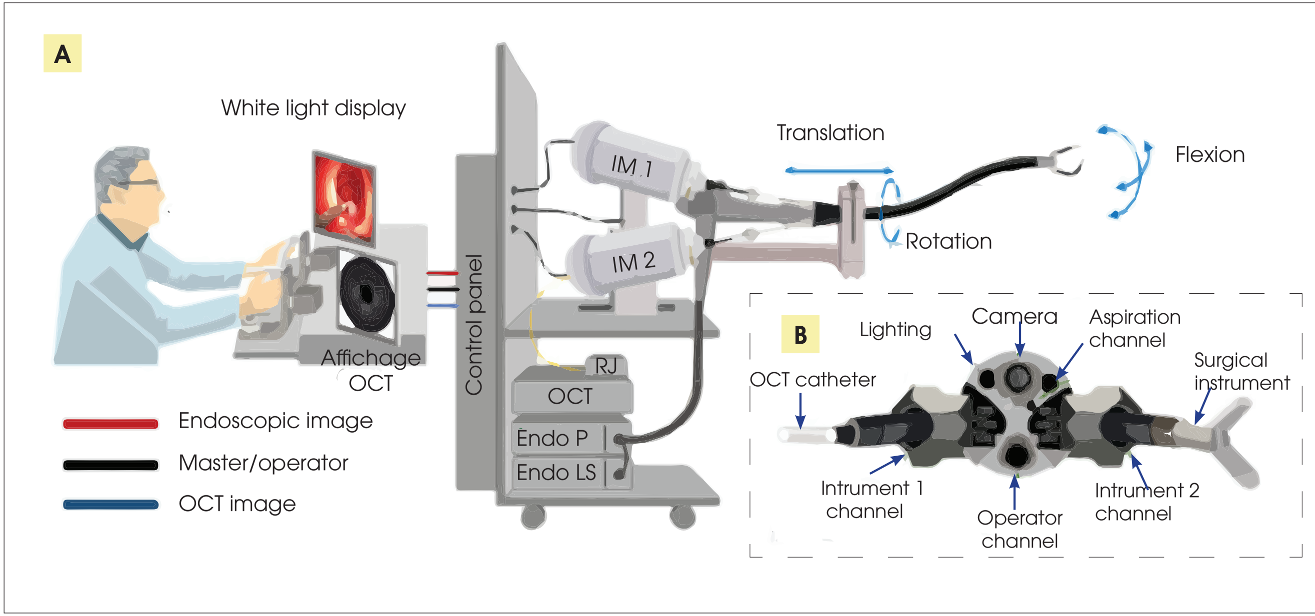


Figure 3. Tomography module for optical coherence imaging adapted on an Anubiscope robot (Karl Storz). A) A scheme depicting the flexible interventional endoscope with the steerable OCT (Optical Coherence Tomography) catheter attached to a «slave» cart is connected to user controllers for teleoperation of the device: instrument driver module (IM1, IM2), volumetric scanning actuator (RJ), OCT system (OCT), Endoscopic processor (Endo P), Endoscopic Light source (Endo LS) ; B) A frontal view of the distal end of the robotised flexible interventional endoscope with steerable OCT catheter. Taken from [13].

autonomy and they are in general controlled by the operator, a problem that could be solved once AI can interpret images in real time. In addition, models that use triangulation have not yet been tested for the resection of large/difficult lesions.

A French team has recently combined robotic and optical technological innovations and set up a system that analyses the colonic mucosa with OCT while a therapeutic procedure is performed [13]. The system consists of a steerable OCT cathether inserted in one of the two instrument channels of a robotised flexible interventional endoscope, the other channel being for example used for surgical forceps (*figure 3*).

Conclusion

Progress in videoendoscopy combined with progress in AI is so spectacular that new technologies are requested to provide practical solutions for problems that are considered impossible with classical endoscopy and in less-invasive conditions. Regarding imaging, AI can reduce the necessary expertise and time for their use. Regarding robotics, the applications of AI are numerous, in order to analyse the digestive tract and the borders of the lesions.

REFERENCES

1. Hamet P, Tremblay J. *Artificial intelligence in medicine.* Metabolism. 2017;69S:S36–S40.
2. Tong L, Wu H, Wang MD. CAESNet : Convolutional AutoEncoder based Semi-supervised Network for improving multiclass classification of endomicroscopic images. *J Am Med Inform Assoc.* 2019;26:1286-96
3. Ştefânescu D, Streba C, Cârtânâ ET, *et al.* Computer Aided Diagnosis for Confocal Laser Endomicroscopy in Advanced Colorectal Adenocarcinoma. *PLoS One.* 2016;11:e0154863.

4. Swager A-F, van der Sommen F, Klomp SR, *et al.* Computer-aided detection of early Barrett's neoplasia using volumetric laser endomicroscopy. *Gastrointest Endosc* 2017;86:839–846.

5. Struyvenberg MR, van der Sommen F, Swager AF, *et al.* Improved Barrett's neoplasia detection using computer-assisted multiframe analysis of volumetric laser endomicroscopy. *Dis Esophagus* 2020;33.

6. van der Putten J, Struyvenberg M, de Groof J, *et al.* Deep principal dimension encoding for the classification of early neoplasia in Barrett's Esophagus with volumetric laser endomicroscopy. *Comput Med Imaging Graph* 2020;80:101701.

7. Trindade AJ, McKinley MJ, Fan C, *et al.* Endoscopic Surveillance of Barrett's Esophagus Using Volumetric Laser Endomicroscopy With Artificial Intelligence Image Enhancement. *Gastroenterology* 2019;157:303–305.

8. Zeng Y, Xu S, Chapman WC, *et al.* Real-time colorectal cancer diagnosis using PR-OCT with deep learning. *Theranostics* 2020;10:2587–2596.

9. Maktabi M, Köhler H, Ivanova M, *et al.* Tissue classification of oncologic esophageal resectates based on hyperspectral data. *Int J Comput Assist Radiol Surg* 2019 ; 14 : 1651–61.

