Applications of artificial intelligence and machine learning in respiratory medicine

Sherif Gonem ,[1,2] Wim Janssens,[3,4] Nilakash Das,[3] Marko Topalovic[3,5]

[1]Department of Respiratory Medicine, Nottingham University Hospitals NHS Trust, Nottingham, UK
[2]Division of Respiratory Medicine, University of Nottingham, Nottingham, UK
[3]Department of Chronic Diseases, Metabolism and Ageing, KU Leuven, Leuven, Belgium
[4]Department of Respiratory Diseases, University Hospitals Leuven, Leuven, Belgium
[5]ArtiQ NV, Leuven, Belgium

Correspondence to
Dr Sherif Gonem, Respiratory Medicine, Nottingham University Hospitals NHS Trust, Nottingham NG5 1PB, UK; sherif.gonem@nottingham.ac.uk

Received 14 January 2020
Revised 19 April 2020
Accepted 22 April 2020

To cite: Gonem S, Janssens W, Das N, *et al*. *Thorax* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/thoraxjnl-2020-214556

ABSTRACT

The past 5 years have seen an explosion of interest in the use of artificial intelligence (AI) and machine learning techniques in medicine. This has been driven by the development of deep neural networks (DNNs)—complex networks residing in silico but loosely modelled on the human brain—that can process complex input data such as a chest radiograph image and output a classification such as 'normal' or 'abnormal'. DNNs are 'trained' using large banks of images or other input data that have been assigned the correct labels. DNNs have shown the potential to equal or even surpass the accuracy of human experts in pattern recognition tasks such as interpreting medical images or biosignals. Within respiratory medicine, the main applications of AI and machine learning thus far have been the interpretation of thoracic imaging, lung pathology slides and physiological data such as pulmonary function tests. This article surveys progress in this area over the past 5 years, as well as highlighting the current limitations of AI and machine learning and the potential for future developments.

INTRODUCTION

'Artificial intelligence (AI)' may be defined as 'the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages'.[1] Machine learning is a subfield of AI in which statistical models are used to learn patterns from data in order to accomplish a specific task. Machine learning techniques range from simple linear models such as logistic regression and naïve Bayes classifiers to complex neural network models with many thousands of parameters.

The explosion of interest in medical applications of AI during the past 5 years may be attributed to the confluence of two key factors:

1. Deep neural networks (DNNs)
   Artificial neural networks (ANNs) are loosely modelled on the human brain and consist of multiple layers of 'neurons' that successively process input data until the output layer is reached. DNNs are a recently developed variant of ANNs that have a large number of intermediate layers (often greater than 10) and process input data in a hierarchical manner, with the first few layers responding to simple low-level features (such as straight lines) and successive layers responding to more abstract high-level features (such as the shape of specific objects).[2] DNNs are typically used to classify the input data into a number of categories. For instance, a chest radiograph image may be classified as 'normal' or 'abnormal'. DNNs have been accompanied by a paradigm shift in AI. In the early years of AI research, the goal was to encode the knowledge of human experts into rule-based 'expert systems' that would be explicitly programmed to look for certain hand-crafted features in the data. However, DNNs are developed using large training datasets and learn in an autonomous manner the features which most discriminate between categories. DNNs therefore have the potential to surpass human experts in classification tasks. Although their accuracy is impressive, a drawback of DNNs is their lack of interpretability. The features that are used to distinguish between data categories are not readily translated into verbal or visual 'rules' that a human can understand.
2. Big data and faster computation
   The success of DNNs has been dependent on the availability of large training datasets that have been correctly labelled with the categories to be distinguished, as well as faster computation to train the DNNs within a reasonable timeframe. For example, a system that can distinguish handwritten numerals requires thousands of examples of each numeral to achieve reasonable accuracy. The widespread use of Electronic Health Records (EHR) and Picture Archiving and Communication Systems has resulted in the availability of large training datasets for healthcare applications, subject to appropriate ethical and information governance safeguards.

DNNs have shown equivalent diagnostic accuracy to expert dermatologists at distinguishing between the macroscopic appearance of malignant and benign skin lesions[3] and to expert pathologists at detecting breast cancer nodal metastases on histological slides.[4] DNNs have been successfully applied to retinal images for the detection of diabetic retinopathy and other retinal pathologies,[5–8] CT images for the detection of acute intracranial events,[9] ECGs for the diagnosis of arrhythmias[10] and cardiac contractile dysfunction,[11] identification of the facial phenotypes of genetic disorders[12] and interpretation of screening mammography.[13] Significant progress has also been made in the analysis of EHRs for medical diagnosis[14] and predicting future events such as acute kidney injury.[15] DNNs have shown an ability to detect subtle features that are undetectable by human observers even in hindsight. For instance, Attia *et al*[16] developed a DNN that accurately predicted the presence of atrial fibrillation occurring previously or in the near future using ECGs recorded during sinus rhythm.

