Time and Cost Savings of Machine Learning (ML) and Artificial Intelligence (AI) in Systematic Reviews: A Case Study

Anne Littlewood[1], Mary Edwards[1], Lavinia Ferrante di Ruffano[1], Rachael McCool[1]

[1] York Health Economics Consortium, University of York, York YO10 5NQ

INTRODUCTION

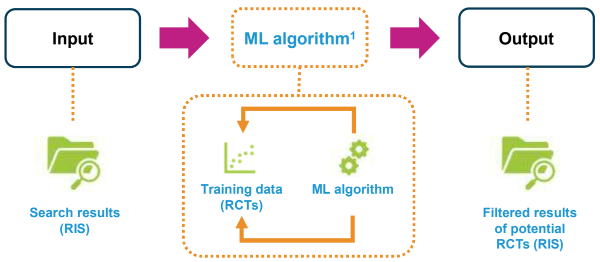
Conducting a systematic review (SR) of clinical trials is labour-intensive and expensive. The rapid and increasing rate of publication makes finding relevant clinical evidence challenging. SR teams must often screen thousands of references to identify eligible studies. Machine learning (ML), a subset of artificial intelligence (AI), has been shown to accelerate the SR process, most notably in the screening of search results for eligible studies for inclusion[1]. York Health Economics Consortium (YHEC) is a research consultancy, specialising in health economics research. YHEC conducted this study to understand whether introducing ML into the screening process can make SRs more cost-effective by reducing the time taken for review production. A bespoke tool (RESbot) was deployed during the screening phase of two review projects. The aim of this case study was to test the tool's performance, calculating the time saved.

METHODS

YHEC have created RESbot using Python, a stand-alone machine learning tool. RESbot was trained on an extensively tested open-source dataset produced by Cochrane for use in its own classifier. The tool identifies randomised controlled trials (RCTs) from a large corpus of records. The training dataset included 280,000 records, 20,000 of which were randomized controlled trials. RESbot was designed to be flexible and to work within a consultancy context. It has a user interface and inputs / outputs to fit into the company's existing workflow at any stage.

Figure 1 illustrates the machine learning screening workflow. Search results were fed into RESbot in research information systems (RIS) format and were processed by the ML algorithm using the training data to evaluate whether each abstract is likely to be an included study or excluded study. RESbot has two settings. The 'sensitive' setting identifies a higher number of possible RCTs with a lower risk of missing eligible studies, while the 'precise' setting is more focused.

Figure 1: ML screening workflow



CASE STUDY

We estimated a reduction in resource required for record screening in two examples of RCT-only reviews using both RESbot settings. Searches were run for two projects, one on postpartum depression and the other on renal denervation. The results were uploaded to RESbot for processing.

We estimated that an experienced systematic reviewer at YHEC could potentially screen 500 records per day at title and abstract stage. We used this figure as a basis to calculate the time saved by each RESbot setting over the two projects.

Table 1: Screening days saved

Project	Days taken (manual)	Days taken (RESbot sensitive)	Days taken (RESbot precise)	Days saved (RESbot sensitive)	Days saved (RESbot precise)
Renal denervation review	5.6	3.0	1.4	2.6	4.2
Postpartum depression review	4.3	3.2	2.5	1.1	1.8

RESULTS

The results of the case study are presented in Table 1 and Figure 2.

RENAL DENERVATION REVIEW

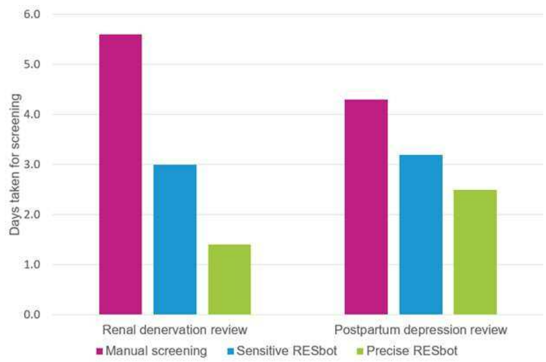
Once all databases were searched and records de-duplicated, the Renal Denervation review returned 2,795 records for screening. We estimated that manually screening the results would take 5.6 days. The sensitive RESbot setting reduced the number to screen by 1,290 records to 1,505, saving 2.6 days. The more precise RESbot setting reduced the number of records by 2,095 to 700, a saving of 4.2 days. No included studies were missed using the sensitive setting, one included study was missed using the precise setting.

POSTPARTUM DEPRESSION REVIEW

For the Postpartum Depression review, a record set of 2,153 was uploaded to RESbot after de-duplication. It would have taken 4.3 days to screen the records manually. The sensitive RESbot setting reduced the set to 1,615, a saving of 1.1 days. The more precise setting reduced the number of records by 882 to 1,271, which is a saving of 1.8 days. All eligible studies were found with both the sensitive and the precise settings.

Resource savings offered by machine learning vary depending on topic, but using RESbot may reduce the time taken to screen records by up to 46% with the sensitive screening option, with a subsequent reduction in cost to the organization commissioning the SR. The sensitive screening option found 100% of the studies included in each review, and the precise option found 98% of the studies across the two reviews.

Figure 2: Manual screening versus RESbot screening



CONCLUSIONS

The use of bespoke machine learning tools in review production has the potential to reduce the time and staff costs involved in producing a review. This case study tested the effect of on a small number of records but for larger reviews retrieving tens of thousands of records, reductions in time and costs could be very significant. Our findings support the deployment of ML tools to accelerate review production.

REFERENCES

1. Marshall *et al.* Machine learning for identifying randomized controlled trials: an evaluation and practitioner's guide. Research synthesis methods. 2018. 9(4), 602-614.

CONTACT US

anne.littlewood@york.ac.uk

+44 1904 323055

York Health Economics Consortium

www.yhec.co.uk

Providing Consultancy & Research in Health Economics