10. Li Z, Chiu PW-Y. Robotic Endoscopy. Visc Med. 2018;34:45–51.

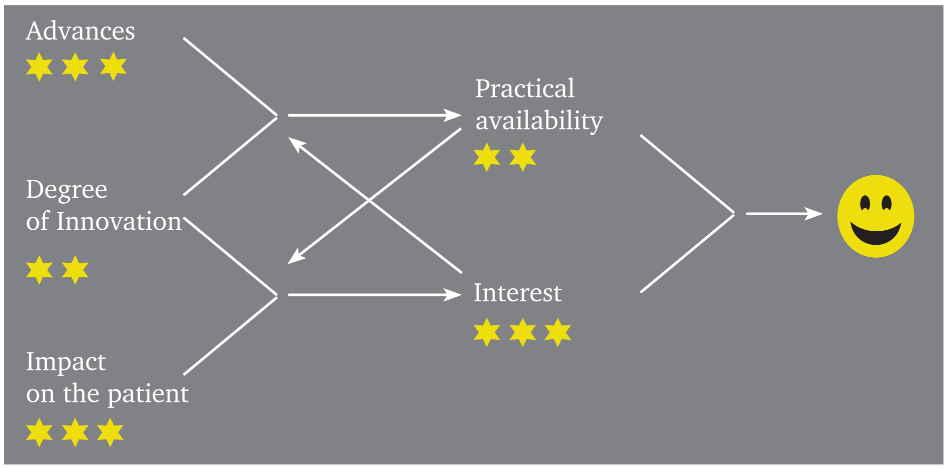
11. Yeung BPM, Chiu PWY. Application of robotics in gastrointestinal endoscopy: A review. *World J Gastroenterol* 2016 ; 22 : 1811–25

12. Visconti TA de C, Otoch JP, Artifon EL de A. Robotic endoscopy. A review of the literature 1. *Acta Cir Bras* 2021 ; 35. : epub. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7184939/.

13. Mora OC, Zanne P, Zorn L, *et al.* Steerable OCT catheter for real-time assistance during teleoperated endoscopic treatment of colorectal cancer. *Biomed Opt Express* 2020;11:12 : 1231-43

At the border of other disciplines: Pathology

This pictorial representation based on the author's assessment has but an indicative value



by Céline Bossard, Raphaël Bourgade
Department of Pathology, University Hospital of Nantes

Definition of digital pathology and introduction to Artificial Intelligence

In a tense context of medical demography, the time of the pathologist is increasingly constrained, in part because of the increasing complexity of the patient diagnosis. In the era of personalised medicine, the diagnosis is increasingly based on the report of the pathologist and it needs to take into account a number of precise diagnostic classifications and a growing number of prognostic and theranostic factors. All of these characteristics place pathology at the heart of the healthcare pathway, and make it a cornerstone of therapeutic decision making, especially in oncology.

Since the year 2000 and the emergence of new scanners able to perform Whole Slide Imaging (WSI), several hundreds of pathological slides can be digitized in a few hours, bringing our discipline to the digital era. Digital pathology, also called virtual microscopy, is an attractive educational tool for young pathologists. It also allows future developments and the use of Artificial Intelligence (AI) in pathology, as evidenced from the growing number of research publications on this topic. All of these developments should soon lead to a revolution in the daily practice of the pathologist.

The so-called convolutional artificial neural networks (CNN), directly inspired by the visual cortex of animals, allow for the extraction and analysis of a considerable amount of information contained in virtual slides (WSI), in contrast to the small amount of information that is actually exploited by the eye of the pathologist. These massive data do not only contribute to the development of diagnostic algorithms (tissue and cell recognition ...), but also to the prediction of the tumor mutational status/its molecular signature and evolution.

Technical constraints of WSI

While the future looks promising, application of AI in pathology is still in its infancy because of various constraints and current limitations.

Due to the very high level of resolution related in particular to their pyramidal architecture (*figure 1*), WSI are complex data to analyse. A slide that is digitized at a magnification of x400 (resolution of about 0.2 μm/pixel), represents more than 200,000 x 100,000 pixels (>20Gpx) and is several gigabytes in size in an uncompressed format [1]. The corresponding data are too large to be processed as such and require a preprocessing step, consisting in particular of cutting them into square tiles usually measuring between 128 and 1,024 pixels on each side (*figure 2*) [2]. The removal of empty tiles reduces the volume of the dataset.

During supervised learning, the CNN are trained with digital images that are labelled by expert pathologists and that are therefore regarded as the «reference», or ground truth. Unfortunately, only a few databases with annotated images exist and their size is usually limited. Indeed, because of the cost required for their storage and implementation, that depends on computers with large storage capacities, only a small number of laboratories are currently using high-speed WSI scanning solutions integrated into their usual workflow. However, the performance and reliability of algorithms, which condition their validation and use in clinical routine, increase with the number and quality of available images. To overcome this current limitation and prevent over-fitting of algorithms, data augmentation techniques are systematically applied in order to virtually

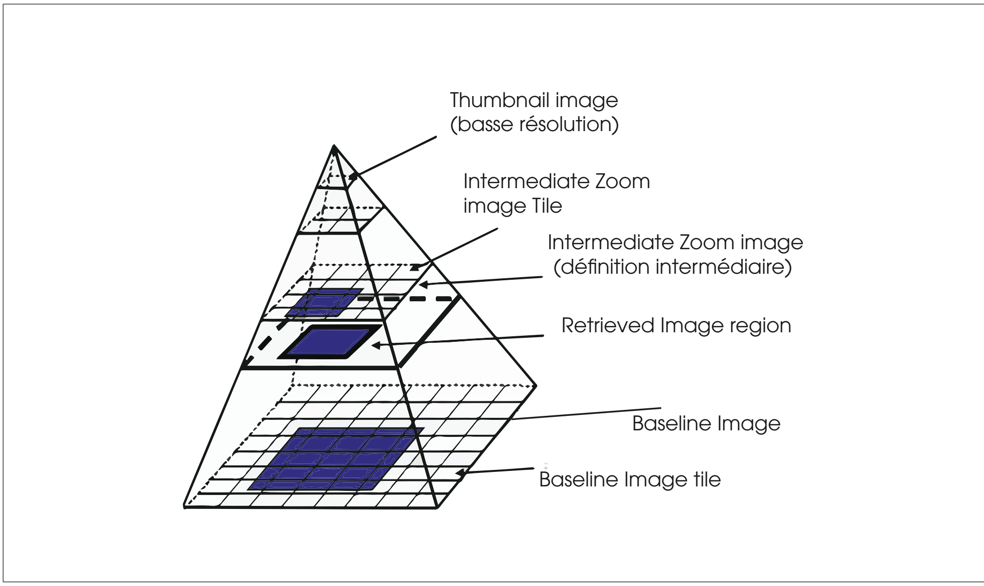


FIGURE 1.. Representation of the pyramidal architecture of WSI (Whole Slide Imaging), adapted from Lajare *et al* [1]

