Artificial intelligence in thoracic surgery: past, present, perspective and limits

Harry Etienne, Sarah Hamdi, Marielle Le Roux, Juliette Camuset, Theresa Khalife-Hocquemiller, Mihaela Giol, Denis Debrosse, Jalal Assouad

Artificial intelligence in thoracic surgery: past, present, perspective and limits

Harry Etienne[1,2], Sarah Hamdi[3], Marielle Le Roux[1], Juliette Camuset[1],
Theresa Khalife-Hocquemiller[1], Mihaela Giol[1], Denis Debrosse[1] and
Jalal Assouad[1,2]

Affiliations: [1]AP-HP, Dept of Thoracic and Vascular Surgery, Tenon University Hospital, Paris, France.
[2]Sorbonne Université, INSERM, UMRS1158 Neurophysiologie Respiratoire Expérimentale et Clinique, Paris,
France. [3]Dept of Thoracic and Vascular Surgery, Le Raincy-Montfermeil Hospital, Montfermeil, France.

Correspondence: Harry Etienne, AP-HP, Dept of Thoracic and Vascular Surgery, Tenon University Hospital, 4
Rue de la Chine, 75020 Paris, France. E-mail: h.etienne@hotmail.fr

🐦 @ERSpublications
**Surgeons should be engaged in assessing quality and applicability of AI advances to ensure
appropriate translation to their clinical practice. This is fundamental as governments are increasing
their funding and adapting their legislation to promote it.** http://bit.ly/2V1yNB3

ABSTRACT   Artificial intelligence (AI) technology is becoming prevalent in many areas of everyday life.
The healthcare industry is concerned by it even though its widespread use is still limited. Thoracic
surgeons should be aware of the new opportunities that could affect their daily practice, by direct use of AI
technology or indirect use *via* related medical fields (radiology, pathology and respiratory medicine). The
objective of this article is to review applications of AI related to thoracic surgery and discuss the limits of
its application in the European Union. Key aspects of AI will be developed through clinical pathways,
beginning with diagnostics for lung cancer, a prognostic-aided programme for decision making, then
robotic surgery, and finishing with the limitations of AI, the legal and ethical issues relevant to medicine.
It is important for physicians and surgeons to have a basic knowledge of AI to understand how it impacts
healthcare, and to consider ways in which they may interact with this technology. Indeed, synergy across
related medical specialties and synergistic relationships between machines and surgeons will likely
accelerate the capabilities of AI in augmenting surgical care.

Introduction

Artificial intelligence (AI) is a suite of technologies that uses adaptive predictive power, autonomous learning
and complex algorithms. AI is able to [1]: recognise patterns; anticipate future events; make good decisions;
learn from its mistakes; assist clinical decision-making; and uncover relevant information from data.

Analysis by the Organisation for Economic Co-operation and Development found that, in the first half of
2018, AI start-ups attracted around 12% of all private equity investments worldwide, a steep increase from
just 3% in 2011 [2].

The technology is now prevalent in many areas of everyday life. Google Maps uses AI to dynamically learn
traffic patterns and create efficient routes; smartphones use AI to recognise faces and verbal commands.
Figure 1 highlights key dates in the history of AI. The healthcare industry makes up only a small part of AI
usage; although AI technologies have advanced rapidly in many fields, their implementation in patient care

settings has yet to become widespread. As government and investors are increasing budget funds for development of AI [4], the future of AI seems bright and full of innovative perspectives.

Thoracic surgeons are affected by recent advances in AI technologies and should consider the new possibilities that could impact their daily practice. AI has shown promising results in thoracic surgery, and related fields such as radiology, pathology or respiratory medicine, which affect patient management from the pre-operative period, to surgery and follow-up. The objective of this article is to review the applications of AI to thoracic surgery, highlight the outlook in robotic surgery, and discuss the limits, ethical and legislative issues of widespread application of AI in thoracic surgery, in the European Union (EU).

This article is structured as a clinical pathway which starts with lung cancer diagnosis, then moves on to prognostic-aided software for decision making and AI in robotic surgery. We finish with the limits and legislative issues related to AI in surgery.

Methods

Pubmed Cochrane databases were searched up to December 2019 for relevant articles for systematic research. Articles were screened by reviewing their abstracts for the following criteria. 1) Topics: AI and lung nodules management, AI and respiratory medicine, AI and pathology, AI and thoracic surgery. 2) Published in English and in a peer-reviewed journal.

Articles of interest that had been cited by the articles identified in the initial search were also reviewed. The *Official Journal of the European Union* was used as the main source for the legislative aspects of AI.