This review will focus on the advances that have been made in AI and machine learning as applied

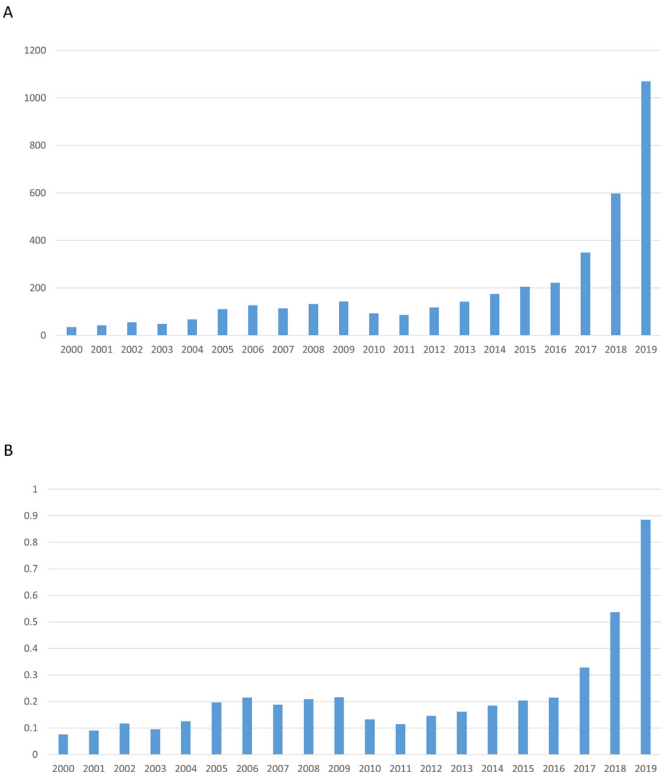

Figure 1 Published articles on artificial intelligence and machine learning in respiratory medicine shown as raw numbers (panel A) and as a percentage of all articles on respiratory medicine (panel B) from 2000 to 2019. Panel A shows the number of retrieved articles on the PubMed database published from 2000 to 2019 using the search terms specified for this review (artificial intelligence related term and respiratory-related term). Panel B shows this number expressed as a percentage of all articles retrieved using just the respiratory-related search terms.

to respiratory medicine in the past 5 years. The inputs that have been subjected to machine learning techniques may be broadly categorised into:

i. Thoracic imaging.
ii. Histopathology or cytology.
iii. Physiological measurements and biosignals.

Search strategy:

The following search was performed on the PubMed database on 19 March 2020:

('artificial intelligence'[All Fields] OR 'machine learning'[All Fields] OR 'deep learning'[All Fields] OR 'neural network'[All Fields]) AND ('chest'[All Fields] OR 'lung'[All Fields] OR 'pulmonary'[All Fields] OR 'respiratory'[All Fields] OR 'thorax'[All Fields] OR 'thoracic'[All Fields] OR 'pneumonia'[All Fields] OR 'pneumonitis'[All Fields] OR 'bronchiectasis'[All Fields] OR 'bronchiolitis'[All Fields] OR 'cystic fibrosis'[All Fields] OR 'tuberculosis'[All Fields] OR 'mycobacteria'[All Fields] OR 'asthma'[All Fields] OR 'copd'[All Fields] OR 'pleural'[All Fields] OR 'sarcoidosis'[All Fields] OR 'sleep'[All Fields] OR 'ventilation'[All Fields])].

A total of 4610 results were returned. All abstracts were reviewed by the first author (SG), and full-text articles were retrieved for papers that described a clinically relevant advance in the field. These papers were reviewed in detail, and examples representing the state of the art in each subfield were selected for inclusion in this narrative review. It was observed that there has been an exponential increase in published papers on this topic starting from 2016 (figure 1).

THORACIC IMAGING

The application of DNNs to chest radiographs and CT scans has resulted in a step change in diagnostic accuracy compared with qualitative semantic features such as tumour spiculation and quantitative features such as shape and texture derived using image analysis software (often referred to as radiomics). The advantage of DNNs is that they derive features directly from the data, resulting in greater accuracy than hand-crafted qualitative or quantitative analyses but with the disadvantage of reduced interpretability. However, some progress has been made towards correlating 'deep features' derived from DNNs with semantic features that are detectable by human radiologists.[17]

A number of algorithms have been developed for automated reporting or triage of plain chest radiographs, in many cases exceeding the accuracy of expert thoracic radiologists. Annarumma *et al*[18] trained a DNN to triage chest radiographs as 'normal', 'non-urgent', 'urgent' and 'critical' using a training dataset of 329 698 images. The AI system detected normal radiographs with a sensitivity of 71%, specificity of 95% and a positive predictive value of 73% in the test dataset. In a simulated radiology reporting pipeline in which the AI was used to prioritise urgent and critical radiographs for reporting by a radiologist, there was an approximately fourfold reduction in the delay to report radiographs with critical findings and a twofold reduction in the delay to report urgent findings. A similar chest radiograph triage system using a binary classification of 'normal' or 'abnormal' was developed by Yates *et al*,[19] with a final model accuracy of 94.6% in the test dataset. These findings suggest a potentially important role for AI in prioritising cases for review by a radiologist, in order to expedite the reporting of cases with critical abnormalities. This could be particularly relevant in resource-poor settings in which there is a shortage of trained radiologists. AI assessment of chest imaging may also have prognostic significance: Lu *et al*[20] developed a DNN that accurately predicted all-cause mortality over a follow-up period of 12 years based on a single plain chest radiograph, even after adjusting for radiologists' diagnostic findings and standard risk factors for mortality.