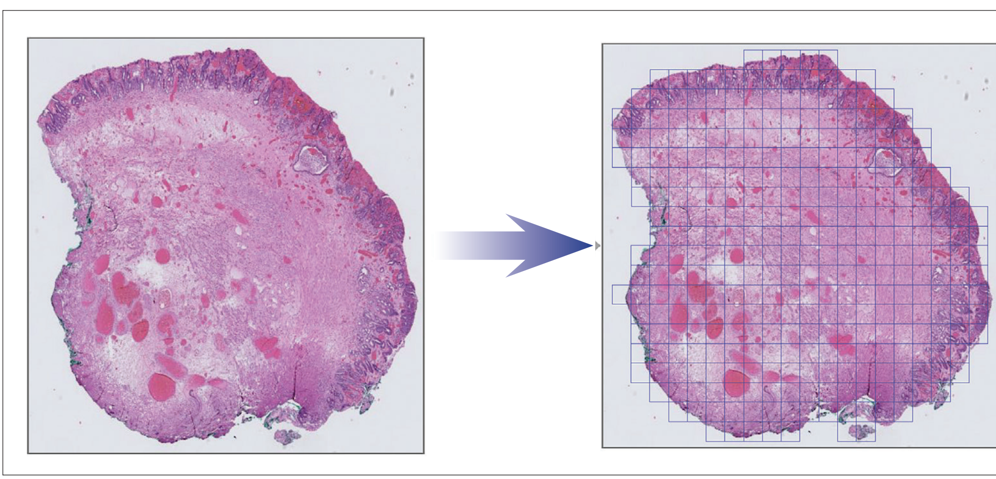


FIGURE 2. Tiling of a histopathological slide, adapted from a figure by Sali *et al* [2].

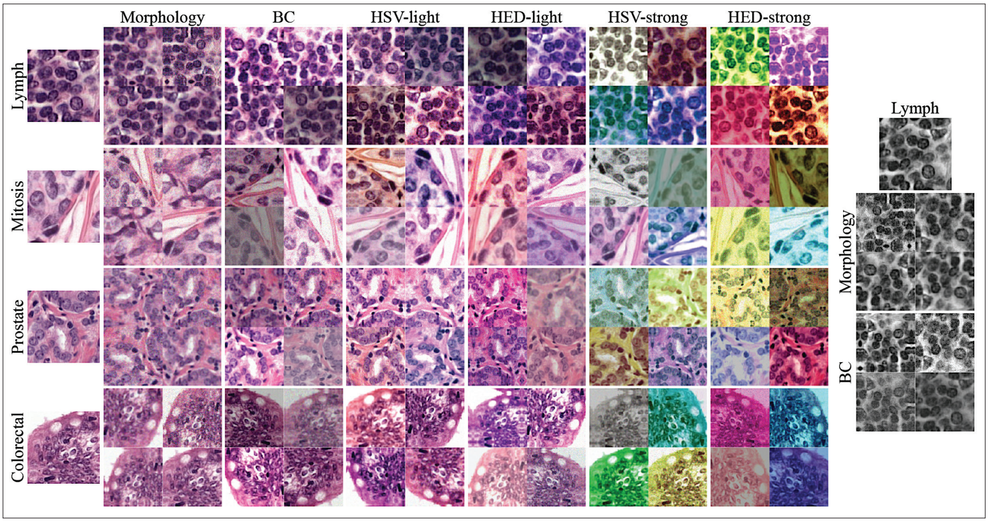


FIGURE 3. Summary of the main techniques of data augmentation applied to 4 tiles. BC : luminosity and contrast. HSV: saturation and tint. HED : Hematoxylin-Eosin-DAB. Adapted from Tellez *et al* [3].

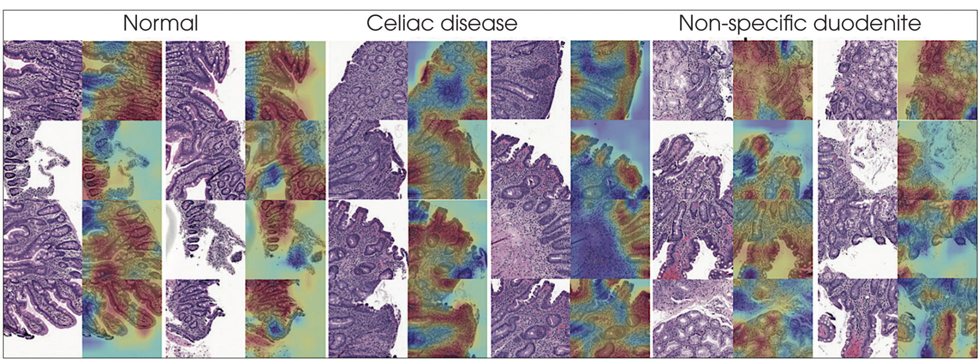


FIGURE 4. Heatmaps illustrating the importance of each tile during WSI classification as «normal», «celiac disease» and «non-specific duodenitis». The more the tile tends towards red, the greater its impact on prediction. Adapted from Wei *et al.* [4].

multiply the training dataset. These techniques generally consist of rotation, inversion, magnification or even variation of staining, saturation or brightness (*figure 3*) [3].

Finally, the validation step, based on a metric and providing a direct measurement of the performance of the system, relies on an external dataset that must be as varied as possible, in order to cope with the heterogeneity of the pre-analytical phases (fixation, impregnation, cutting, staining, mounting ...) found between different pathology laboratories. This current limitation is likely to be overcome in the near future, thanks to the automation of the workflow (that still remains mostly manual these days) for the fixed samples as they are put in a cassette.

Applications of artificial intelligence to digestive pathology

Diagnostic classification

Diagnostic classifications based on the recognition of tissues/cells are at the heart of the work of a pathologist, and they are a research topic in artificial intelligence, especially in digestive oncology, as summarised in a recent review [4]. For example, O. Iizuka and his team became interested in the classification of gastric and colonic epithelial tumors from surgical biopsies. They set up a CNN trained with >4,000 slides in order to classify lesions as «adenocarcinoma», «adenoma» and «no neoplastic lesion». The corresponding ROC curves were determined on an external validation cohort obtained from a different center, and found >0.972 for colon adenocarcinoma and >0.966 for gastric adenocarcinoma. In addition, in order to assess the broad applicability of their model and to define its possible limitations, a second external evaluation was carried out using data from TCGA (The Cancer Genome Atlas) cohort, consisting of surgical samples. The ROC curves were determined to be 0.982 (0.968 to 0.991 95% CI) for colon adenocarcinoma and 0.924 (0.887-0.952 95% CI) for gastric adenocarcinoma, demonstrating the model's ability to analyse images of different sizes and different levels of complexity without direct supervision.