Diagnostic for lung cancer

Table 1 lists basic terminology used in AI. AI has been used in diagnosis of lung nodules for decades [5–8]. Computer-aided diagnostics (CAD) was developed as a tool to assist radiologists detect lesions on chest radiographs [9, 10]. EDWARDS *et al.* [11] used AI to predict if pulmonary nodules were benign or
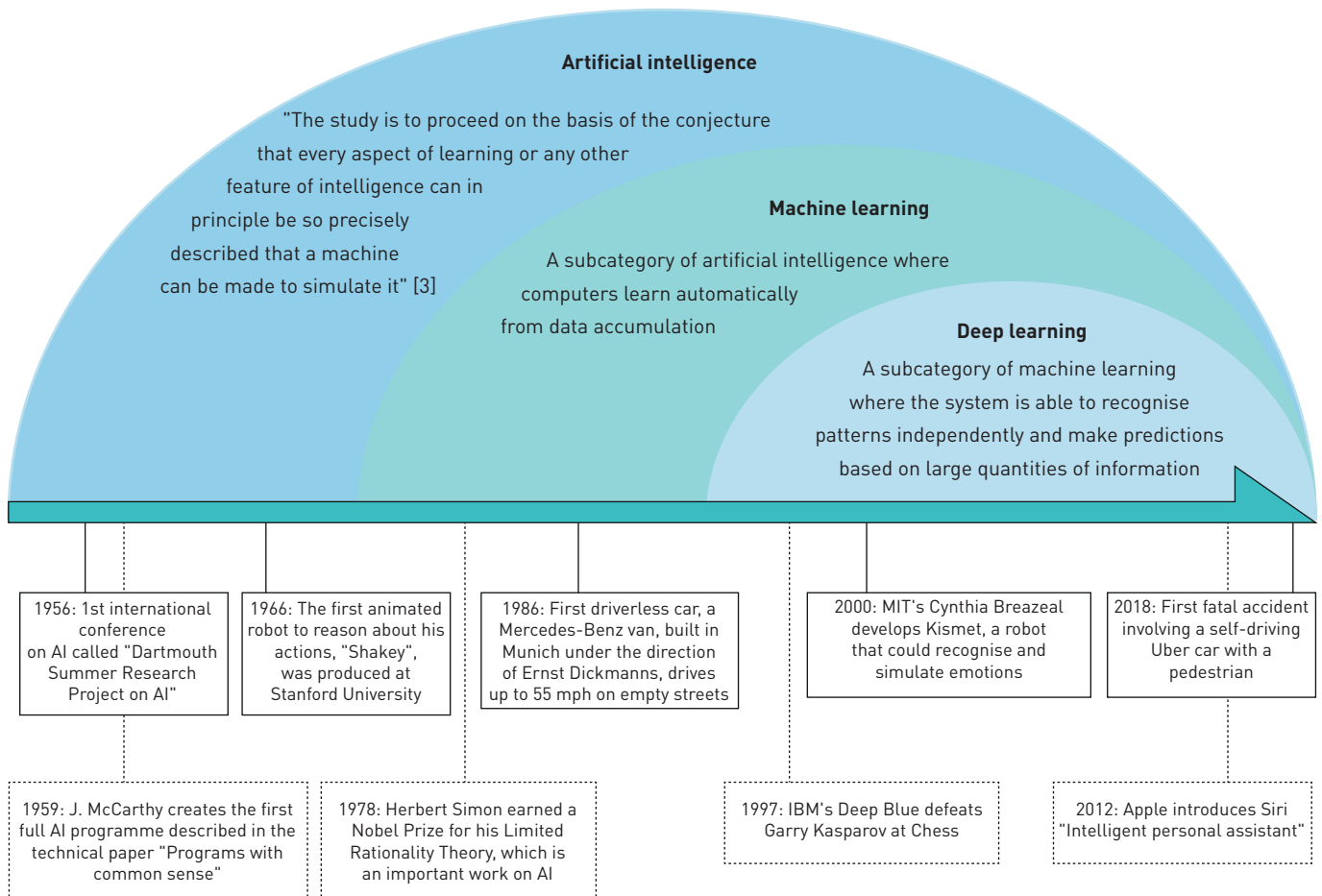

FIGURE 1 Brief history of artificial intelligence (AI).

malignant. To build the algorithm for their AI, they used clinical and radiological data from 165 consecutive patients who had had surgery in their centre. They then tested the algorithm on 100 consecutive patients who underwent surgery for a suspicious lung nodule, without a pre-operative diagnosis. The model correctly categorised the lesions as benign or malignant in 95 of the 100 cases, yielding a 95% accuracy. 16 of the 18 benign lesions and 79 of the 82 malignant lesions were correctly assigned. This yielded a sensitivity of 96% and a specificity of 89%.

CAD was subsequently used for similar purposes along with computed tomography (CT), magnetic resonance imaging and nuclear medicine. CAD uses algorithms based on machine learning technologies to provide second opinions to assist radiologists [12]. Although AI can be used to assess images from various regions of the human body, its use to detect lung nodules and differentiate malignant from benign nodules seems to be a key strength. The introduction of the computed neural network in 2012 was a big change in the used of AI in radiology [13]. Krizhevsky *et al.* [13] designed an artificial neural network for large-scale object detection and image classification based on deep learning for the ImageNet competition in order to identify and classify 1.2 million high-resolution images to a previously unseen level of accuracy. The authors used a neural network of eight layers, with 60 million parmeters and 650 000 neurons. There were 1.2 million training images, 50 000 validation images and 150 000 testing images. The Top-5 error (the fraction of test images for which the correct label is not among the five labels considered most probable

TABLE 1 Keywords in artificial intelligence

Big data
Data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges which makes it difficult to process it using on-hand data management tools or traditional data processing applications. A multitude of technology platforms may be involved in the generation or collection of this data, including network servers, digital imaging and communications in medicine (DICOM images) stored in picture archiving and communication system (PACS), electronic health records (EHRs), genomic data, personal computers, smart phones and mobile applications, and wearable devices and sensors. Artificial intelligence algorithms present the ability to search through such databases and extract the necessary data for specific task.

Algorithm
A set of rules provided to artificial intelligence that allows the machine to perform certain tasks, such as classification.

Machine learning
A field of artificial intelligence where a system is taught to interpret various features related to an objective and to flag it when it comes up. It is also trained to make predictions by recognising patterns. Automatic learning enables the machine to improve without continuous support and guidance by employing algorithms based on comparison logic, research and mathematical probability.

Classifier
A machine learning algorithm that sorts data into predefined categories. It can be:
Supervised which means it uses data that has been identified by the programmers to generate predictive models to identify unlabelled data. Often, the algorithm attempts to mimic the ability of a highly trained individual with domain expertise.
Unsupervised which requires no data labelling. An output or a target is not defined. The computer looks for patterns or grouping in the data. This approach may be useful in facilitating analysis of big data.
Reinforcement learning: the inputs and the outputs pairs are not specified, and the focus of the machine is to perform a task while learning from its own successes and mistakes to improve its performance.

Neural network
Neural networks mimic human brain function with each neuron performing its own simple calculation and the network formed by all of the neurons multiplying the potential of these calculations. This technology is particularly useful in predictive analysis, image recognition and speech processing.

