



SHAREABLE PDF

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>].

This single-page version can be shared freely online.

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.