Another important demand of pathologists regarding the performances of AI in daily practice, in addition to the reliability of the algorithms, is the medical time required for diagnosis. In this same study, 23 pathologists were thus evaluated for their interpretation of 45 gastric adenocarcinoma slides, with a maximum time of 30 seconds per case. Human intelligence reached an average level of accuracy of 85.9% of correct answers, well below the trained model, which takes between 5 and 30 seconds to make a diagnosis with an accuracy of >95%.

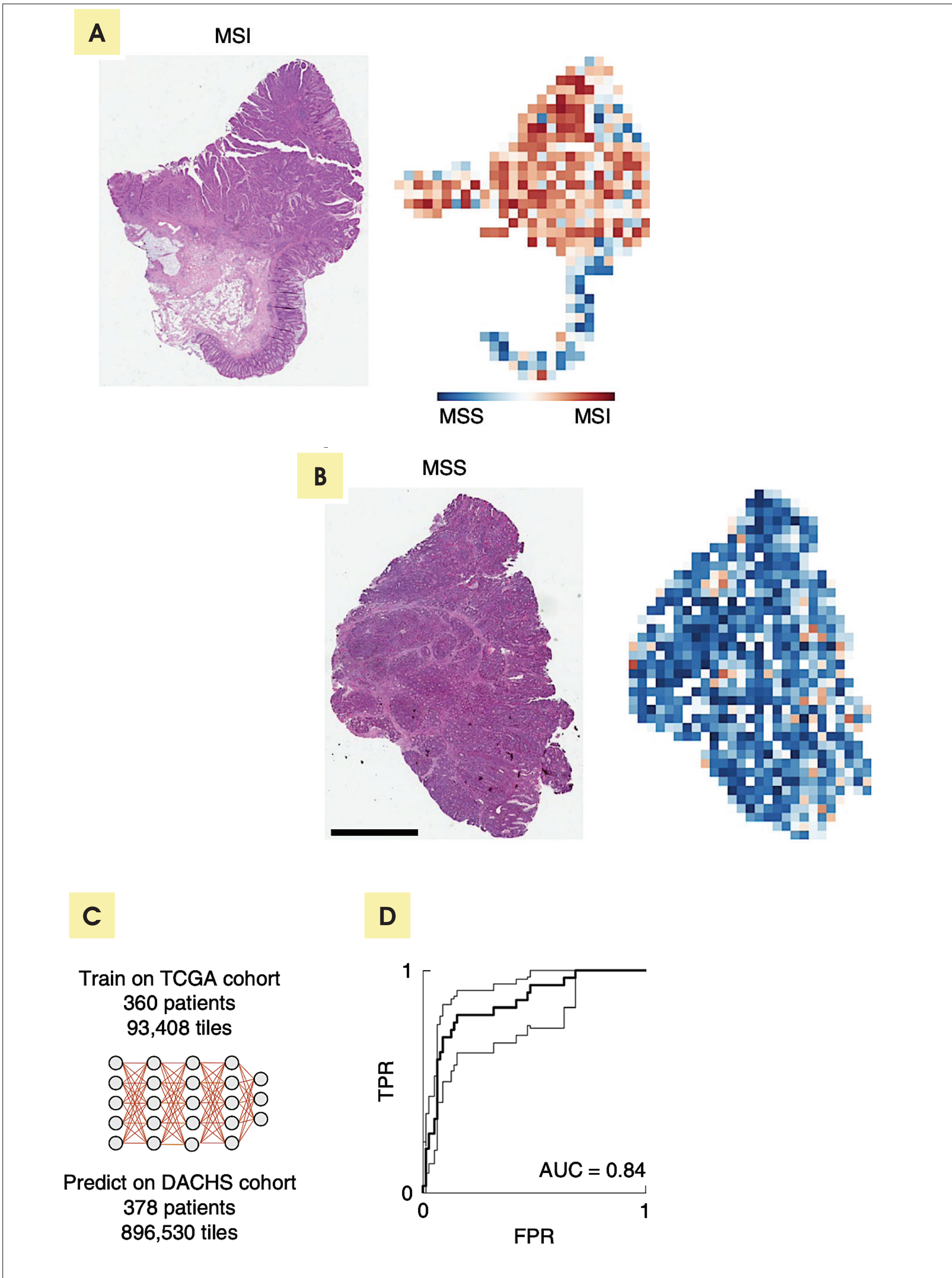
Other works focused on Barrett's oesophagus and the recognition of metaplastic and/or dysplastic lesions or adenocarcinomas from sample biopsies from a dataset consisting of more than 500 slides and had an average accuracy of 0.83 (0.80-0.86 95% CI) [5]. A. Kiani studied the impact of a trained model on the performance of 11 pathologists regarding the ability to distinguish cholangiarcinomas from hepatocellular carcinomas, proving the complementarity of this association [6].

The application of deep learning to diagnostic classification is also possible outside of the field of tumor pathology. Some teams managed to train an algorithm to distinguish nonspecific duodenitis lesions from autoimmune lesions linked to celiac disease on duodenal biopsies with an accuracy of >95% [7,8]. Heatmaps representing the importance of each tile in the final prediction are shown in Figure 4.

Nevertheless, while these classifications are mainly based on the global analysis at the tissue level, they could also be based on cell analyses. This was shown by Y. H. Chang in 2017, by correlating WSI analysis of pancreatic tumors stained with HES (white light) with their counterparts stained with DAPI (fluorescent light). By combining segmentation performed with machine learning and classification with deep learning, they were able to develop a model able to distinguish tumor cell nuclei from normal cell nuclei with an accuracy >90%.