DNNs have been trained to recognise specific pathologies on chest radiographs including tuberculosis,[21–24] malignant pulmonary nodules,[25] congestive cardiac failure[26] and pneumothorax.[27] Hwang *et al*[28] developed a DNN that could recognise lung cancer, tuberculosis, pneumonia and pneumothorax on chest radiographs as well as providing visual localisation of the abnormality. In a head-to-head comparison using the same test dataset, the DNN achieved an area under the receiver operating curve (AUC; a measure of the accuracy of a diagnostic test) of 0.983, exceeding that of thoracic radiologists (0.932), general radiologists (0.896) and non-radiology physicians (0.614). The same research group subsequently tested this DNN algorithm in an emergency department setting and found that it improved the sensitivity of radiology residents in the detection of clinically significant abnormalities when used as a second reader.[29] However, the DNN was not trained to interpret radiographs with multiple pathologies, nor to interpret images in the context of background clinical information. Therefore, while current DNNs cannot replace radiologist reporting of chest radiographs, they may act as a competent second reader to reduce perceptual errors. Prospective studies incorporating DNNs as a second

reader in routine clinical practice are warranted to determine whether they can reduce the rate of reporting errors.

Evidence that lung cancer screening can reduce mortality is steadily accumulating, and a recent European Union position statement has concluded that implementation of low-dose CT screening should start throughout Europe as soon as possible.[30] An important limiting factor in this implementation is the availability of radiologists to report the large volume of screening CT scans. There has therefore been substantial interest in developing AI systems that can detect and accurately diagnose malignant pulmonary nodules on CT imaging.[31–35] Ardila *et al*[31] trained a DNN to predict the risk of lung cancer based on current and previous chest CT scans using cases from the National Lung Cancer Screening Trial. The DNN achieved an AUC of 0.944 for predicting biopsy-proven cancer in the test dataset. The accuracy of the DNN was higher than that of six board-certified radiologists when only the current CT scan was available and was equivalent to the radiologists when both current and previous CT scans were available for review. Similarly, Baldwin *et al* developed a DNN to predict malignancy in incidentally detected pulmonary nodules measuring 5–15 mm and achieved an AUC of 0.896 in the test dataset, which was significantly higher than that of the Brock model currently recommended in UK guidelines.[32] In order to improve the interpretability and clinical acceptability of DNN predictions, Shen *et al*[33] merged deep learning techniques with more traditional semantic features such as nodule calcification and margin definition. Incorporating semantic features into DNN predictions did not significantly affect model accuracy but may have improved interpretability of the model output. A number of investigators have taken a hybrid approach, in which radiomic features are entered into machine learning models in order to derive the best combination of features to optimise classification accuracy.[34 35] Delzell *et al*[34] measured 416 quantitative imaging biomarkers in CT scans of pulmonary nodules from 200 patients and entered these radiomic features into a variety of machine learning models. The best performing models were elastic net and support vector machine, which achieved an AUC of 0.72 for distinguishing benign from malignant nodules.

Beyond lung cancer diagnosis, there is evidence that DNNs can be used to predict prognosis and tumour type based on CT images. Hosny *et al*[36] trained a DNN to predict survival based on CT appearances in patients with non-small cell lung cancer undergoing surgery or radiotherapy. The DNN distinguished between early (<2 years) and late (≥2 years) mortality with an AUC of 0.71 and 0.70 in patients undergoing surgery and radiotherapy, respectively. Wang *et al*[37] found that a DNN could predict epithelial growth factor receptor mutation status in patients with lung adenocarcinoma based on CT images, with an AUC of 0.81 in the validation dataset. The accuracy of the DNN significantly exceeded that of predictive models using clinical features alone, semantic features or radiomics features.

Machine learning techniques have also been used for the diagnosis of interstitial lung disease.[38 39] Walsh *et al*[38] trained a DNN using a total of 420 096 montages each consisting of four transverse CT images. These were derived from full high-resolution CT scans of 210 patients with usual interstitial pneumonia (UIP), 392 with possible UIP and 327 whose scans were considered inconsistent with UIP. The reference standard for each CT scan was determined by an experienced thoracic radiologist with a specialist interest in interstitial lung disease. In a test set of 68 093 montages derived from 139 separate patients, the algorithm achieved an accuracy of 76.4%. A second test set consisted of 150 four-slice montages from CT scans that had been previously evaluated by 91 thoracic radiologists, with the reference standard being the majority opinion of the radiologists. The algorithm achieved an accuracy of 73.3% in this test set that was comparable with the median accuracy of the individual radiologists (70.7%). Moreover, in a Cox regression analysis, an algorithm diagnosis of UIP was associated with a HR for mortality of 2.88, compared with a diagnosis of 'not UIP', with the equivalent HR for a majority radiologist opinion diagnosis of UIP being 2.74.

González *et al*[40] trained a DNN using four-slice CT montages from 7983 smokers who took part in the COPDGene study and found that the algorithm could accurately diagnose COPD, with an AUC of 0.856. A subsequent study using the same dataset found that the staging of emphysema from 'absent' to 'advanced destructive' using a DNN was highly predictive of survival and lung function.[41] DNNs have also been developed to diagnose and evaluate the burden of thrombus in acute pulmonary embolism. The algorithm developed by Liu *et al*[42] achieved an AUC of 0.926 for the diagnosis of pulmonary embolism, and the clot burden measured by the DNN correlated significantly with manual (Qanadli and Mastora) scores and with measures of right ventricular function.