Deep learning
A subcategory of machine learning where the system is made of digitised inputs, such as an image or speech, which go through multiple layers (ranging from 5 to 1000) of connected ˝neurons˝ (neural network), that progressively detect different features, and ultimately provides an output. The results from the first layer of neurons serve as a point of departure for calculating subsequent results. The connected neurons constitute the neural network with autodidactic quality. The system recognises patterns independently and makes predictions on a large quantity of information.

Natural language processing
A sub field of artificial intelligence dealing with the way to program computers to process and analyse large amounts of natural language data. Challenges are set by speech recognition, natural language understanding and natural language generation. It employs a sequence of mathematical processes and comparisons to identify the user's input, possibly correcting certain errors or applying synonyms, in order to identify all the information necessary to understand the needs of the user.

by the model) was 15.2%. Deep Learning sparked interest for a broader use of AI, especially in radiology [14]. This technique was used by NAM *et al.* [15] to detect malignant pulmonary nodules on chest radiographs. The deep learning-based algorithm outperformed physicians in radiograph classification and nodule detection performance. But, most importantly, it enhanced physician performance when used as a second reader.

In 2011, the National Lung Screening Trial (NLST) was conducted to determine if screening with low-dose CT could reduce mortality from lung cancer compared with chest radiography, in patients aged between 55 and 74 years, with a history of cigarette smoking of at least 30 pack-years or former smokers who had quit within the previous 15 years [16]. It found a relative reduction in mortality from lung cancer with low-dose CT screening of 20% (95% CI 6.8–26.7; p=0.004). Large clinical screening for early detection of lung cancer by low-dose CT scan (NLST trial and the NELSON trial) provided a large data set to explore using AI [16, 17]. ARDILA *et al.* [18] took advantage of this and developed a deep convolutional neural network to do a full assessment of low-dose CT volume, focusing on regions of concern, with comparison to prior imaging when available and calibration against biopsy-confirmed outcomes. To do so, they used a NLST dataset consisting of 42 290 CT cases from 14 851 patients, 578 of whom developed biopsy-confirmed cancer within the 1 year follow-up period; 70% of the data was used to train the algorithm and 15% was used as a cross-validation set. The final 15% was used to test the algorithm against radiologists. When prior CT imaging was not available, the model outperformed the radiologists with absolute reductions of 11% in false positives and 5% in false negatives. When prior CT imaging was available, the model performance was on par with the radiologists. This could lead to future approaches to lung cancer screening as well as to support assisted- or second-read workflows. Because detecting nodules or masses in chest radiographs or CT by radiologists can be a hard task when numerous examinations have to be interpreted, AI can serve as a diagnostic aid for suspicious nodules. By automating simple tasks, physicians are freed to tackle more complex tasks. This represents an improvement in the use of human capital; clinicians and AI cooperate in a synergistic way.

Prognostic-aided programme for decision-making

Lobectomies for primary nonsmall cell lung cancer have a mortality rate of 2.3% compared with pneumonectomies where it increased to 6.8% in specialised centres with high volume activity [19]. It is important to correctly evaluate patients who are candidates for surgery in order to assess their risks. The 2009 European Society of Thoracic Surgeons/European Respiratory Society guidelines and the American Thoracic Society have set recommendations for the necessary exams needed for thorough cardiac and respiratory assessments [20]; however, these guidelines do not take into consideration other comorbidities that might impact post-operative recovery such as past medical history, diabetes, smoking history, *etc*.

AI has shown promising results as an aid to decision making for thoracic surgery patients by assessing their surgical risk and individually evaluating their prognoses. Prognosis plays an important role in patient management in surgery because it implies not only the usefulness of the surgical act planned and proposed to the patient, but also the impact it will have on the patient's short- and long-term outcomes. Risk indexes classify surgical risk by identifying groups with statistical probability outcomes [21, 22]. These are based on large multivariate statistical analyses designed to compare population groups and the relative contributions of different risk factors to specific outcomes. With AI, the objective is to evaluate an individual risk factor with precision and to adapt decision making individually and not by group risk. AI personalises patient management for each individual, thus increasing accuracy because the machine is more flexible and can learn by adaptation [23].

SANTOS-GARCIA *et al.* [24] evaluated the performance of an artificial neural network ensemble to predict cardio-respiratory morbidity after pulmonary resection for nonsmall cell lung cancer (NSCLC). They trained their AI with 348 cases and validated their model on 141 cases. Sensitivity and specificity for predictions were 0.67 (95% CI 0.49–0.79) and 1.00 (95% CI 0.97–1.00), respectively. The area under the receiver operating characteristic was 0.98. Similarly, ESTEVA *et al.* [25] used four different probabilistic artificial neural network models, trained by data from 113 cases of NSCLC. Three of the models were programmed to estimate post-operative prognosis (dead/alive) with the following information: Model 1: using 96 clinical and laboratory variables; Model 2: using only clinical and laboratory variables related to the risk scores of GOLDMAN *et al.* [21] and TORRINGTON *et al.* [22]; Model 3: using all variables except for the ones in the risk scores of GOLDMAN *et al.* [21] and TORRINGTON *et al.* [22]. A fourth model was programmed with all variables to estimate major post-operative complications.

All four models were tested on 28 patients. Models 1 and 3 were able to correctly classify all 28 test cases whereas model 2 made six errors in classification from the 28 test cases. Model 4 estimated major post-operative complications correctly in all 28 test cases.