Substitution for complementary techniques: prediction of abnormalities and molecular signatures

Although being extensively studied, diagnostic classification from WSI is not the only potential application of deep learning. Some teams have indeed developed algorithms that are able to predict the mutational status of a tumor or its gene expression profile by the mere analysis of WSI stained with HES. This very promising approach would make it possible to skip heavy complementary techniques that can be cumbersome and time-consuming, such as immunohistochemistry or even molecular biology analyses based on NGS (Next Generation Sequencing), WES (Whole Exome Sequencing), etc. These techniques are essential for the identification of diagnostic, prognostic and theranostic markers, that are essential for the therapeutic management of the patient, especially in the era of personalised medicine. Recent publications have included the determination of the consensus molecular subtypes of colorectal cancer, with a ROC curve determined to be at 0.84 [9]. Similarly, J.N. Kather and his team have shown that deep learning analysis of HES WSI could predict the presence of certain mutations (TP53, KRAS, BRAF...) as well as the presence of molecular signatures related to the tumor phenotype by AI [10]. Using these models in routine for diagnosis would permit a low-cost screening and might be interesting in order to identify the samples in which selected molecu-


A
MSI
MSS
MSI
B
MSS
C
Train on TCGA cohort
360 patients
93,408 tiles
Predict on DACHS cohort
378 patients
896,530 tiles
D
1
0
TPR
AUC = 0.84
0
1
FPR

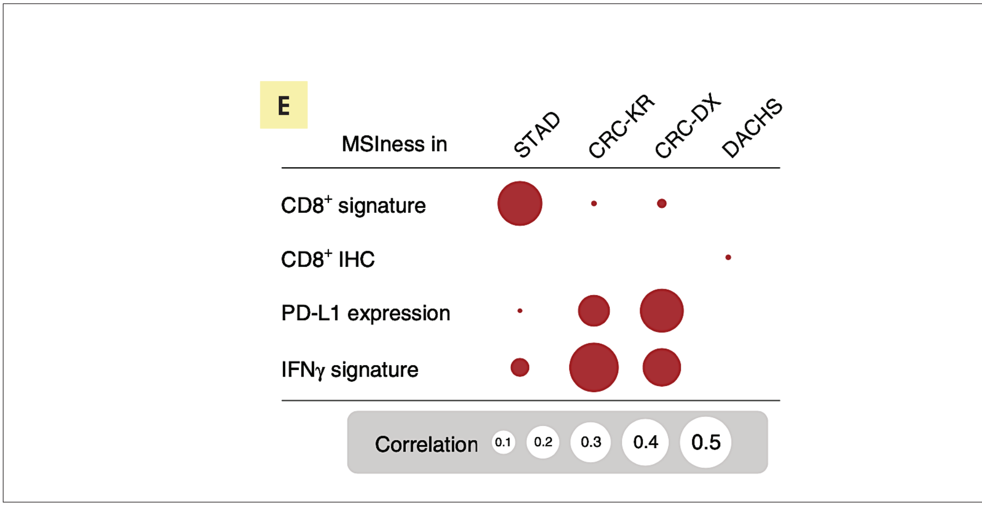


FIGURE 5. A) and B), heatmaps showing the distribution of tiles predicted to be MSI (microsatellite instability) or MSS (microsatelite stability) in colon cancer ; C) the model was trained on a TCGA cohort, comprising 360 patients with colorectal cancer, then confirmed with an independent cohort of 378 patients ; D) ROC curve detemined to be at 0.84 (0.73-0.91 95% CI) using the external validation cohort. TPR : true positive rate (sensitivity) ; FPR : false positive rate (1-specificity) ; E) Pearson coefficient measuring the correlation between the ratio of tiles predicted to be MSI (MSIness) and the expression of immunohistochemical and transcriptomic markers in gastric adenocarcinoma (STAD) and colorectal cancer (CRC-KR, CRC-DX & the cohort DACHS). Adapted from Kather *et al* [11].

lar analyses should be performed. J. N. Kather has also addressed the possibility of predicting the microsatellite status of gastric and colorectal adenocarcinomas with a ROC curve of 0.84 (0.72-0.92 CI 95 %) [11]. In order to strengthen their results, the authors have compared the proportion of tiles predicting MSI with the expression of the markers known to be associated with this status (Figure 5). The spatial visualisation of the molecular heterogeneity, as shown using heatmaps, will improve the understanding of the mechanisms of carcinogenesis.

Prediction of the course of the disease

Today, all morphological markers (pTNM, mitosis, budding, differentiation, grade and vascular emboli), both phenotypic and molecular, are considered essential to establish a reliable prognosis, but many studies suggest that AI could help to skip some of these analyses. O-J. Skrede has for example set up a prognostic model based on WSI of surgically-resected colorectal adenocarcinoma that can predict patient survival [12]. This model, based on a dozen CNN was trained with more than 12,000,000 tiles taken from the slides of 828 patients, and it was able to differentiate between two groups of patients according to their prognosis with a hazard ratio to 3.84 (2.72-5.43; $p<0.0001$). In a univariate analysis, the hazard ratio was at 3.04 (2.07-4.47; $p<0.0001$) after adjustment for known prognostic factors (pT, pN and venolymphatic emboli).

D. Bychkov performed a similar study, by comparing three different approaches to differentiate between patients with good or bad prognosis from a WSI analysis of colorectal cancer with deep learning [13]. The first approach was based on tumor grade estimation and Duke classification by three experienced pathologists. The second approach relied on the use of classical algorithms of machine learning, such as Logistic Regression, Naive Bayes classifier or support vector machine (SVM). Finally, the third approach was based on deep learning and recurrent neural network (LSTM: Long Short-Term Memory) trained with patterns extracted by a dedicated neural network (ACV-16). The performances of these three methods, measured at different resolutions and illustrated in Figure 6, demonstrate the superiority of automated approaches, when the resolution is sufficient.

Conclusion and outlook

Digital pathological analyses by AI open up a limitless field of investigations for optimised and personalised care of patients. By facilitating the integration and processing of all data, either morphological or phenotypic, clinical, biological, radiological or molecular, deep learning will certainly dramatically improve the diagnostic, prognostic and theranostic approaches, generating new markers and classifications.