HISTOPATHOLOGY AND CYTOLOGY

Deep learning techniques have been successfully applied to digital histology images, particularly in the field of lung cancer. Coudray *et al*[43] found that a DNN could distinguish between adenocarcinoma and squamous cell carcinoma of the lung with comparable accuracy with expert pathologists (AUC of 0.97). This significantly exceeded the performance of traditional image-processing techniques with hand-crafted features, which achieved an AUC of approximately 0.75 for the same task.[44] Moreover, the DNN could also predict the presence or absence of six common gene mutations of therapeutic significance (STK11, EGFR, FAT1, SETBP1, KRAS and TP53) with AUC values ranging from 0.73 to 0.86. Similarly, Sha *et al*[45] trained a DNN to predict programmed death-ligand 1 status in non-small cell lung cancer based on morphological appearances on standard H&E stained tumour sections, with an AUC of 0.80. DNNs have also been trained to accurately differentiate between lung adenocarcinoma growth patterns (acinar, micropapillary, solid, lepidic and cribriform),[46 47] as well as to detect lung cancer metastases in lymph node slides.[48] Courtiol *et al*[49] trained a DNN (MesoNet) to accurately predict overall survival of patients with malignant mesothelioma based on whole slide digitised images. The predictions made by MesoNet cut across traditional histological boundaries (such as epithelioid, sarcomatoid and biphasic) and moreover identified the specific regions within the slides that most contributed to patient outcome prediction.

Kim *et al*[50] used machine learning methods (support vector machines and penalised logistic regression) to develop classifiers for interstitial lung diseases based on high-dimensional transcriptional data from surgical lung biopsies. In a subsequent prospective study, the investigators found that the molecular classifier they developed could accurately distinguish between UIP and non-UIP in less invasive transbronchial biopsy samples, suggesting that the technique could avoid the need for surgical biopsy in some cases.[51]

Xiong *et al*[52] trained a DNN to recognise acid fast-stained *Mycobacterium tuberculosis* bacilli on digital cytology slides. The small size of the bacilli (20×4 pixels) and the loss of resolution when scanning the digital images resulted in some technical challenges. Although good sensitivity of 98% was achieved following modifications to the algorithm, there were a number

of false positive results due to contaminant bacilli and slide artefacts, resulting in a specificity of 84%.

PHYSIOLOGICAL MEASUREMENTS AND BIOSIGNALS

Interpretation of pulmonary function tests including spirometry, body plethysmography and measurement of diffusing capacity has traditionally been considered an important aspect of the expertise of respiratory physicians. Topalovic *et al*[53] developed a random forest machine learning model using 1430 historical cases that could accurately differentiate between eight categories of respiratory disease. In a head-to-head comparison using 50 test cases, the model displayed an accuracy of 82% and outperformed 120 European pulmonologists by a wide margin.

Machine learning approaches have also been applied to the forced oscillation technique (FOT), which measures respiratory impedance non-invasively using sound waves, with minimal effort from the subject.[54] Amaral *et al*[55] applied a variety of machine learning models (including K nearest neighbour, decision trees, ANNs and support vector machines) to FOT measurements to detect COPD (AUC >0.95) to discriminate between different Global initiative for Obstructive Lung Disease stages of airflow obstruction[56] and to identify early smoking-induced changes in healthy subjects,[57] as well as to identify airflow obstruction in patients with asthma.[58]

Breath analysis offers excellent potential to phenotype respiratory disorders because exhaled breath contains a mixture of gases and traces of many volatile organic compounds (VOCs) that emanate from the respiratory tract itself. Several techniques exist to measure VOCs in the breath, such as gas chromatography–mass spectroscopy, electronic nose and chemical sensors, each of which require advanced pattern recognition methods to identify abnormal signatures in measured VOCs.[59] Machine learning methods such as decision trees and support vector machines on VOC data have been used to discriminate COPD and healthy individuals[60] and to detect lung cancer.[61] Brinkman *et al*[62] used an electronic nose to classify inflammatory asthma phenotypes using K-means and Ward clustering. These unsupervised learning techniques do not rely on prelabelling but instead group the cases into novel categories or clusters based on similarity of exhaled breath metabolites.

Computerised lung sound analysis involves discriminating between normal and adventitious lung sounds obtained during auscultation. Although machine learning has become a standard method to classify adventitious sounds, these sound events are intermittent and highly variable from one person to another presenting a challenge in generalising these algorithms to a general population.[63] Machine learning approaches (including ANNs and support vector machines) have been applied to classify adventitious sounds associated with asthma,[64] COPD[65] and interstitial lung disease[66] and to detect common respiratory disorders in children using cough sounds.[67] Bardou *et al*[68] found that DNNs outperformed traditional machine learning techniques in the classification of lung sounds into seven categories (normal, coarse crackle, fine crackle, monophonic wheeze, polyphonic wheeze, squawk and stridor).

Analysis of biosignals using machine learning may permit a superior understanding of the dynamics of physiological regulation in health and disease. Examples of biosignal monitoring in the respiratory sphere include polysomnography, which is used to diagnose obstructive sleep apnoea and other sleep disorders. Nikkonen *et al*[69] developed an ANN that accurately determined the oxygen desaturation index (ODI) and apnoea–hypopnoea index (AHI) using only the oxygen saturation signal as input. The median absolute error was 0.78 events/hour for AHI and 0.68 events/hour for ODI, using manual scoring of events as the gold standard. Allocca *et al*[70] developed an automated sleep-stage classification programme that achieved high accuracy against a gold standard of manual visual scoring in human, rodent and pigeon polysomnography data. Mousavi *et al*[71] developed a DNN to annotate various sleep stages using an openly accessible electroencephalogram (EEG) dataset, achieving an accuracy of 84%. Automated monitoring of biosignals has also been proposed as a solution to patient-ventilator asynchrony, which is a mismatch between ventilator delivery and patient demand. Gholami *et al*[72] developed a random forest machine learning model to detect cycling asynchrony based on waveform analysis with positive and negative predictive values above 90%.

