



Unravelling machine learning: insights in respiratory medicine

Elsa Angelini¹, Simon Dahan¹ and Anand Shah^{2,3}

Affiliations: ¹NIHR Imperial Biomedical Research Centre, Institute of Translational Medicine and Therapeutics (ITMAT), ITMAT Data Science Group, Imperial College London, London, UK. ²National Heart and Lung Institute, Imperial College London, London, UK. ³Royal Brompton and Harefield NHS Foundation Trust, London, UK.

Correspondence: Anand Shah, Adult Cystic Fibrosis Centre, Royal Brompton Hospital, Sydney Street, London, SW3 6NP, UK. E-mail: a.shah2@rbht.nhs.uk

 @ERSpublications

Machine learning is the brain of an artificial intelligence machine. We describe how it is performed and explore its current use in respiratory medicine. Potential future applications and possible issues in integration into clinical practice are discussed. <http://bit.ly/31XVruW>

Cite this article as: Angelini E, Dahan S, Shah A. Unravelling machine learning: insights in respiratory medicine. *Eur Respir J* 2019; 54: 1901216 [<https://doi.org/10.1183/13993003.01216-2019>].

What is machine learning?

Artificial intelligence (AI) is considered the science of creating intelligent programmes. Data scientists try to incorporate some features of human intelligence into machines in order to achieve specific tasks. These powerful and trained machines are able to solve problems quickly, robustly and in a reproducible manner. AI is not a new concept, but its evolution has recently undergone a massive transformation due to the drastic improvement of learning methods, computational power and access to large datasets. Machine learning is the brain of an AI machine. The core of machine learning is to create algorithms which learn from input data in order to automatically perform a targeted task, for example, making a decision or a prediction. In machine learning, inputs are numerical features (raw data or derived features). For instance, if the task is to detect a skin lesion, we might use a whole picture, while if the task is to label a detected lesion as benign or malignant, we might use the width, colour, and regularity of the lesion as features. Given a pre-selected type of machine learning model, a type of input data and a targeted task, the data scientist trains the algorithm on a cohort of example cases (training data) to perform the task as “accurately” as possible. Inside the algorithm, a variety of data manipulation, comparison and aggregation techniques is deployed and optimised, to transform the input numerical values into a final instruction or decision (the “task”). Distinction is made between supervised and unsupervised learning. While the former within healthcare can alleviate a diagnostic task and lead to extended understanding of key informative factors of a given pathology, the latter opens the way to discovering new phenotypes.

Deep learning (DL) is currently the most powerful machine learning algorithm. DL algorithms are able to learn from raw (or with little pre-processing) input data and build by themselves sophisticated abstract feature representations (useful patterns) that enable very accurate task decision making. This is a major revolution, from traditional machine learning which relied on explicit formulation of a decision-making processes and rules. DL is particularly performant in handling complex input data (large size, large number of variables, large variability) for categorising, finding patterns, and extracting discriminating information. The inherent limitation of any machine learning is the dependence on the “training” data and the risk of making errors with any future “unseen” cases if the training cohort did not include “similar” cases. Testing the reproducibility of the task performance quality on an independent cohort is therefore essential.

Received: 20 June 2019 | Accepted after revision: 26 Sept 2019

Copyright ©ERS 2019

How does it work?

The base of DL is the use of neural networks. These are stacks of units (the “neurons”) organised into multiple connected layers (the network). Connections use weights to control how neurons exchange information within and across layers. Between layers, non-linear operations are performed to aggregate data into more abstract representation based on data components that are seen as most “informative” for the target task. The learning process adjusts the weights in order to optimise the task performance quality, according to an explicit quality metric. The learning stage is composed of multiple iterations over which the algorithm improves itself incrementally, adjusting the weights between neurons. The final architecture with the learned weights forms the predictive model that can be used on new unseen data (figure 1). The number of layers defines the depth of the network. The deeper the network, the more abstract the representation will be and the more complex the task being learned can be, but at the cost of longer training (can be several hours) and need for larger training cohorts.