Other medical specialists with whom thoracic surgeons collaborate regularly have taken an interest in AI and are aware of its potential to change daily practice. Yu *et al.* [26] published an article in which AI software distinguished between primary lung adenocarcinoma and squamous cell carcinoma using quantitative histopathology features extracted from 2186 haematoxylin and eosin whole-slide pathology images from the Cancer Genome Atlas. The average area under the receiver operating characteristic was 0.72. The authors used seven different classifiers in all; performance did not differ significantly between these classifiers. The secondary objective of the study was to detect image features which would help to predict stage I adenocarcinoma survival. However, this was limited by the fact that only haematoxylin and eosin stains were used; no immune-histochemical stains or molecular biology were used for adenocarcinomas. Moreover, different adenocarcinoma subtypes were not identified. Similarly, Coudray *et al.* [27] used a convolutional neural network to distinguish between adenocarcinomas and squamous cell carcinomas; the performance was comparable to pathologists with an area under the curve of 0.97. Furthermore, they attempted to predict the most commonly mutated genes in adenocarcinomas using the same software [27].

Topalovic *et al.* [28] showed that AI outperformed pulmonologists in the interpretation of pulmonary function tests (PFT); 120 pulmonologists from 16 hospitals in five European countries participated in this multicentre noninterventional study. Pattern recognition of PFTs by pulmonologists matched the guidelines in 74.4 ± 5.9% of cases and the pulmonologists made a correct diagnoses in 44.6 ± 8.7% of cases, whereas AI perfectly matched the PFT pattern interpretations (100%) and assigned a correct diagnosis in 82% of all cases. The AI-based software's performance was superior compared with the performance of pulmonologists. AI commands a certain level of fascination in the medical world because of the potential of such technology to improve clinical practice. To gain broader acceptance, the challenge for AI is to answer complex clinical questions taking into account, for example, patients' comorbidities.

AI in robotic surgery

Minimally invasive surgery decreases the surgery-related trauma by using smaller incisions compared with conventional surgical approaches [29]. Video-assisted thoracoscopic surgery has been shown to decrease length of hospital stays and post-operative complications in thoracic surgeries for major lung resections [30]. Robotic-assisted surgery, with the introduction of the da Vinci Surgical System from Intuitive Surgery (Sunnyvale, CA, USA), offered new possibilities by allowing the surgeon to have seven degrees of liberty in wrist movement inside the thorax and high definition three-dimensional vision. The system aids surgeons by increasing precision at each step of surgery and eliminating any tremors during the intervention. While the general term "robotic surgery" is often applied to the da Vinci Surgical System, it is not robotic surgery in the strict sense of the term since it cannot be programmed, nor can it make decisions on its own to move or perform any type of surgical manoeuvre [31]. Robotic-assisted surgery is then by definition not AI, although the term is often mistakenly used in this context. The da Vinci Surgical System offers lots of intelligent solutions for clinical practice, but it is not autonomous robotic surgery.

The robot is a tool set to augment surgeons' capabilities and is not intended as a replacement surgeon. The surgeon has control of the robot's every move; the system mimics the surgeon's hand movements in real time. In case of any difficulty that might put the patient at risk, the surgeon is able to convert and switch to a conventional incision. But with the advent of AI, in the future it seems possible that the robotic system could fully carry out the intervention by itself and adapt to different scenarios. Indeed, in 2016, Shademan *et al.* [32] compared metrics of intestine anastomosis (consistency of the suturing informed by the average suture spacing, the pressure at which the anastomosis leaked, the number of mistakes that required removing the needle from the tissue, completion time and lumen reduction in intestinal anastomoses) between a supervised autonomous system (Smart Tissue Autonomous Robot (STAR)), manual laparoscopic surgery and clinically used robotic-assisted surgery. The supervised autonomy with STAR surpassed the performance of robotic-assisted surgery, laparoscopic approach and manual surgery, based on some metrics,. This study demonstrated that autonomous surgery is feasible. The interactive adaptative decision making and execution between the surgeon and the STAR system pave the way for interesting new and innovative possibilities to augment surgeons. The STAR system proved the feasibility of performing suture in pigs but, despite further research, so far there has been no application for Food and Drug Administration approval.

Limits of AI

AI shows promising results in many fields of medicine and has the potential to drastically change daily practice. However, we must be critical of the various applications proposed for it. New software must be thoroughly evaluated, preferably in randomised controlled, trials to highlight its benefits. A programme developed in one institution may not be directly applicable to other institutions. The way in which the algorithm is developed will effectively tailor the model to reflect the clinical experience of that institution. This is both an advantage and a disadvantage to potential users [11].

The quality of the answers by the AI largely depends on the quality and the amount of data available. The generalisability of AI algorithms across subgroups is critically dependent on factors such as representativeness of included populations, missing data and outliers. Indeed, it is important to continuously update and provide new patient data to algorithms so the decision making can be adapted. Statistical bias may be inherent due to suboptimal sampling, measurement error in predictor variables and heterogeneity of effects [33]. This outlines the importance of transparency to enable the technology to be as accurate as possible and to avoid potential bias [34]. It is important for surgeons to understand how the machine comes to a decision in order to review them [35]: to do so, it is vital to take interest in this field and collaborate with technicians, informaticians and engineers who are involved.

The outputs of the algorithm are limited by the types and accuracy of available data; for example, lung cancers that affect Caucasians in the EU do not present the same epidemiologic characteristics as lung cancers in people living in Asia. In Asia, a higher proportion of cancers are attributable to nonsmoking causes, particularly in women [36]. Therefore, under-representation of certain population or sexes presents a risk of selection bias which will impact predictions.

Last, but not least, we must overview whatever decisions the AI software may suggest. IBM's Watson for Oncology Cognitive Computing system, created in 2012, uses AI algorithms to generate treatment recommendations for various cancers including lung cancers. The software was trained by oncologists at the Memorial Sloan Kettering Cancer Center (New York, NY, USA) to learn key data associated with a patient's cancer: results from blood tests; pathology and imaging reports; and presence of genetic mutations. The treatment options are generaly consistent with the National Comprehensive Cancer Network guidelines and the algorithm therapeutic recommandations are classified in three categories: recommended, for consideration and not recommended.