As the use of smartphones, sensors and wearables proliferates, telemedicine may become an important tool for the self-management of respiratory disorders. By monitoring clinical outcomes at an individual level, such technologies facilitate preventive and pre-emptive care while providing medical expertise remotely. Machine learning offers a powerful solution to analyse patterns associated with respiratory outcomes in data collected by telemonitoring devices.[73] Machine learning methods such as naïve Bayes classifiers and support vector machines have been applied to home peak expiratory flow measurements and symptom scores to predict exacerbations a week early in adults[74] and children[75] with asthma. Similar studies have also been published to predict exacerbations in patients with COPD.[76 77]

CONCLUSION AND FUTURE DEVELOPMENTS

Research into AI in medicine has accelerated markedly since 2015, with the field of respiratory medicine being well represented. DNNs are emerging as a key tool to develop imaging biomarkers for diagnosis, prognosis and prediction of response to therapy. Figure 2 summarises the process by which machine learning models may be developed and incorporated into routine clinical practice in the near future. There remains an enormous potential for DNNs to embrace domains outside of imaging such as pulmonary function tests and physiological biosignals. However, a major limitation for such computational approaches is a shortage of sufficiently large medical training datasets. Overcoming this will require large-scale collaborations such as the recently formed Open-Source Imaging Consortium (https://www.osicild.org/), a collaboration between academia and industry to develop imaging biomarkers for idiopathic pulmonary fibrosis and other interstitial lung diseases using AI.

Large clinical databases from multicentre randomised controlled trials are another underexplored domain. Applying DNNs to these detailed datasets, potentially including merged data from multiple similar trials, carries the potential to predict treatment effects for individual patients, ushering in a new era of personalised medicine. The benefits of sharing and reuse of individual participant data from clinical trials are increasingly recognised but will require a robust internationally recognised ethical and legal framework to gain wider adoption and acceptance.[78] Similarly, data collected during the course of routine clinical practice has great potential for training AI algorithms for patient benefit. However, clear legal and ethical guidelines are needed to maximise the benefit of such datasets while preserving the confidentiality of individual patients.[79]

Natural language processing (NLP) is still at an early stage of development but in future may be deployed to extract clinical insights from the vast pool of unstructured EHR[80] or to extract relationships between concepts from the rapidly expanding

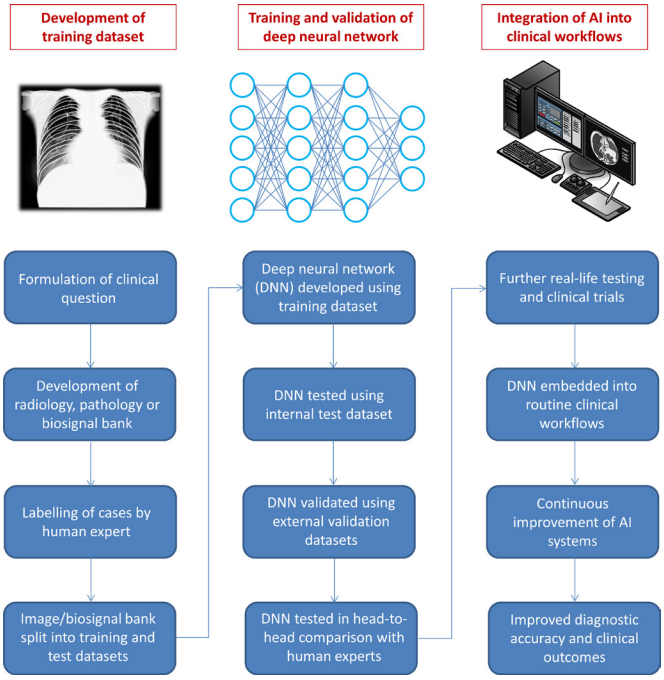

Figure 2 Typical path for the development of a machine learning model and its incorporation into clinical practice. Chest radiograph and imaging workstation images are from www.shutterstock.com (reproduced under licence). Contributing artists were sfam_photo and Zern Liew, respectively.

body of medical research.[81] NLP may also be used to accelerate the development of AI algorithms for interpreting radiological or histological images, by automatically converting free-text radiology or pathology reports into a structured format suitable for training DNNs, potentially obviating the need for manual labelling of cases.[82–84]

While the advances made over the past 5 years have been impressive, a number of challenges must still be overcome before AI can be widely adopted into routine clinical practice.[79 85 86] These include intrinsic problems with the machine learning algorithms themselves, logistical difficulties and social or cultural barriers. It is known that DNNs have the potential to misclassify examples that have been subtly altered, even by the addition of a few extra pixels.[87] In a clinical setting, this could manifest as a lack of generalisability; for instance, a DNN model trained on imaging data from the latest scanner at an advanced care facility may not function properly at a hospital that has older machines. Similarly, machine learning models are prone to perpetuating biases that may exist in the training dataset, as well as spurious associations in which confounding factors are used as predictors.[86] A related problem with DNNs and other complex machine learning algorithms is their lack of interpretability, which may be defined as an ability to provide reasons for their output. DNNs often act like 'black boxes' with the reasons for their output remaining opaque, even to their developers. In the medical sphere, interpretability is critical for gaining trust, particularly if important management decisions are being made based on the evaluation of a DNN. Fortunately, progress is now being made towards more interpretable AI. Several techniques have been developed that can generate explanations by estimating how the input features or different regions within an input image contributed to the output.[85] These techniques should allow a closer inspection of DNN outputs by clinicians so that decisions based on faulty or biased explanations can be over-ruled.

