Commissioned Editorial for the accepted article: Accurate classification of secondary progression in multiple sclerosis using decision tree by Ramanujam et al, manuscript ID: MSJ-20-0721.R2

Towards an objective classification of multiple sclerosis

Arman Eshaghi[1,2], MD, PhD

1 Queen Square Multiple Sclerosis Centre, Department of Neuroinflammation, Queen Square Institute of Neurology, University College London, United Kingdom

2 Centre for Medical Image Computing, Department of Computer Science, University College London, United Kingdom

Correspondence address:

Russell Square House

10-12 Russell Square

London WC1B5EH

UK

Email: a.eshaghi@ucl.ac.uk

A more objective classification of multiple sclerosis (MS) has long been the aim of our field, or such wrote Charles Poser in March 1965: "[an aim of this study] was to establish a more objective system of scoring [and classification], based upon clinical signs and symptoms" [1]. This ambition, however, has remained elusive half a century later.

The classification of MS into clinically isolated syndrome, relapsing-remitting (RRMS), secondary progressive (SPMS) and primary progressive is based on a retrospective recollection of medical history. New descriptors of '*active'* and '*progressive'* introduced in 2013 provide a better understanding of the clinical status of each phenotype [2]. However, Lack of clear and objective boundaries or biological distinction across these phenotypes have introduced a misalignment in clinical practice and across research studies. This issue is also a challenge for pharmaceutical regulators as treatments for "progressive" MS are becoming increasingly available [3]. Such ambiguity has been more challenging in RRMS while transitioning to a progressive course because an "SPMS" label can drastically change how patients are treated.

In this issue of *Multiple Sclerosis Journal,* Ramanujam et al. introduce a machine learning model that endeavours to classify SPMS 'objectively' from a uniquely large dataset of health records [4]. In this work, the authors looked at several large datasets from Swedish and Canadian cohorts to develop and validate their newly developed model. Authors combine Expanded Disability Status Scale (EDSS), age, age at onset and disease duration in their model. The model was taught to predict whether the patient was *SPMS* or *RRMS* and was assessed against a "gold standard" classification assigned by the treating neurologist. Authors found that in the final iteration of their model using only EDSS and age, they could predict with an accuracy—defined as the concordance between model's prediction and the neurologists' classification—of between 82-89% depending on the dataset.

This work has several novelties. First, it uses an automatic classifier to bring a standardised classification system to future research studies. Second, this study uses large, collaborative, real-world cohorts of patients, upending the new era of big-data analytics in MS. Third, this study uses independent datasets to replicate and validate the results. This robust methodological design should set a standard in our field to test and externally validate newly developed models. However, the results should be cautiously interpreted, as this work has many limitations. First, the model was trained on the 'subjective' classification of neurologists; therefore, at best, this can standardise how patients are classified rather than providing a genuinely objective classification system. Second, the final prediction results set specific thresholds for EDSS and age, which were expected even before this study. Since SPMS is related to the change in EDSS, this might be perceived as a circular argument, even though external datasets have been used for validation. So, the main question remains, as where exactly machine learning can bring novelties to the field. Third, the authors rightly mention that the biological markers of the phenotype transition are missing in MS. However, the 'supervised' classification used in this study does not tell us anything about underlying biological underpinnings, which remain to be elucidated. Finally, it remains to be seen whether such models have any use in real-world clinical practice. This requires a prospective validation, in carefully designed clinical trials for artificial intelligence [5].

We are standing at the crossroad of history: the availability of large datasets and new digital technologies are coming together to change how we understand MS. This exciting study heads in the right direction but leaves an important question to be addressed in future: Can machine learning, after all, deliver on Poser's objectives?

Disclosure of conflicts of interest

I have no competing interest with respect to this work.

I have received speaker's honoraria from Biogen and At The Limits educational programme. I have received travel support from the National Multiple Sclerosis Society and honorarium from the *Journal of Neurology, Neurosurgery and Psychiatry* for Editorial Commentaries. I serve on the Editorial Board of *Neurology.*

References

1. **Poser, CM**. CLINICAL DIAGNOSTIC CRITERIA IN EPIDEMIOLOGICAL STUDIES OF MULTIPLE SCLEROSIS. *Annals of the New York Academy of Sciences*. 1965; **122**(1):506–519.

2. **Lublin FD, Reingold SC, Cohen JA, *et al.*** Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology*. 2014; **83**(3):278–286.

3. **Coetzee T**, **Thompson AJ**. Unified understanding of MS disease is required for drug development. *Nature Reviews. Neurology*. 2018.

4. **Ramanujam R**. Accurate classification of secondary progression in multiple sclerosis using decision tree.

5. **Cruz Rivera S, Liu X, Chan A-W, *et al.*** Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *The Lancet Digital Health*. 2020; **2**(10):e549–e560.