It is used in many hospitals around the world. However, in 2018, IBM's Watson came under fire for suggesting treatments in some cases that were erroneous and which might even have put patients' lives at risk [37]. The system suggested that a patient with a lung squamous cell carcinoma should take bevacizumab, which is strictly contraindicated. An abstract on IBM's Watson for Oncology Cognitive Computing system was presented at the American Society for Clinical Oncology in June 2019, suggesting it may help in decision making during multidisciplinary tumour board [38]. Zhou et al. [39] in a retrospective study in China, evaluated the concordance between the treatment recommandation proposed by IBM's system and clinical decisions by oncologists: concordance for lung cancer was slightly above 80% whereas it was only 12% for gastric cancer. Differences in concordance did not reflect the accuracy of the AI nor the physicians'abilities but rather the fact that the same drugs are not necessarily approved and commercialized in both China and the USA, respoonsible for the differences in cancer protocols.

Moreover, surgical decision-making involves complex, high-stake analysis of a patient's health chart along with intuition based on experience, while taking into consideration past medical history, modifiable risk factors, potential complications, prognosis, and the patient's emotional experience [40]. Each step of decision-making introduces variability and opportunities for errors, especially in an acute surgical condition. Electronic health records are a source of big data that are continuously updated by surgeons, nurses, nutritionists and physical therapists. Safavi et al. [41] used electronic health records to develop a machine learning model to aid discharge processes for inpatient surgical care, including for thoracic surgery patients. The estimated out-of-sample area under the receiver operating characteristic curve of the model was 0.84 (sd 0.008; 95% CI 0.839–0.844). The neural network model had a sensitivity of 56.6%, a specificity of 82.6%, a positive predicted value of 51.7% and a negative predicted value of 85.2%. Overall, 838 (27.8%) patients were discharged later than the neural network predicted because of clinical barriers, variation in clinical practice and no clinical reason. As William Osler said, "Medicine is an art of uncertainty and an art of probability" [42].

Topol et al. [43] insists that the validation of the performance of an algorithm in terms of its accuracy is not equivalent to demonstrating clinical efficacy. Keane and Topol [44] refer to it as the "AI chasm"; that is, an algorithm with an area under curve of 0.99 is not worth very much if it is not proven to improve clinical outcomes. This is probably the reason why clinical application of AI technologies is not more developed.

Figure 2 provides an overview of the potential applications of AI in thoracic surgery as previously described through the clinical pathway.

Ethical and legislative issues to implementation

AI offers a wide range of possibilities that have already started to change daily practices in some specialties, but AI could also pose significant risks in terms of inappropriate or inaccurate risk assessment, treatment recommendations, diagnostic error or privacy breaches [45]. In 2019, the National Academy of Medicine in the USA published a reference document on the responsible development, implementation and maintenance
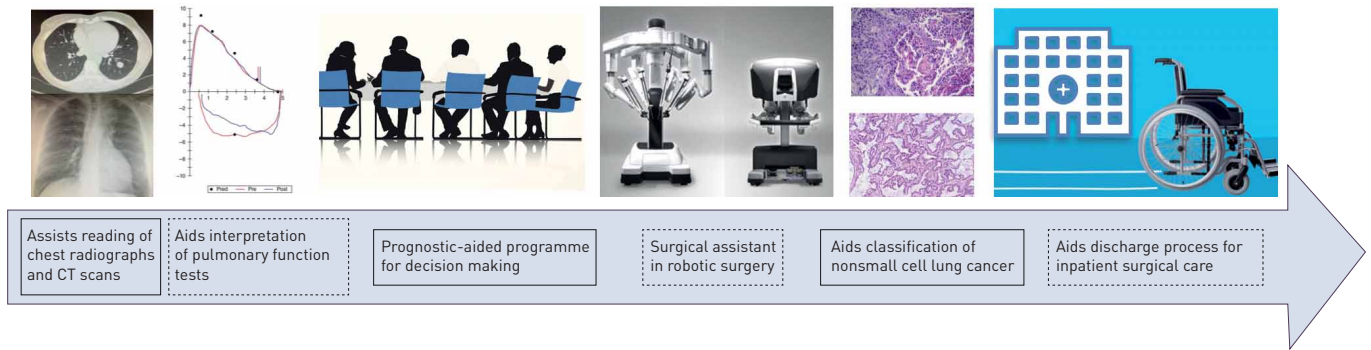
FIGURE 2 Potential applications of artificial intelligence in thoracic surgery in a clinical pathway. CT: computed tomography. The third image is courtesy of https://pngimage.net/. The fourth image represents the Da Vinci Xi surgical robot system (patient's cart and surgeon's console) from Intuitive Surgical.

of AI in the clinical enterprise [46]. The International Medical Device Regulators Forum [47], whose members include Australia, Brazil, Canada, China, EU, Japan, Russia, Singapore and the USA, is a group of medical device regulators working toward harmonising medical device regulation. They have also defined software using AI as a medical device. In the past couple of years, the US Food and Drug Administration approved a number of algorithms using AI for clinical use; for example, in April 2018, IDx-DR, an AI algorithm designed for screening diabetic retinopathy patients using retinal images prior to referring patients to ophthalmologists [48]. Other algorithms, mainly in radiology, have also been approved. None of the approved algorithms are directly applicable to thoracic surgery as yet. These software have been approved in the USA, mainly through two procedures [49]. 1) De novo process [50, 51]: authorises for marketing clearance novel medical devices for which general controls alone, or general special controls, provide reasonable assurance of safety and effectiveness for the intended use, but for which there is no legally marketed predicate device. It opens the path for future medical devices, class I or II, to get Food and Drug Administration clearance through the 510 k process. Device classification is based on the risk it poses to the patient and the intended use. 2) 510 k process clearance (analogous to the generic drug concept): marketing clearance for a medical device is obtained by stipulating that it is substantially equivalent in safety and effectiveness to another lawfully marketed device or to a standard recognised by the Food and Drug Aministration when used for the same intended purpose.