In conclusion, AI and machine learning have the power to transform many aspects of respiratory medicine. The emergence of DNNs developed using big training datasets has resulted in a number of novel applications, particularly in the field of thoracic imaging. However, DNN models still suffer from problems with interpretability, generalisability and potential bias. Rigorous validation strategies combined with the development of new standards for reporting machine learning models are required to address these issues before AI can take its place in the clinic.[88]

Contributors SG conceived the idea for the manuscript, undertook literature search, co-wrote the first draft, and prepared the final draft and figures; WJ, ND and MT undertook literature search, co-wrote the first draft and critically appraised the final draft.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests SG has received speaker's fees from Teva and consultancy fees from Anaxsys and 3M. WJ has received grants from AstraZeneca, Chiesi and GSK. WJ and MT are co-founders of ArtiQ, a spinoff company of KU Leuven. ND has no competing interests to declare.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iD
Sherif Gonem http://orcid.org/0000-0002-6080-2246

REFERENCES

1 Oxford University Press, 2019. Available: https://www.lexico.com [Accessed 1 Jan 2020].
2 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–44.
3 Esteva A, Kuprel B, Novoa RA, *et al*. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115–8.
4 Ehteshami Bejnordi B, Veta M, Johannes van Diest P, *et al*. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017;318:2199–210.
5 Gulshan V, Peng L, Coram M, *et al*. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus Photographs. *JAMA* 2016;316:2402–10.
6 Ting DSW, Cheung CY-L, Lim G, *et al*. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 2017;318:2211–23.
7 Poplin R, Varadarajan AV, Blumer K, *et al*. Prediction of cardiovascular risk factors from retinal fundus Photographs via deep learning. *Nat Biomed Eng* 2018;2:158–64.
8 De Fauw J, Ledsam JR, Romera-Paredes B, *et al*. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* 2018;24:1342–50.
9 Titano JJ, Badgeley M, Schefflein J, *et al*. Automated deep-neural-network surveillance of cranial images for acute neurologic events. *Nat Med* 2018;24:1337–41.
10 Hannun AY, Rajpurkar P, Haghpanahi M, *et al*. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 2019;25:65–9.
11 Attia ZI, Kapa S, Lopez-Jimenez F, *et al*. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med* 2019;25:70–4.
12 Gurovich Y, Hanani Y, Bar O, *et al*. Identifying facial phenotypes of genetic disorders using deep learning. *Nat Med* 2019;25:60–4.
13 McKinney SM, Sieniek M, Godbole V, *et al*. International evaluation of an AI system for breast cancer screening. *Nature* 2020;577:89–94.
14 Liang H, Tsui BY, Ni H, *et al*. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat Med* 2019;25:433–8.
15 Tomašev N, Glorot X, Rae JW, *et al*. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 2019;572:116–9.
16 Attia ZI, Noseworthy PA, Lopez-Jimenez F, *et al*. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet* 2019;394:861–7.
17 Paul R, Schabath M, Balagurunathan Y, *et al*. Explaining deep features using Radiologist-Defined semantic features and traditional quantitative features. *Tomography* 2019;5:192–200.
18 Annarumma M, Withey SJ, Bakewell RJ, *et al*. Automated triaging of adult chest radiographs with deep artificial neural networks. *Radiology* 2019;291:196–202.
19 Yates EJ, Yates LC, Harvey H. Machine learning "red dot": open-source, cloud, deep convolutional neural networks in chest radiograph binary normality classification. *Clin Radiol* 2018;73:827–31.