As a practical example, a generic network architecture (called VGG19) coming from the computer vision community interested in detecting object presence in photographs (*e.g.* is it a picture of a “cat?”), was successfully trained on high-resolution computed tomography (CT) lung imaging to assess if images are from patients with chronic pulmonary aspergillosis (CPA) (yes/no) [1]. Input data consisted of CT scans (pre-segmented to only show the lungs and transformed in maximum intensity projections to compress image information), and a label for each scan indicating if the subject has CPA or not. The pipeline in figure 1 enables sequence generation of “deep features” (set of values which “encode” the visual information contained in the input image) and the last blue layers perform the “classification” task, which returns as an output two numerical values (red and green dots) which are the probabilities of the subject with the input CT scan having CPA or not. Implementing such network architecture is technically made easier with open-source dedicated software libraries [2]. However, success in training the network to the specific classification task depends on two factors: 1) clinicians carefully supervising the composition of the image database used for training, to ensure a representative cohort of patients and balanced number of control cases, and avoiding biases in image quality (*e.g.* single scanner type, body mass index (BMI)); 2) data scientists carefully preparing the data to balance properly the number of examples used per class (*e.g.* balancing gender, scanner type, BMI), augment (*i.e.* transform) the cohort to reflect various sources of variability (*e.g.* rotating scans to simulate varying patient’s position in the scanner, re-scaling to simulate varying morphologies), and show robustness of the trained network for accuracy of classification decisions on an independent cohort of subjects.

Machine learning within respiratory medicine

The use of machine learning within healthcare is rapidly expanding, both within a research setting and, more recently, emerging within clinical practice. Respiratory medicine is no exception. To date, there have been numerous examples of integration of machine learning for prediction of sepsis, lung cancer prognosis

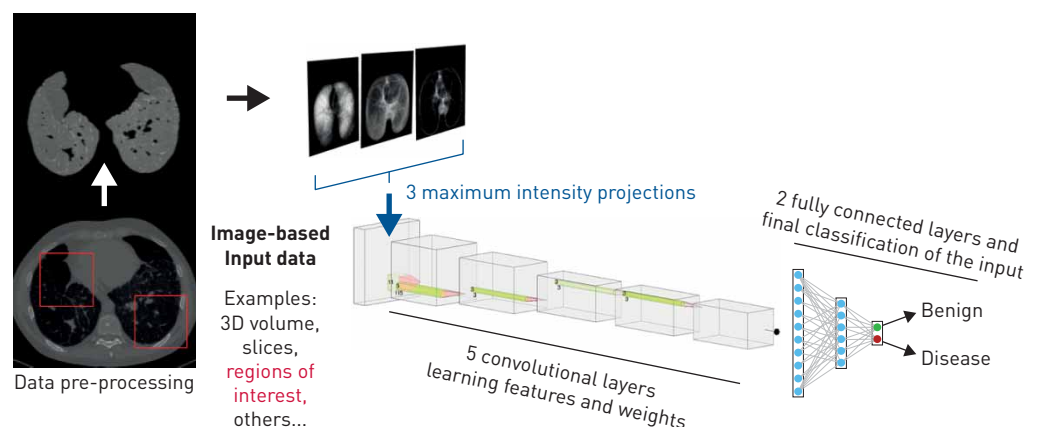


FIGURE 1 Typical architecture of a deep convolutional neural network for classification purpose. The input images enter the network on the left. Each grey box (deep features) represents a convolution layer made of “neurons” connected *via* weights. Small filters (yellow) perform convolution operations over the input (input or previous layer), attributing values to neurons of the next convolution layer, after integration of spatial information and simplification (red “operator”). Finally, the last “deep features” layer is vectorised and processed into a classifier made of fully connected layers to end up in two neurons (red and green dots) returning the probability values of the input images to be of subject type “benign” and “disease” (sum of green+red=1).

and risk of hospital admission with chronic obstructive lung disease [3–5]. The use of electronic health records (EHR) and wearable medical sensors has enabled access to large datasets that can be used to accurately phenotype disease, enable risk classification and predict treatment outcome [6, 7]. A recent study in *Nature Medicine* has proposed a data mining framework for EHR data that integrates prior medical knowledge and data-driven modelling. A DL system was built to extract clinically relevant information and subsequently establish a diagnostic system including respiratory disease (e.g. asthma) based on extracted clinical features which achieved accuracy comparable to experienced clinicians [8].