None of the approved medical devices have been through the pre-market approval process, which is based on demonstration of safety and effectiveness through clinical trials, preferably randomised and controlled. This procedure is mostly reserved for class III medical devices, which designate the device with the highest risk for the patients.

In the EU, AI technologies considered as medical devices are under regulation by the General Data Protection and Regulation adopted by the EU parliament in 2016 [52–54]: article 13 stipulates that the citizen has the right to explicit and informed consent which means a citizen is entitle to receive a full explanation on how the algorithm comes up with a decision. This has to be done before any collection of the patient's data. The patient also has the right to track what data is being collected and be able to request removal of their data. This legislation became enforced on 25 May 2018. Processing of data is not allowed unless one of the following conditions has been fulfilled: the data subject gave consent; the data is necessary for archiving purposes in the public interest, scientific or historical research purposes, or statistical purposes.

Moreover, Article 25 requires "Data protection by design and by default" from all data controllers, regardless of their size. The controllers must be able to demonstrate that they have implemented dedicated measures to protect these principles and that they have integrated specific state-of-the-art safeguards that are necessary to secure rights and freedoms of data subjects [55]. Therefore, medical professionals using this software must consequently have an understanding of the mechanisms underlying their function [56, 57].

Conscious of the potential of AI, the French government has taken a set of measures necessary to promote and broaden its implementation. As medical devices implicated in diagnostic or therapeutic approaches, AI algorithms are listed as medical devices which necessitate an evaluation by the National Commission for Evaluating Medical Devices and Health Technologies (CNEDiMTS) in order for them to be reimbursed by social security. CNEDiMTS intervenes once the medical device has obtained a conformity assessment (CE marking) to demonstrate that it meets legal requirements to ensure it is safe and performs as intended. Under Regulation (EU) 2017/745 on Medical Devices, adopted in 2017, the European

TABLE 2 Criteria for public consultation analysis by CNEDiMTS on the use of artificial intelligence (AI) software in medicine

Final use
  1. Detail the benefit of the information that will be given by the system relying on AI
  2. Describe the characteristics of the population for whom it is intended

Description of the learning process for the AI
  3. Type of learning method
      Continuous (autonomous learning and adaptation)?
      Initial (algorithm is trained initially with no further updates)?
      Or incremental (algorithm updated through learning process)?
  4. Describe the model used
      Supervised?
      Semi-supervised?
      Unsupervised?
      By reinforcement?
  5. Describe the algorithm used
      Classification
      Regression
      Clustering
  6. Describe selection method of the model
  7. Describe the different steps of learning phase
  8. Describe the strategy for updating the algorithm
  9. Describe if necessary in which cases humans intervene in the learning process

Detail the data entered in the initial learning phase or data involved for the updates or autonomous learning process
Describe the characteristics of the sample from the targeted population, used to develop the model
  10. Detail the characteristics of the sample
  11. Detail the modalities for separating training data from testing data and validating data
  12. Justify representativeness of the sample chosen compared with the targeted population
Description of the variables
  13. Characteristics of the variables
        Type
        Distribution
  14. Origins of the variables and methods of acquisition
Detail handling of the data before their use for the learning phase
  15. Describe the statistical tests used
  16. Describe methods of transformation use for the data
  17. Describe handling of missing information
  18. Explain detection of erroneous or aberrant data and their handling

Detail entry data implicated in decision-making
  19. Origins of the variables and methods of acquisition
  20. Characteristics of the variables
        Type
        Distribution

Performance
  21. Describe and justify the method of measurement adopted for the performance
  22. Describe the potential impact of adjustments measures
  23. Characterise overfitting and underfitting
  24. Describe the methods to handle overfitting and underfitting

Validation
  25. Describe the methods of validation
  26. Report the performance of the algorithms on the data set

Resilience of the system
  27. Describe the mechanisms set up in order to understand model drift
  28. Detail the thresholds chosen
  29. Detail if there is a system to detect any anomaly in the entry data implicated in the decision
  30. Describe the potential impacts of those aberrant entry data
  31. Detail the measure set up in case of model drift
  32. Detail the situations susceptible to alter the system function

TABLE 2 Continued

Explainability/interpretability
  33. Does the algorithm benefit from a technique of explainability/interpretability for the patients and/or the physicians?
  34. Detail the elements of explainability available
  35. Identify the influential parameters
  36. Detail if the decision-making in the system follows guidelines when they exist

Medicines Agency, and also national competent authorities from EU state members, can conduct conformity assessments. Because of its ability to learn from its mistakes, AI software is different to other medical devices. This means that specific checklists are necessary to evaluate their added value for patients. CNEDiMTS issued eight criteria for a public consultation on evaluation of AI software (table 2), from 20 November 2019 to 15 January 2020 [56]. The results are not yet available.

The study by SHADEMAN *et al.* [32] compared the metrics of intestine anastomosis by an autonomous robot to accomplish sutures of an intestine compared to with the metrics for humans. Similar comparisons between the abilities of autonomous robots and humans to perform other surgical interventions carried out in the operating room will be made in the future. As with the automotive industry, the objective is not full automation with no potential for human back-up or high automation with human back-up in very limited conditions, but rather a collaboration combining functions that machines do best with those that are best suited for surgeons [43, 58]. Current EU legislation is not adapted for this. Under current law, AI or robots cannot be held responsible for damage caused to a patient; liability is assumed either by: the manufacturer, if it's a manufacturing defect (in the Directive on Liability for Defective Products and Product Safety); the operator, if it's a medical error; or the person responsible for the maintenance of the system if the damages result from its failure [59, 60].

Conscious of this, the EU have debated the scenarios in which a robot can take autonomous decisions: "the traditional rules would not make it possible to identify the party responsible for providing compensation and require that party to make good any damage it has caused". The solution would be to consider a specific legal status for robots as "electronic persons", responsible for making good any damage they may cause [60].