20 Lu MT, Ivanov A, Mayrhofer T, *et al*. Deep learning to assess long-term mortality from chest radiographs. *JAMA Netw Open* 2019;2:e197416.
21 Hwang EJ, Park S, Jin K-N, *et al*. Development and validation of a deep Learning-based automatic detection algorithm for active pulmonary tuberculosis on chest radiographs. *Clin Infect Dis* 2019;69:739–47.
22 Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using Convolutional neural networks. *Radiology* 2017;284:574–82.
23 Pasa F, Golkov V, Pfeiffer F, *et al*. Efficient deep network architectures for fast chest X-ray tuberculosis screening and visualization. *Sci Rep* 2019;9:6268.
24 Qin ZZ, Sander MS, Rai B, *et al*. Using artificial intelligence to read chest radiographs for tuberculosis detection: a multi-site evaluation of the diagnostic accuracy of three deep learning systems. *Sci Rep* 2019;9:15000.
25 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;290:218–28.
26 Seah JCY, Tang JSN, Kitchen A, *et al*. Chest radiographs in congestive heart failure: visualizing neural network learning. *Radiology* 2019;290:514–22.
27 Taylor AG, Mielke C, Mongan J. Automated detection of moderate and large pneumothorax on frontal chest x-rays using deep convolutional neural networks: a retrospective study. *PLoS Med* 2018;15:e1002697.
28 Hwang EJ, Park S, Jin K-N, *et al*. Development and validation of a deep Learning-Based automated detection algorithm for major thoracic diseases on chest radiographs. *JAMA Netw Open* 2019;2:e191095.
29 Hwang EJ, Nam JG, Lim WH, *et al*. Deep learning for chest radiograph diagnosis in the emergency department. *Radiology* 2019;293:573–80.
30 Oudkerk M, Devaraj A, Vliegenthart R, *et al*. European position statement on lung cancer screening. *Lancet Oncol* 2017;18:e754–66.
31 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;25:954–61.
32 Baldwin DR, Gustafson J, Pickup L, *et al*. External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules. *Thorax* 2020;75:306–12.
33 Shen S, Han SX, Aberle DR, *et al*. An interpretable deep hierarchical semantic Convolutional neural network for lung nodule malignancy classification. *Expert Syst Appl* 2019;128:84–95.
34 Delzell DAP, Magnuson S, Peter T, *et al*. Machine learning and feature selection methods for disease classification with application to lung cancer screening image data. *Front Oncol* 2019;9:1393.
35 Tu S-J, Wang C-W, Pan K-T, *et al*. Localized thin-section CT with radiomics feature extraction and machine learning to classify early-detected pulmonary nodules from lung cancer screening. *Phys Med Biol* 2018;63:065005.
36 Hosny A, Parmar C, Coroller TP, *et al*. Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study. *PLoS Med* 2018;15:e1002711.
37 Wang S, Shi J, Ye Z, *et al*. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *Eur Respir J* 2019;53. doi:10.1183/13993003.00986-2018. [Epub ahead of print: 28 Mar 2019].
38 Walsh SLF, Calandriello L, Silva M, *et al*. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. *Lancet Respir Med* 2018;6:837–45.
39 Christe A, Peters AA, Drakopoulos D, *et al*. Computer-Aided diagnosis of pulmonary fibrosis using deep learning and CT images. *Invest Radiol* 2019;54:627–32.
40 González G, Ash SY, Vegas-Sánchez-Ferrero G, *et al*. Disease staging and prognosis in smokers using deep learning in chest computed tomography. *Am J Respir Crit Care Med* 2018;197:193–203.
41 Humphries SM, Notary AM, Centeno JP, *et al*. Deep learning enables automatic classification of emphysema pattern at CT. *Radiology* 2020;294:434–44.
42 Liu W, Liu M, Guo X, *et al*. Evaluation of acute pulmonary embolism and clot burden on CtpA with deep learning. *Eur Radiol* 2020. doi:10.1007/s00330-020-06699-8. [Epub ahead of print: 16 Feb 2020].
43 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;24:1559–67.
44 Yu K-H, 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.
45 Sha L, Osinski BL, Ho IY, *et al*. Multi-Field-of-View deep learning model predicts nonsmall cell lung cancer programmed Death-Ligand 1 status from Whole-Slide hematoxylin and eosin images. *J Pathol Inform* 2019;10:24.
46 Gertych A, Swiderska-Chadaj Z, Ma Z, *et al*. Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides. *Sci Rep* 2019;9:1483.
47 Wei JW, Tafe LJ, Linnik YA, *et al*. Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. *Sci Rep* 2019;9:3358.
48 Pham HHN, Futakuchi M, Bychkov A, *et al*. Detection of lung cancer lymph node metastases from Whole-Slide histopathologic images using a two-step deep learning approach. *Am J Pathol* 2019;189:2428–39.
49 Courtiol P, Maussion C, Moarii M, *et al*. Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nat Med* 2019;25:1519–25.
50 Kim SY, Diggans J, Pankratz D, *et al*. Classification of usual interstitial pneumonia in patients with interstitial lung disease: assessment of a machine learning approach using high-dimensional transcriptional data. *Lancet Respir Med* 2015;3:473–82.
51 Raghu G, Flaherty KR, Lederer DJ, *et al*. Use of a molecular classifier to identify usual interstitial pneumonia in conventional transbronchial lung biopsy samples: a prospective validation study. *Lancet Respir Med* 2019;7:487–96.
52 Xiong Y, Ba X, Hou A, *et al*. Automatic detection of Mycobacterium tuberculosis using artificial intelligence. *J Thorac Dis* 2018;10:1936–40.
53 Topalovic M, Das N, Burgel P-R, *et al*. Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests. *Eur Respir J* 2019;53:pii: 1801660:1801660.
54 Oostveen E, MacLeod D, Lorino H, *et al*. The forced oscillation technique in clinical practice: methodology, recommendations and future developments. *Eur Respir J* 2003;22:1026–41.
55 Amaral JLM, Lopes AJ, Jansen JM, *et al*. Machine learning algorithms and forced oscillation measurements applied to the automatic identification of chronic obstructive pulmonary disease. *Comput Methods Programs Biomed* 2012;105:183–93.
56 Amaral JLM, Lopes AJ, Faria ACD, *et al*. Machine learning algorithms and forced oscillation measurements to categorise the airway obstruction severity in chronic obstructive pulmonary disease. *Comput Methods Programs Biomed* 2015;118:186–97.
57 Amaral JLM, Lopes AJ, Jansen JM, *et al*. An improved method of early diagnosis of smoking-induced respiratory changes using machine learning algorithms. *Comput Methods Programs Biomed* 2013;112:441–54.
58 Amaral JLM, Lopes AJ, Veiga J, *et al*. High-Accuracy detection of airway obstruction in asthma using machine learning algorithms and forced oscillation measurements. *Comput Methods Programs Biomed* 2017;144:113–25.
59 van der Schee MP, Paff T, Brinkman P, *et al*. Breathomics in lung disease. *Chest* 2015;147:224–31.
60 Phillips CO, Syed Y, Parthaláin NM, Mac PN, *et al*. Machine learning methods on exhaled volatile organic compounds for distinguishing COPD patients from healthy controls. *J Breath Res* 2012;6:036003.
61 Huang C-H, Zeng C, Wang Y-C, *et al*. A study of diagnostic accuracy using a chemical sensor array and a machine learning technique to detect lung cancer. *Sensors* 2018;18:pii: E2845:2845.
62 Brinkman P, Wagener AH, Hekking P-P, *et al*. Identification and prospective stability of electronic nose (eNose)-derived inflammatory phenotypes in patients with severe asthma. *J Allergy Clin Immunol* 2019;143:1811–20.
63 Pramono RXA, Bowyer S, Rodriguez-Villegas E. Automatic adventitious respiratory sound analysis: a systematic review. *PLoS One* 2017;12:e0177926.
64 Islam MA, Bandyopadhyaya I, Bhattacharyya P, *et al*. Multichannel lung sound analysis for asthma detection. *Comput Methods Programs Biomed* 2018;159:111–23.
65 Jácome C, Marques A. Computerized respiratory sounds in patients with COPD: a systematic review. *COPD* 2015;12:104–12.
66 Flietstra B, Markuzon N, Vyshedskiy A, *et al*. Automated analysis of crackles in patients with interstitial pulmonary fibrosis. *Pulm Med* 2011;2011:1–7.
67 Porter P, Abeyratne U, Swarnkar V, *et al*. A prospective multicentre study testing the diagnostic accuracy of an automated cough sound centred analytic system for the identification of common respiratory disorders in children. *Respir Res* 2019;20:81.
68 Bardou D, Zhang K, Ahmad SM. Lung sounds classification using convolutional neural networks. *Artif Intell Med* 2018;88:58–69.
69 Nikkonen S, Afara IO, Leppänen T, *et al*. Artificial neural network analysis of the oxygen saturation signal enables accurate diagnostics of sleep apnea. *Sci Rep* 2019;9:13200.
70 Allocca G, Ma S, Martelli D, *et al*. Validation of 'Somnivore', a Machine Learning Algorithm for Automated Scoring and Analysis of Polysomnography Data. *Front Neurosci* 2019;13:207.
71 Mousavi S, Afghah F, Acharya UR. SleepEEGNet: automated sleep stage scoring with sequence to sequence deep learning approach. *PLoS One* 2019;14:e0216456.
72 Gholami B, Phan TS, Haddad WM, *et al*. Replicating human expertise of mechanical ventilation waveform analysis in detecting patient-ventilator cycling asynchrony using machine learning. *Comput Biol Med* 2018;97:137–44.
73 Sanchez-Morillo D, Fernandez-Granero MA, Leon-Jimenez A. Use of predictive algorithms in-home monitoring of chronic obstructive pulmonary disease and asthma: a systematic review. *Chron Respir Dis* 2016;13:264–83.
74 Finkelstein J, Jeong IC. Machine learning approaches to personalize early prediction of asthma exacerbations. *Ann N Y Acad Sci* 2017;1387:153–65.
75 Luo G, Stone BL, Fassl B, *et al*. Predicting asthma control deterioration in children. *BMC Med Inform Decis Mak* 2015;15:84.
76 Shah SA, Velardo C, Farmer A, *et al*. Exacerbations in chronic obstructive pulmonary disease: identification and prediction using a digital health system. *J Med Internet Res* 2017;19:e69.
77 Orchard P, Agakova A, Pinnock H, *et al*. Improving prediction of risk of hospital admission in chronic obstructive pulmonary disease: application of machine learning to telemonitoring data. *J Med Internet Res* 2018;20:e263.