Perhaps the biggest area of expansion, however, within respiratory research has been the application of machine learning within imaging [9]. Image analysis using convolutional neural networks (CNN) is highly suitable for lesion detection, segmentation and classification. The lung, however, presents specific challenges. It is a deformable organ, with high levels of normal anatomical variability, complex parenchymal texture, multi-form lesions and often progressive degenerative disease. This requires an ability to identify lesions of variable shapes, sizes, intensity and with high variability of surrounding tissues. A further challenge relates to the large variability in image quality which is often scanner-specific and inflation level dependent. Nevertheless, significant progress has been made [10].

Within lung imaging, a large effort has been made in developing and applying computer-aided detection (CAD) systems. DL using chest radiographs has recently enabled classification of pulmonary tuberculosis with high accuracy [11]. The detection of lung nodules on CT is another promising field which is highly studied with the introduction of lung cancer screening programmes. Recent DL CAD approaches have reported promising performance [12]. Although the sensitivity and specificity of a DL CAD system for distinguishing solid nodules >5 mm (90.3% and 100.0%, respectively) and ground-glass nodules (100.0% and 96.1%, respectively) is getting close to that of double reading by independent radiologists, this dropped to 55.5% and 93%, respectively, when discriminating part solid nodules from solid ones [13]. Hence, further refinement is still required before implementation into clinical practice.

Further active fields of lung imaging research using DL include optimising airway segmentation, as well as exploiting very large cohorts such as COPD-Gene to perform genetic association studies to identify and characterise imaging phenotypes (*i.e.* genomic imaging) [14]. Within emphysema and interstitial lung disease, classic machine learning approaches have improved traditional visual scoring using *ad hoc* designed image-based features, to characterise 10 novel emphysema radiological subtypes [15], and enabled low-cost, reproducible, near-instantaneous classification of fibrotic lung disease with improved prognostication of mortality [16, 17].

With the emergence of large datasets including EHRs, radiographic imaging and wearable sensors, open source software tools, lower costs of computer hardware and graphics processing units, alongside community sharing of codes and pre-trained models, the impact of machine learning within respiratory medicine and healthcare in general is steadily increasing [6]. With this ability to generate increasingly large and complex datasets, an important consideration is the sharing of databases between institutions and accurate annotation to enable training of DL models [18, 19]. Quality labelling and annotation without compromising patient privacy could enable significant advancement. Integrating radiomic data with complementary omics datasets (e.g. genomics, proteomics) will enable a highly personalised diagnostic and therapeutic approach with recent advances in oncology including lung cancer highlighting this potential [20, 21]. This evolution, however, will create increasingly complex scenarios for clinicians and clinical trial methodological design.

Although machine learning is a powerful technique, it is understandably seen as a “black box” [22]. Machine learning can either be used to enable automation or as decision support. This distinction is critical when designing and incorporating a system. Clinicians and patients are understandably mistrustful of automation without explanation. “Explainable” AI has hence been a recent focus, as the long term success will likely depend on the ability of both patients and clinicians to understand and explain predictions or diagnosis [23]. An obvious solution therefore is to use machine learning algorithms as a decision support tool. The key to success, however, will likely be data visualisation, interpretability in a time-efficient manner and trust to enable a shared decision-making process with the patient [24].

Conflict of interest: None declared.