Conclusion

The scale to which AI may change approaches to screening, diagnosis and healthcare management is enormous as its full potential to analyse unstructured data from big data, learn from its mistakes and draw relevant conclusions which might be a valuable aid in decision-making becomes apparent. Thoracic surgeons should improve their knowledge in this field and understand the way AI technologies could impact their clinical practice and their collaboration with pulmonologists, pathologists and radiologists. Truly autonomous robotic surgery does not exist yet and no application seems possible in the near future. Nonetheless, taking interest in the capabilities of AI technology and its progression in related medical specialties gives us a lead to enhance patient management in a dedicated clinical pathway as described in this article. Surgeons should be key actors in assessing AI's relevance and validating its application in daily practice in order to augment surgical care. This will require a close partnership with AI engineers, AI programmers, AI developers, hospital administration and healthcare management leaders. This is fundamental as governments are increasing their funding and adapting their legislation to promote a safe and secure introduction of AI into our daily practice.

References

1   Jiang F, Jiang Y, Zhi H, *et al.* Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2017; 2: 230–243.
2   OECD. Private Equity Investment in Artificial Intelligence. https://www.oecd.org/going-digital/ai/private-equity-investment-in-artificial-intelligence.pdf Date last updated: July 2018; date last accessed: 20 June 2020.
3   McCarthy J, Minsky M, Rochester N, *et al.* A proposal for the Dartmouth Summer Research Project on artificial intelligence, August 31, 1955. *AI Magazine* 2006; 27: 12.
4   Niu JT W, Xu F, Zhou X, *et al.* Global research on artificial intelligence from 1990–2014: spatially-explicit bibliometric analysis. *ISPRS Int J Geo-Inf* 2016; 5: 66.
5   Rotte KH, Meiske W. Results of computer-aided diagnosis of peripheral bronchial carcinoma. *Radiology* 1977; 125: 583–586.

6   Rogers W, Ryack B, Moeller G. Computer-aided medical diagnosis: literature review. *Int J Biomed Comput* 1979; 10: 267–289.

7   Cummings SR, Lillington GA, Richard RJ. Estimating the probability of malignancy in solitary pulmonary nodules. A Bayesian approach. *Am Rev Respir Dis* 1986; 134: 449–452.

8   Edwards FH, Schaefer PS, Callahan S, *et al.* Bayesian statistical theory in the preoperative diagnosis of pulmonary lesions. *Chest* 1987; 92: 888–891.

9   Doi K, MacMahon H, Katsuragawa S, *et al.* Computer-aided diagnosis in radiology: potential and pitfalls. *Eur J Radiol* 1999; 31: 97–109.

10   Giger ML. Computerized analysis of images in the detection and diagnosis of breast cancer. *Semin Ultrasound CT MR* 2004; 25: 411–418.

11   Edwards FH, Schaefer PS, Cohen AJ, *et al.* Use of artificial intelligence for the preoperative diagnosis of pulmonary lesions. *Ann Thorac Surg* 1989; 48: 556–559.

12   Li Q, Li F, Suzuki K, *et al.* Computer-aided diagnosis in thoracic CT. *Semin Ultrasound CT MR* 2005; 26: 357–363.

13   Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Communications ACM* 2017; 60: 84–90.

14   Murphy A, Skalski M, Gaillard F. The utilisation of convolutional neural networks in detecting pulmonary nodules: a review. *Br J Radiol* 2018; 91: 20180028.

15   Nam JG, Park S, Hwang EJ, *et al.* Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. *Radiology* 2019; 290: 218–228.

16   National Lung Screening Trial Research Team, Aberle DR, Adams AM, *et al.* Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; 365: 395–409.

17   Horeweg N, van der Aalst CM, Thunnissen E, *et al.* Characteristics of lung cancers detected by computer tomography screening in the randomized NELSON trial. *Am J Respir Crit Care Med* 2013; 187: 848–854.

18   Ardila D, Kiraly AP, Bharadwaj S, *et al.* End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 2019; 25: 954–961.

19   Brunelli A, Salati M, Rocco G, *et al.* European risk models for morbidity (EuroLung1) and mortality (EuroLung2) to predict outcome following anatomic lung resections: an analysis from the European Society of Thoracic Surgeons database. *Eur J Cardiothorac Surg* 2017; 51: 490–497.

20   Brunelli A, Charloux A, Bolliger CT, *et al.* ERS/ESTS clinical guidelines on fitness for radical therapy in lung cancer patients (surgery and chemo-radiotherapy). *Eur Respir J* 2009; 34: 17–41.

21   Goldman L, Caldera DL, Nussbaum SR, *et al.* Multifactorial index of cardiac risk in noncardiac surgical procedures. *N Engl J Med* 1977; 297: 845–850.

22   Torrington KG, Henderson CJ. Perioperative respiratory therapy (PORT). A program of preoperative risk assessment and individualized postoperative care. *Chest* 1988; 93: 946–951.

23   Bellotti M, Elsner B, Paez De Lima A, *et al.* Neural networks as a prognostic tool for patients with non-small cell carcinoma of the lung. *Mod Pathol* 1997; 10: 1221–1227.

24   Santos-Garcia G, Varela G, Novoa N, *et al.* Prediction of postoperative morbidity after lung resection using an artificial neural network ensemble. *Artif Intell Med* 2004; 30: 61–69.

25   Esteva H, Marchevsky A, Nunez T, *et al.* Neural networks as a prognostic tool of surgical risk in lung resections. *Ann Thorac Surg* 2002; 73: 1576–1581.

26   Yu KH, Zhang C, Berry GJ, *et al.* Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun* 2016; 7: 12474.

27   Coudray N, Ocampo PS, Sakellaropoulos T, *et al.* Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* 2018; 24: 1559–1567.

28   Topalovic M, Das N, Janssens W. Artificial intelligence for pulmonary function test interpretation. *Eur Respir J* 2019; 53: 1801660.

