SHAREABLE PDF

Machine learning can predict disease manifestations and outcomes in lymphangioleiomyomatosis

Saisakul Chernbumroong[1,2], Janice Johnson[3], Nishant Gupta[4], Suzanne Miller [3], Francis X. McCormack [4], Jonathan M. Garibaldi [2,5] and Simon R. Johnson [1,3,6]

Affiliations: [1]Nottingham Molecular Pathology Node, Nottingham, UK. [2]Advanced Data Analysis Centre, University of Nottingham, Nottingham, UK. [3]Respiratory Medicine and NIHR Biomedical Research Centre, University of Nottingham, Nottingham, UK. [4]Division of Pulmonary, Critical Care and Sleep Medicine, University of Cincinnati, Cincinnati, OH, USA. [5]School of Computer Science, University of Nottingham, Nottingham, UK. [6]National Centre for Lymphangioleiomyomatosis, Nottingham University Hospitals NHS Trust, Nottingham, UK.

Correspondence: Simon Johnson, Division of Respiratory Medicine and Biomedical Research Centre and Nottingham Biodiscovery Institute, University of Nottingham, University Park, Nottingham, NG7 2RD, UK. E-mail: simon.johnson@nottingham.ac.uk

ABSTRACT

Background: Lymphangioleiomyomatosis (LAM) is a rare multisystem disease with variable clinical manifestations and differing rates of progression that make management decisions and giving prognostic advice difficult. We used machine learning to identify clusters of associated features which could be used to stratify patients and predict outcomes in individuals.

Patients and methods: Using unsupervised machine learning we generated patient clusters using data from 173 women with LAM from the UK and 186 replication subjects from the US National Heart, Lung, and Blood Institute (NHLBI) LAM registry. Prospective outcomes were associated with cluster results.

Results: Two- and three-cluster models were developed. A three-cluster model separated a large group of subjects presenting with dyspnoea or pneumothorax from a second cluster with a high prevalence of angiomyolipoma symptoms (p=0.0001) and tuberous sclerosis complex (TSC) (p=0.041). Patients in the third cluster were older, never presented with dyspnoea or pneumothorax (p=0.0001) and had better lung function. Similar clusters were reproduced in the NHLBI cohort. Assigning patients to clusters predicted prospective outcomes: in a two-cluster model the future risk of pneumothorax was 3.3 (95% CI 1.7–5.6)-fold greater in cluster 1 than cluster 2 (p=0.0002). Using the three-cluster model, the need for intervention for angiomyolipoma was lower in clusters 2 and 3 than cluster 1 (p<0.00001). In the NHLBI cohort, the incidence of death or lung transplant was much lower in clusters 2 and 3 (p=0.0045).

Conclusions: Machine learning has identified clinically relevant clusters associated with complications and outcome. Assigning individuals to clusters could improve decision making and prognostic information for patients.