References

- 1 Zusag M, Desai S, Paolo MD, *et al.* SAPSAM – Sparsely Annotated Pathological Sign Activation Maps – A Novel Approach To Train Convolutional Neural Networks On Lung CT Scans Using Binary Labels Only. *IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. 2019; 298–302.
- 2 Wikipedia-contributors. Comparison of deep-learning software, 2019. https://en.wikipedia.org/w/index.php?title=Comparison_of_deep-learning_software&oldid=907831362

- 3 Barton C, Chettipally U, Zhou Y, *et al.* Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. *Comput Biol Med* 2019; 109: 79–84.
- 4 Burki TK. Predicting lung cancer prognosis using machine learning. *Lancet Oncol* 2016; 17: e421.
- 5 Min X, Yu B, Wang F. Predictive modeling of the hospital readmission risk from patients' claims data using machine learning: a case study on COPD. *Sci Rep* 2019; 9: 2362.
- 6 Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol* 2019; 20: e262–e273.
- 7 Ravi D, Wong C, Lo B, *et al.* A deep learning approach to on-node sensor data analytics for mobile or wearable devices. *IEEE J Biomed Health Inform* 2017; 21: 56–64.
- 8 Liang H, Tsui BY, Ni H, *et al.* Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat Med* 2019; 25: 433–438.
- 9 Saba L, Biswas M, Kuppli V, *et al.* The present and future of deep learning in radiology. *Eur J Radiol* 2019; 114: 14–24.
- 10 van Ginneken B. Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning. *Radiol Phys Technol* 2017; 10: 23–32.
- 11 Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 2017; 284: 574–582.
- 12 Ciompi F, Chung K, van Riel SJ, *et al.* Towards automatic pulmonary nodule management in lung cancer screening with deep learning. *Sci Rep* 2017; 7: 46479.
- 13 Li L, Liu Z, Huang H, *et al.* Evaluating the performance of a deep learning-based computer-aided diagnosis (DL-CAD) system for detecting and characterizing lung nodules: comparison with the performance of double reading by radiologists. *Thorac Cancer* 2019; 10: 183–192.
- 14 Bhatt SP, Washko GR, Hoffman EA, *et al.* Imaging advances in chronic obstructive pulmonary disease. Insights from the Genetic Epidemiology of Chronic Obstructive Pulmonary Disease (COPDGene) Study. *Am J Respir Crit Care Med* 2019; 199: 286–301.
- 15 Yang J, Angelini ED, Balte PP, *et al.* Unsupervised discovery of spatially-informed lung texture patterns for pulmonary emphysema: the MESA COPD Study. *Med Image Comput Comput Assist Interv* 2017; 10433: 116–124.
- 16 Jacob J, Bartholmai BJ, Rajagopalan S, *et al.* Mortality prediction in idiopathic pulmonary fibrosis: evaluation of computer-based CT analysis with conventional severity measures. *Eur Respir J* 2017; 49: 1601011.
- 17 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: 837–845.
- 18 Van Panhuis WG, Paul P, Emerson C, *et al.* better data integration and sharing can accelerate research and policy for global population health. *Am J Trop Med Hyg* 2015; 93: 205–205.
- 19 Tsopra R, Wyatt JC, Beirne P, *et al.* Level of accuracy of diagnoses recorded in discharge summaries: a cohort study in three respiratory wards. *J Eval Clin Pract* 2019; 25: 36–43.
- 20 Matsubara T, Ochiai T, Hayashida M, *et al.* Convolutional neural network approach to lung cancer classification integrating protein interaction network and gene expression profiles. *J Bioinform Comput Biol* 2019; 17: 1940007.
- 21 Huang Z, Zhan X, Xiang S, *et al.* SALMON: Survival Analysis Learning With Multi-Omics Neural Networks on Breast Cancer. *Front Genet* 2019; 10: 166.
- 22 Watson DS, Krutzinna J, Bruce IN, *et al.* Clinical applications of machine learning algorithms: beyond the black box. *BMJ* 2019; 364: l886.
- 23 Choi KS, Choi SH, Jeong B. Prediction of IDH genotype in gliomas with dynamic susceptibility contrast perfusion MR imaging using an explainable recurrent neural network. *Neuro-oncology* 2019; 21: 1197–1209.
- 24 Jamieson T, Goldfarb A. Clinical considerations when applying machine learning to decision-support tasks *versus* automation. *BMJ Qual Saf* 2019; 28: 778–781.