29   Bendixen M, Jorgensen OD, Kronborg C, *et al.* Postoperative pain and quality of life after lobectomy *via* video-assisted thoracoscopic surgery or anterolateral thoracotomy for early stage lung cancer: a randomised controlled trial. *Lancet Oncol* 2016; 17: 836–844.

30   Falcoz PE, Puyraveau M, Thomas PA, *et al.* Video-assisted thoracoscopic surgery *versus* open lobectomy for primary non-small-cell lung cancer: a propensity-matched analysis of outcome from the European Society of Thoracic Surgeon database. *Eur J Cardiothorac Surg* 2016; 49: 602–609.

31   Hubbard FP. Sophisticated robots: balancing liability, regulation, and innovation. *Florida Law Review* 2014; 66: 1803–1872.

32   Shademan A, Decker RS, Opfermann JD, *et al.* Supervised autonomous robotic soft tissue surgery. *Sci Transl Med* 2016; 8: 337ra364.

33   Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. *JAMA* 2019; 322: 2377–2378.

34   He J, Baxter SL, Xu J, *et al.* The practical implementation of artificial intelligence technologies in medicine. *Nat Med* 2019; 25: 30–36.

35   Burrell J. How the machine "thinks": understanding opacity in machine learning algorithms. *Big Data & Society* 2016; 3: 2053951715622512.

36   Youlden DR, Cramb SM, Baade PD. The International Epidemiology of Lung Cancer: geographical distribution and secular trends. *J Thorac Oncol* 2008; 3: 819–831.

37   Ross C Swetlitz I. IBM's Watson supercomputer recommended "unsafe and incorrect" cancer treatments, internal documents show. www.statnews.com/wp-content/uploads/2018/09/IBMs-Watson-recommended-unsafe-and-incorrect-cancer-treatments-STAT.pdf. Date last updated: 25 July 2018. Date last accessed: 20 June 2020.

38   Somashekhar SP, Sepúlveda M-J, Shortliffe EH, *et al.* A prospective blinded study of 1 000 cases analyzing the role of artificial intelligence: Watson for oncology and change in decision making of a Multidisciplinary Tumor Board (MDT) from a tertiary care cancer center. *J Clin Oncol* 2019; 37: Suppl. 15, 6533.

39   Zhou N, Zhang C, Lv H. Concordance Study Between IBM Watson for Oncology and Clinical Practice for Patients With Cancer in China. *Oncologist* 2019; 24: 812–819.

40   Loftus TJ, Tighe PJ, Filiberto AC, *et al.* Artificial intelligence and surgical decision-making. *JAMA Surg* 2019; 10.1001.

41   Safavi KC, Khaniyev T, Copenhaver M, *et al.* Development and validation of a machine learning model to aid discharge processes for inpatient surgical care. *JAMA Netw Open* 2019; 2: e1917221.

42   Simpkin AL, Schwartzstein RM. Tolerating uncertainty – the next medical revolution? *N Engl J Med* 2016; 375: 1713–1715.

43   Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019; 25: 44–56.

44   Keane PA, Topol EJ. With an eye to AI and autonomous diagnosis. *NPJ Digit Med* 2018; 1: 40.

45   Matheny ME, Whicher D, Thadaney Israni S. Artificial intelligence in health care: a report from the National Academy of Medicine. *JAMA* 2019 in press [DOI: 10.1001/jama.2019.21579]

46   Matheny ME, Israni ST, Ahmed M, *et al.* Artificial intelligence in healthcare: the hope, the hype, the promise, the peril. https://nam.edu/artificial-intelligence-special-publication Date last updated: 2020. Date last accessed: 20 June 2020.

47   International Medical Device Regulators Forum. www.imdrf.org Date last accessed: 20 June 2020.

48   US Food and Drug Administration. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. https://wayback.archive-it.org/7993/20190423071557/https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm604357.htm Date last updated: 11 April 2018. Date last accessed: 20 June 2020.

49   Hwang TJ, Kesselheim AS, Vokinger KN. Lifecycle regulation of artificial intelligence and machined learning-based software devices in medicine. *JAMA* 2019; 322: 2285–2286.

50   US Food and Drug Administration. De Novo Classification Request. www.fda.gov/medical-devices/premarket-submissions/de-novo-classification-request Date last updated: 20 November 2019. Date last accessed: 20 June 2020.

51   Publications Office of the EU. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Off J Eur Union* 2016; L1191.

52   Goodman B, Flaxman S. European Union regulations on algorithmic decision-making and a "Right to Explanation". *AI Magazine* 2017; 38: 50–57.

53   Wachter S, Mittelstadt B, Floridi L. Transparent, explainable, and accountable AI for robotics. *Science Robotics* 2017; 2: eaan6080.

54   European Data Protection Board. Guidelines 4/2019 on Article 25 Data protection by design and default. Brussels, European Data Protection Board, 2019.

55   Veale M, Binns R, Van Kleek M. Some HCI priorities for GDPR-compliant machine learning. *arXiv* 2018; preprint [arXiv:1803.06174].

56   Vellido A. Societal issues concerning the application of artificial intelligence in medicine. *Kidney Dis (Basel)* 2019; 5: 11–17.

57   Haute Autorité de Santé. Projet de grille d'analyse pour l'évaluation de dispositifs médicaux avec intelligence artificielle. www.has-sante.fr/jcms/p_3118247/fr/projet-de-grille-d-analyse-pour-l-evaluation-de-dispositifs-medicaux-avec-intelligence-artificielle. Date last updated: 15 January 2020.

58   Pesapane F, Volonte C, Codari M, *et al.* Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. *Insights Imaging* 2018; 9: 745–753.

59   O'Sullivan S, Nevejans N, Allen C, *et al.* Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. *Int J Med Robot* 2019; 15: e1968.

60   Publications Office of the EU. Civil law rules on robotics: European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). *Off J Eur Union* 2018; C252/239.