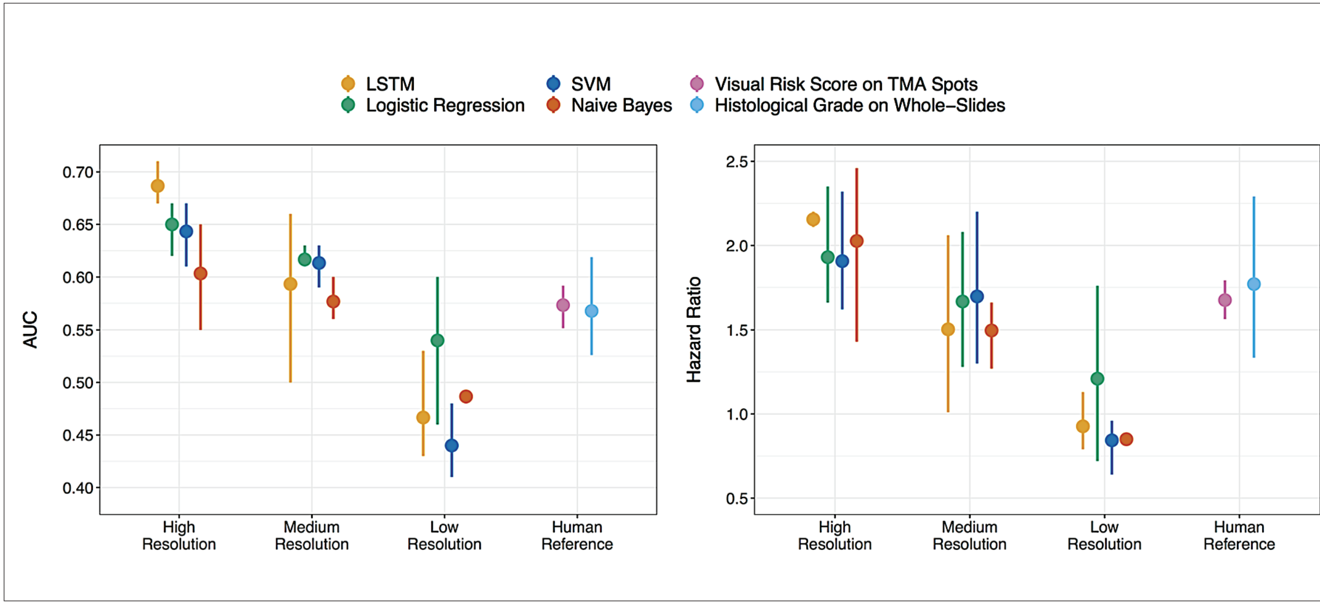


FIGURE 6. Representation of the area under the curve (AUC) and the hazard ratio, comparing the predictive performance of 4 automated models with human evaluation. The vertical bars indicate the possible variation, while the points show the mean. These measurements were done on microarray tissue images at different resolution (« high » corresponds to a resolution of 0.22 μm/px, while « medium » and « low » correspond to compressed images by a factor of 4 and 16, respectively. Adapted from Bychkov *et al.* [13].

However, the rise of these new tools inevitably generates ethical questions. Their reliability will be a prerequisite and will be a sine qua non condition for the confidence of the pathologist and the patient. This reliability directly depends on the quality of the data used in their training, as well as on their validation before their routine deployment. Their foundation and the basis of their reasoning must be understood and criticised by the pathologist in order to be safely delivered to the patients.

This new era of digital pathology will profoundly improve the work of the pathologist, who will not be replaced, but assisted by AI. Pathologists must become key players in this digital transformation, in order to benefit from the augmented performance of a third eye!

REFERENCES

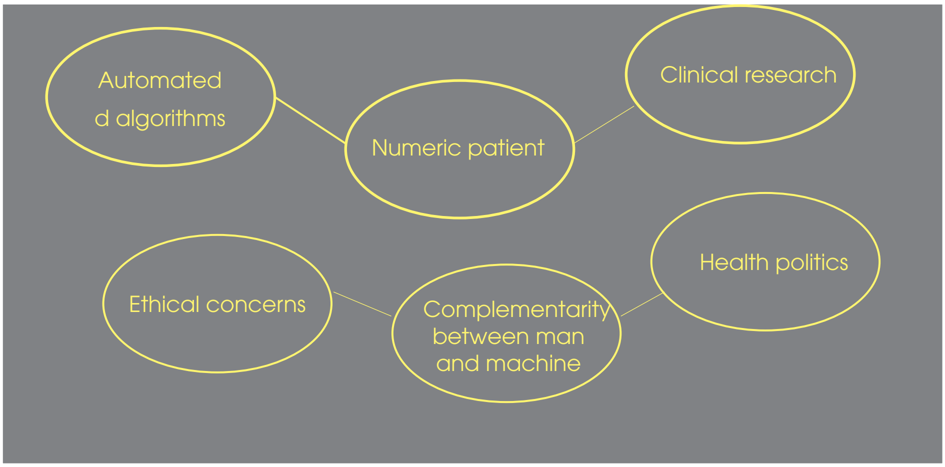
1. Lajara N, Espinosa-Aranda JL, Deniz O et al. Optimum web viewer application for DICOM whole slide image visualization in anatomical pathology. Computer Methods and Programs in Biomedicine 2019 ; 179 : 104983.
2. Sali R, Moradinasab N, Guleria S et al. Deep Learning for Whole-Slide Tissue Histopathology Classification: A Comparative Study in the Identification of Dysplastic and Non-Dysplastic Barrett's Esophagus. Journal of Personalized Medicine 2020 ; 10 : 141.
3. Tellez D, Litjens G, Bandi P et al. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Medical Image Analysis 2019 ; 58 : 101544.
4. Calderaro J, Kather JN. Artificial intelligence-based pathology for gastrointestinal and hepatobiliary cancers. Gut 2020 : epub. doi : gutjnl-2020-322880.
5. Tomita N, Abdollahi B, Wei J et al. Attention-Based Deep Neural Networks for Detection of Cancerous and Precancerous Esophagus Tissue on Histopathological Slides. JAMA Netw Open 2019 ; 2 : e1914645.
6. Kiani A, Uyumazturk B, Rajpurkar P et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer. npj Digit Med. déc 2020 ; 3 : 23.
7. Wei J, Wei J, Jackson C et al. Automated detection of celiac disease on duodenal biopsy slides: A deep learning approach. J Pathol Inform. 2019;10 : 7.
8. Srivastava A, Sengupta S, Kang S-J et al. Deep Learning for Detecting Diseases in Gastrointestinal Biopsy Images. In: 2019 Systems and Information Engineering Design Symposium (SIEDS). Charlottesville (Etats-Unis) : IEEE; 2019
9. Sirinukunwattana K, Domingo E, Richman SD et al. Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. Gut. 20 juill 2020;gutjnl-2019-319866.
10. 0. Kather JN, Heij LR, Grabsch HI et al. Pan-cancer image-based detection of clinically actionable genetic alterations. Nat Cancer. août 2020;1 : 789-99

11. Kather JN, Pearson AT, Halama N et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med. juill 2019;25(7):1054-6.

12. Skrede O-J, De Raedt S, Kleppe A et al. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. The Lancet. févr 2020;395(10221):350-60.

13. Bychkov D, Linder N, Turkki R et al. Deep learning based tissue analysis predicts outcome in colorectal cancer. Sci Rep. déc 2018;8(1):3395.

