

Thursday, 22 October 2015

Dear Editors,

We are writing to ask that you consider our manuscript titled “**Does ensembling classifiers always confer a quantifiable advantage of selecting a single best classifier for medical diagnosis or prognosis?**” for publication in Bioinformatics.

Medicine has become a data-rich discipline. This opens the way for data-driven predictive algorithms to improve healthcare, for example through diagnostic or prognostic predictions for a given disease. Any systematic approaches that provide improvements to such predictions will lead directly to improvements in patient care. One such example is ensembling of multiple classification algorithms, to improve prediction quality. There is significant evidence in real-world applications such as data science competitions of the value of ensembling methods, but they remain an under-studied approach at a deeper level, and in specifically medical contexts.

This paper carries out a quantitative study of 9 different ensembling methods, each combining 11 different base classification algorithms. The ensembling methods are assessed on 10 different medical data sets, in each case providing separate test set validation of classification performance. We produce the surprising result that the gains from ensembling are overshadowed by the gains from simply using a subset of the available data to choose the best-performing single classifier for a given data set. We also confirm that there is no consistently superior single classifier (an expected result, due to the No Free Lunch theorem). Our results lead us to the interesting new hypothesis that the apparent benefits of ensembling which are seen in data science competitions may be more due to combining different feature representations, than combining multiple classifiers.

Thank you for your consideration,

Dr Richard Savage
Systems Biology Centre
University of Warwick, UK

Dr Richard Savage
Associate Professor
Warwick Systems Biology Centre
Coventry House
University of Warwick
Coventry CV4 7AL
United Kingdom
Tel: +44-24761-50254
Fax: +44-24765-75795
Email: r.s.savage@warwick.ac.uk