78 Ohmann C, Banzi R, Canham S, *et al*. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ Open* 2017;7:e018647.
79 Panch T, Mattie H, Celi LA. The "inconvenient truth" about AI in healthcare. *NPJ Digit Med* 2019;2:77.
80 Sheikhalishahi S, Miotto R, Dudley JT, *et al*. Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR Med Inform* 2019;7:e12239.
81 Zhang Y, Lin H, Yang Z, *et al*. Neural network-based approaches for biomedical relation classification: a review. *J Biomed Inform* 2019;99:103294.
82 Sorin V, Barash Y, Konen E, *et al*. Deep learning for natural language processing in Radiology-Fundamentals and a systematic review. *J Am Coll Radiol* 2020. doi:10.1016/j.jacr.2019.12.026. [Epub ahead of print: 28 Jan 2020].
83 Weikert T, Nesic I, Cyriac J, *et al*. Towards automated generation of curated datasets in radiology: application of natural language processing to unstructured reports exemplified on CT for pulmonary embolism. *Eur J Radiol* 2020;125:108862.
84 Gao S, Qiu JX, Alawad M, *et al*. Classifying cancer pathology reports with hierarchical self-attention networks. *Artif Intell Med* 2019;101:101726.
85 Kelly CJ, Karthikesalingam A, Suleyman M, *et al*. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019;17:195.
86 Challen R, Denny J, Pitt M, *et al*. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf* 2019;28:231–7.
87 Heaven D. Why deep-learning AIS are so easy to fool. *Nature* 2019;574:163–6.
88 Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *Lancet* 2019;393:1577–9.