SWAR 25: A comparison of artificial intelligence (AI) aided and manual reviewing in abstract screening

Objective of this SWAR
This SWAR aims to answer following research questions:
1. What is the comparative performance of AI-aided tools versus manual reviewing by experts for screening abstracts for eligibility for a systematic review, and which AI tools demonstrate the highest level of performance among the tested tools?
2. What are the strengths and limitations of different AI methods in abstract screening?
3. What are the key characteristics and features of abstracts that are more likely to be misclassified by AI algorithms?

Study area: Abstract screening
Sample type: Reviewers
Estimated funding level needed: Low

Background
Systematic reviews provide an important source of knowledge for evidence-based decision making. However, they often demand substantial time and resources. Recently, the adoption of artificial intelligence (AI) has gained importance as a potential solution to enhance the efficiency of specific systematic review stages, including the automation of abstract screening, which identifies potentially relevant articles for subsequent evaluation. It is a critical component of the study identification phase for all systematic reviews.

A variety of AI-aided methods and tools have been used to support semi-automated abstract screening. These approaches include supervised classification models using machine learning techniques and web-based tools such as AbstrackR [1], Distiller [2], and EPPI Reviewer [3]. However, there remains a research gap pertaining to a comprehensive comparative analysis of all these methods within the context of interdisciplinary evidence synthesis.

This Study Within a Review (SWAR) aims to address this gap in the literature by offering a systematic and interdisciplinary evaluation of AI-facilitated abstract screening tools and machine learning models. Through this research, we aim to provide valuable insights into the potential advantages and limitations of these AI-driven methodologies within the domain of systematic reviews. Ultimately, our goal is to contribute to the refinement of evidence synthesis procedures, fostering efficiency in this essential process.

Interventions and comparators
Intervention 1: We will use different AI-aided screening methods for abstract screening, such as AbstractR and Distiller. The methods to be tested will be determined based on a scoping review we will undertake to identify the most frequently used AI-aided screening tools.
Intervention 2: Manual abstract screening results will be used as comparator.

Index Type:

Method for allocating to intervention or comparator

Outcome measures
Primary: Primary outcomes will be the following metrics:
Sensitivity: the ability of the screening method to correctly identify relevant abstracts. It is calculated as true positive / (true positive + false negative)
Specificity: the ability of the screening method to correctly identify irrelevant abstracts. It is calculated as true negative / (true negative + false positive)
Precision: the proportion of correctly identified relevant abstracts out of all abstracts identified as relevant. It is calculated as true positive / (true positive + false positive).
Accuracy: the overall correctness of the screening method. It is calculated as (true positive + true negative) / (true positive + true negative + false positive + false negative).

Secondary: We will also collect qualitative information from users of each of the tested tools. This will include aspects such as time efficiency and user-friendliness.

Analysis plans
We will begin with undertaking a scoping review to identify potential web-based AI-aided tools for abstract screening. Subsequently, we will proceed with the development of our own machine learning model by employing text feature engineering techniques.

Experts will be instructed to screen abstracts for potential relevance to the evidence synthesis question using AI-aided abstract screening tools identified through a scoping review. Results will be examined manually by experts. Results will be visualized using a confusion matrix for each tool, which will include true positives, true negatives, false positives, and false negatives. This will help to compare the primary outcomes across the tools tested versus manual screening results. Additionally, we will create a concise survey to be completed by experts who will be using AI-aided screening tools to capture information such as the time spent using the tool, ease of adaptation, and the learning curve.

Survey results will be used to compare the user experience of web-based AI tools and manual screening. By comparing expert assessments of the different screening methods (AI-aided versus manual), we will be able to evaluate their relative efficiency. Furthermore, we will analyse a diverse set of abstracts to identify patterns in language, style, and keywords that may impact the accuracy and efficiency of AI-aided screening. Subsequently, we will compare these patterns with the performance of AI algorithms in abstract screening to identify the most significant characteristics associated with misclassification.

Possible problems in implementing this SWAR
While some web applications are free or offer trial versions, they may have limitations or require paid subscriptions to provide their full functionality. These factors should be considered when budgeting for a review. Another important consideration is that if a review involves multiple reviewers using the web application, providing adequate training would be necessary to ensure consistent and reliable screening results. Additionally, not all web applications offer customer support and maintenance, so this should also be considered when selecting web applications for a specific study.

References
1. abstrackr: home. Available from http://abstrackr.cebm.brown.edu/account/login (accessed 3 October 2023).
2. DistillerSR. Available from https://www.distillersr.com/products/distillersr-systematic-review-software (accessed 3 October 2023).
3.EPPI-Reviewer Web (Beta). Available from https://eppi.ioe.ac.uk/eppireviewer-web/home (accessed 3 October 2023)

Publications or presentations of this SWAR design

Examples of the implementation of this SWAR

People to show as the source of this idea: Selin Akaraci
Contact email address: s.akaraci@qub.ac.uk
Date of idea: 28/FEB/2023
Revisions made by: Leandro Garcia, Maureen Dobbins, Mike Clarke, Morteza Namvar, Ruth Hunter, Selin Akaraci
Date of revisions: 30/MAR/2023