Perspectives in the domain of AI & expected impacts



This pictorial representation based on the author's assessment has but an indicative value.

par Robert Benamouzig
Department of Gastroenterology and Digestive Oncology,
Hospital Avicenne,AP-HP, University Paris Nord Sorbonne

Artificial Intelligence has already proven to be of practical interest for the prediction of the diagnosis of dermatological lesions, such as melanoma, or the recognition of certain retinal diseases. An artificial intelligence system with an interactive robot has been implemented this year in some emergency departments in Israel, the US and Great-Britain to help sort and streamline the flow of patients. Artificial intelligence is rising as a powerful means of exploiting data from radiological imaging and more recently, endoscopy and pathology. The previous chapters of this book devoted to AI in our discipline have certainly convinced you that major advances are expected with this approach in digestive endoscopy and other areas. The assistance in the detection and characterisation of digestive lesions gained with AI, while still in its infancy, will soon be essential.

Healthcare is a particularly important domain of application for the algorithms that specialise in the automated analysis of digital data, i.e. Artificial Intelligence.

This is explained by two important observations: the economic and societal stakes and the increase in health data digitization seen in the past 15 years. This digitization, estimated to be under 15% in the early 2000s, is now at over 95%. A recent analysis by the French Ministry of Economy identified artificial intelligence in health as a strategic economic stake for France, with around 150 existing companies of varying size [1]. The importance of this issue is illustrated by the great French public investments that are underway, such as those given for the implementation of a health data platform (HDP), i.e. the «Health Data Hub» (HDH) [2]. This HDP was created on November 29th, 2019, in order to facilitate sharing and comparison of health data originating from multiple sources, and to promote research. Its creation follows the recommendations of the report delivered in March 2018 by the French deputy Cédric Villani entitled «Making sense of artificial intelligence: for a national and European strategy». The evolution of the status of this structure, from a public interest group with a complex governance inhibiting its action to that of a company with simplified shares where the French state keeps the control with the majority of shares, is discussed at the parliamentary level. The existence of a privileged partnership between the HDP and a key private actor of the sector, Microsoft Health, implies that individual data are exported outside of the European Union, which is the subject of a debate regarding issues of law and sovereignty. On the other hand, the development of this approach will require addressing the challenges of technical interoperability between the many computer tools used, as well as issues regarding data stability over time.

Artificial Intelligence will help to improve the diagnosis of diseases and new classifications are expected that will not only be based on «conventional» data, such as clinical, radiological, biological, pathological and molecular data, but that will also take into account the analysis, without a priori, of big data gathered in medico-administrative databases [3].

The combination of these big data produced in Health culminates in the global concept of the «digital» patient. The patient's identity is associated with a large array of data consisting of images, biological data including genetic, metabolomic, microbiotic as well as environmental or behavioral data.

Supervised or unsupervised analyses of these data will allow for a better diagnosis and a more accurate prediction of prognosis and therapeutic response. By so doing, this approach will contribute to the development of Precision Medicine. The data will also perhaps allow targeted preventive actions.

Clinical research is also already concerned. From this year on, an artificial intelligence system developed by Microsoft Health will permit an interaction with the patients consulting the clinicaltrial.gov website in real-time, which presents most of the clinical trials that are underway in a synthetic form. A conversational agent or chatbot will allow the collection of information via keyword analysis. The chatbot will guide patients toward the clinical trials that may interest them, noting the information collected during the electronic exchange. This democratisation of the approach could lead to increased awareness among patients and perhaps might facilitate their recruitment. Other chatbots have been developed to help improve patient follow-up during clinical trials. They will reduce the time between the occurrence of a possible side effect and its declaration. They will also replace some of the follow-up notebooks, which should improve monitoring, adherence to treatment and data collection, leading to optimisation of the time of research technicians.

Artificial intelligence could also be an instrument of political health management at the local level. The study of healthcare pathways, the study of their prognostic implications should lead to structural improvements. Comparisons between territories will refine the choice of care and improve the management of the resources that are already available.

These developments involve patients and more widely citizens that will need to be informed if any diagnostic or therapeutic system involving artificial intelligence is proposed. Clear information should be considered, as is already done when a therapy is chosen, by honestly presenting the expected beneficial effects and the potential secondary effects.

A global human supervision of artificial intelligence systems is and will remain necessary. Maintaining a high level of independent expertise in these systems is essential to ensure sufficient mastery. Administrative marking, while not yet anticipated, should be considered and maybe should be inspired by the model of the French «Agence nationale de sécurité du médicament et des produits de santé» (National Agency for the safety of Drugs and Health products, ANSM).

Doctors will remain essential in the clinical decision process and they will never be replaced by a machine, no matter how important its contribution or its degree of «Intelligence» may be. The doctors should keep control over machines, a major objective that is hardly attainable but that must be kept in mind during the development and adoption of a system. Alongside the implementation of these systems, appropriate initial teaching must be considered as well as continued training. This book will help.

Once its usefulness is established, the use of AI in Medicine will naturally prevail. It will imply issues regarding training on "primitive" technologies for young colleagues that are born with these new technologies already being available. How can new generations be trained for the optimal use of these systems while maintaining independent human intellectual expertise outside the system? These questions may seem superfluous today, but they must be anticipated to avoid major setbacks. The currently available sophisticated aeronautical systems do not preclude future pilots from learning the basics of flying.

The time spent by doctors on mastering and using these new technologies should also be considered, having in mind the current medical demographics. Currently, physicians spend on average twice as much time consulting and filling in information on computer systems as they do on directly interacting with their patients. This unanticipated development is often overlooked and constitutes a significant cause of burnout because these additional tasks are often performed after a day of consultation. AI systems should allow automation of some of these tasks, such as recording of clinical information and automated coding of acts.

Finally these systems open up new types of problems regarding data property and the necessary authorisations required, raising issues regarding human rights and individual freedom. The development of poorly understood or non-mastered algorithms could become an important ethical problem. The need of a legal framework has only recently been considered at the French and European levels, but its principles and its application remain to be developed. The subject is economically and socially important, and it is therefore likely that the development of AI systems will be at the heart of international competition, with different groups of humans with different values, and this might be a difficult aspect to manage.

To enable the harmonious development of these technologies and put them at the service of medical progress and our patients, we will need to expand our knowledge and master these challenges. The present book is preparing us for these challenges.

