



# Quality and Quantity: The Future of Machine Learning for Health Research Appraisal

Yoah Sui<sup>1</sup> and Anna Rudkovska<sup>1</sup>

<sup>1</sup>University of Western Ontario

As the number of published primary academic literature is growing, the timely production of review papers is more important than ever. Between January 1<sup>st</sup>, 2017 and December 31<sup>st</sup>, 2017, a total of 322,732 journal articles were published on the PubMed database under the keyword “Health”, with 307,620 articles being published the year prior. Comparatively, only 42,797 and 41,443 review articles were published during those same time periods, respectively. Research from 2010 suggests that 75 trials and 11 reviews of trials are published each day [1]. This disparity between journal articles and review papers is important, since most review papers will include a method of appraising the quality (a.k.a. risk of bias) of each included study, and hence, are the primary means by which researchers evaluate the quality of a study – and by extension the state of research in a field [2]. Given that findings derived from a single study are rarely definitive, multiple replications followed by a review and appraisal of said primary studies is needed to establish the reliability of the findings [3]. These reviews are ultimately what help guide changes in policy, in practice, and in recommendations [4]. Recent drives for research informed clinical practice and policy setting has exponentially increased the demand for reviews that illustrate a clear relationship between healthcare inputs and outputs [3].

Unfortunately, this combination of high-volume research publication and reliance upon reviews for research appraisal raises two main issues: one, quality appraisal of new health research is limited by how quickly a review paper can be published; and two, new health research is likely being published more often than review papers that can assess their quality, given that traditionally, quality appraisal is conducted manually by a group of researchers using a pre-specified criteria. To this end, one major recommendation has been to reduce the number of “unnecessary trials” and prioritize systematic reviews [1]. However, this recommendation implies several obstacles and drawbacks, such as a standardized methods and training for writing (and evaluating) reviews, universal agreement in determining “unnecessary trials”, and ultimately resulting in less “new” research. Arguably, even if researchers were to collectively regress the amount of new research they would publish,

in favor for review papers, a considerable amount of time would still likely need to be dedicated to the task before all new research was appraised.

One alternative solution to this issue is machine learning. At its core, machine learning “automatically learns programs from data” [5]. More specifically, classification (*i.e.*, a subset of machine learning) takes inputs of data and outputs a discrete variable. For example, most malicious software, or malware, share 90-98% of code with previous iterations; hence, machine learning can identify this shared code, and classify a software as malware or not. Machine learning has already seen successful application in industries that involve huge amounts of data, such as spam filters, stock trading, and Web searches [5]. The main advantage to machine learning over manpower is obvious: the program/algorithm can process far more data than a human within a given timespan. However, machine learning also has advantages over manual programming, in that a machine learning algorithm is able to generalize from the data it is provided; in other words, the more data an algorithm is given to “learn from”, the stronger its classification becomes [5].

When applied to a health research appraisal context, machine learning offers an elegant solution to the issue of quantity. Firstly, established tools/instruments to measure the quality of a study available (*e.g.*, Cochrane Risk of Bias Tool [6]; Downs and Black Checklist [7]), and responses to individual items can be classified relatively easily (*e.g.*, yes/no, 0/1, low/high). This allows an algorithm to have clear-cut parameters in how to classify an item, and hence, how to appraise an article. Additionally, there is a gross amount of data through which an algorithm can learn from. Cross-referencing how previous studies have been appraised in review papers and how the algorithm classifies those same studies allows the accuracy of an algorithm to be constantly evaluated. The ability for the algorithm to learn from data also means minimal human oversight is needed after the initial algorithm is tested and successfully implemented. This automatic learning also minimizes the amount of potential human bias that is introduced when appraising articles. Finally, and perhaps most importantly, is the immense capacity for an algorithm to process data

(both new and existing), relative to a human researcher.

Despite these advantages, there are several barriers that present a challenge to implementing machine learning to a health research context, such as: determining appropriate and representative classifiers based off of existing instruments and tools, a standardized method of distinguishing good vs. bad classifiers, and optimization of the algorithm [5]. It is also worth noting that machine learning will not be ideal for assessing qualitative studies as well as studies that cannot be conducted as randomised control trials due to ethical or logistical considerations. Addressing these obstacles is imperative for machine learning to be effectively used for research appraisal purposes. However, when taken as a whole, machine learning presents an efficient and effective response to the issue of both quality-appraisal and quantity of appraisal in the ever-growing field of modern – and future – health research.

## REFERENCES

- [1] Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS medicine*. 2010;7(9):e1000326.
- [2] Glasziou P, Vandenbroucke J, Chalmers I. Assessing the quality of research. *Bmj*. 2004;328(7430):39–41.
- [3] Clancy CM, Cronin K. Evidence-based decision making: global evidence, local decisions. *Health affairs*. 2005;24(1):151–162.
- [4] Higgins JP, Green S, et al. *Cochrane handbook for systematic reviews of interventions*. 2006;.
- [5] Domingos PM. A few useful things to know about machine learning. *Commun acm*. 2012;55(10):78–87.
- [6] Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *Bmj*. 2011;343:d5928.
- [7] Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology & Community Health*. 1998;52(6):377–384.



Yoah received his B.Sc from the University of Waterloo, and his M.A. from Western University. He is currently a PhD candidate at Western University. His research focuses on the impact of excessive sedentary behaviour on subjective well-being in university students. He enjoys playing guitar, mountain biking, and chasing after his two dogs.



Anna received a B. Sc from the University of Toronto and a Masters degree in Rehabilitation Sciences from the University of Western Ontario. Currently, she is enrolled in her Doctoral studies at Western and focuses on women's reproductive health and autonomy.