REFERENCES

1. *Panorama des startups santé françaises utilisant l'IA*. BPI, 2020. https://lehub.bpifrance.fr/panorama-startups-sante-francaises-ia/
2. *La Plateforme des données de santé (Health Data Hub)*. Cnil, 2021. https://www.cnil.fr/fr/la-plateforme-des-donnees-de-sante-health-data-hub
3. *[Grand Défi] Santé: améliorer les diagnostics médicaux à l'aide de l'intelligence artificielle*. Gouvernement français, 2021. https://www.gouvernement.fr/grand-defi-sante-ameliorer-les-diagnostics-medicaux-a-l-aide-de-l-intelligence-artificielle

What Professor Coron thinks of the CAD EYE system

ADR is a quality indicator. How do you see that CAD EYE in combination with LCI can help to improve ADR rates and benefits the patient?

Detection of precancerous lesions is key to prevent colorectal cancer. It is now well established that ADR reflects the quality of the procedure and the expertise of the physician. However, there is a wide variability among physicians in terms of both endoscopic technique and motivation. This results in major differences in the use of ADR and the detection of interval cancers. One major interest of artificial intelligence is that it will help in the standardization of the diagnostic performances of endoscopists, and optimize the detection and treatment of precancerous lesions worldwide. LCI has been shown to help detect small flat lesions, in particular in the right colon. CAD EYE is a fantastic tool to detect subtle lesions that can be missed even by skilled endoscopists. For instance, a recent randomized study has shown that lesions that are small in size, isochromatic, flat in shape, with an unclear boundary, partly concealed behind colon folds or on the edge of the visual field, benefit from computer-aided detection (doi: 10.1016/S2468-1253(20)30051-0). This is also what we see in our clinical experience with CAD EYE. Of course, CAD EYE can only detect what is in the field of view, which means that mucosal exposure is also a key parameter. Endoscopic withdrawal technique must also be taught in combination with the use of LCI and CAD EYE.

Professor Coron, a leading figure of European endoscopy

What is the importance for you of having the complete package: CAD EYE detection & CAD EYE characterization in one device?

CAD EYE detection combined with CAD EYE characterization was the first « all in one » device developed by a company, and therefore represents a major step in the history of endoscopy. As early users, I must say that we were amazed by the real-time efficacy of the system. Within milliseconds, the system is able to predict lesion histology, which is of paramount importance to guide the decision-making of the physician during colonoscopy. The study published in *Endoscopy* in 2021 by Weight *et al.* (doi: 10.1055/a-1372-0419) showed that the high accuracy of CAD EYE detection & characterization is similar to that of an expert. Of course, we need more data from multicenter prospective studies to confirm these results, as well as cost-efficacy studies but it is more than likely that upgrading detection and removal of precancerous lesions by endoscopy will decrease the burden of colorectal cancer on a large scale. These first results will also help non expert physicians to adopt AI as a tool to provide the best screening modality for their patients in daily practice. Furthermore, it is likely that the cost of AI systems will dramatically decrease overtime as more and more systems will equip endoscopy units.

·How do you see the benefit of CAD EYE to support young/starting endoscopists?

Training young endoscopists to detect and characterize colonic lesions requires time and motivation from experts. It is one of the most challenging procedures since the risk of missing a potential lesion (and seeing an interval cancer occur years later) is an important responsibility for the trainee. In addition, because colonoscopy is the most frequently performed examination under sedation or general anesthesia in most centers, it is impossible to have an expert assisting the trainee at every procedure. The reality is that, at least in France, the availability and the level of expertise of the supervisor is very variable according to the workload in the different endoscopy rooms. Therefore, some procedures are performed by trainees without the oversight of the expert focused on their procedure all the time, and real-time decision in a short period of time makes it unethical and unrealistic for the expert to just review the video afterwards and schedule another colonoscopy if a lesion was missed…Therefore, having CAD EYE detection and characterization is like offering the trainee a 'virtual expert' for every colonoscopy, and it also dramatically shortens their learning curve. My advice is that, beyond the use of CAD EYE, and according to the ESGE guidelines, physicians (trainees or experts) continue documenting lesions with high quality images during colosconoscopy. The aim is to review these images with the final pathology in order to progress in optical diagnosis and not rely only on AI as a 'simple technician' would do, because the development of their human expertise is finally what matters the most, with patient's optimal management. This is very easy to do with novel softwares that provide access to the whole integrated medical files of the patient from our offices, hospitals, and even from home!

·Do you think that the simple user interface of CAD EYE is beneficial for physicians and easy in use, not interfering with the endoscopic image? How important is that for you?

As previously said, it is amazing to see how CAD EYE is easy to use, and provides instantaneous detection and characterization. Once the system is set up on the tower, it is immediately adopted by the majority of users. Of course, there still might be a minority of 'refractory people' but I'm quite confident that it will finally equip the majority of units. It does not interfere with the endoscopic images provided the colon is sufficiently inflated. The only disadvantage is that it continues to highlight the lesion during the interventional part, i.e. endoscopic resection. Therefore, I prefer to switch it off as soon as the lesion has been detected and characterized, to better focus on the different steps of endoscopic resection, then switch it on again for the rest of withdrawal. It will probably continue to progress, with specific characterization of sessile serrated adenomas/polyps (SSAPs) and superficial cancers, and the future looks bright!

THE EYE, THE MACHINE, THE MIND : COVER ILLUSTRATIONS

Duck of Vaucanson, 1734

Laws of perspective, Abraham Bosse

« The Golem », israelian computer 1965

Robot from Metropolis , 1927

John MacCarthy

Match Kasparov/ Deep Blue, 1996

Probabilities B. Pascal

Pespective, Abraham Bosse

« Typographie raisonnée », Geoffroy Tory, XVIe siecle

Thomas Bayes

The « Turk », (false automatic, player XVIIIe siecle)

Bristol's Turtle (cybernetics), 1947

Cartesian Dioptrics

Jacquard Machine, 1801

Serendip (cf notion of serendipity)

Boolean Logic, Elements

Marvin Minsky

Analytical machine (Babbage) XIXe siècle

Tree of knowledge, R. Lulle, XIVe siecle

This book is the result of an initiative
by the Nantes Institute of Diseases of the Digestive System
with the help and the support of FUJIFILM Europe.

It is published by Le Grand Métier editions and can be cited as follows:

Artificial intelligence and digestive diseases.
Ed: Coron E, Vanbiervliet G. LGM: Rouen, 2021.

It is the english version of :

Intelligence artificielle et maladies digestives.
Ed : Coron E, Vanbiervliet G. LGM : Rouen, 2021.

L'Institut des
Maladies de l'Appareil Digestif

FUJIFILM
Value from Innovation