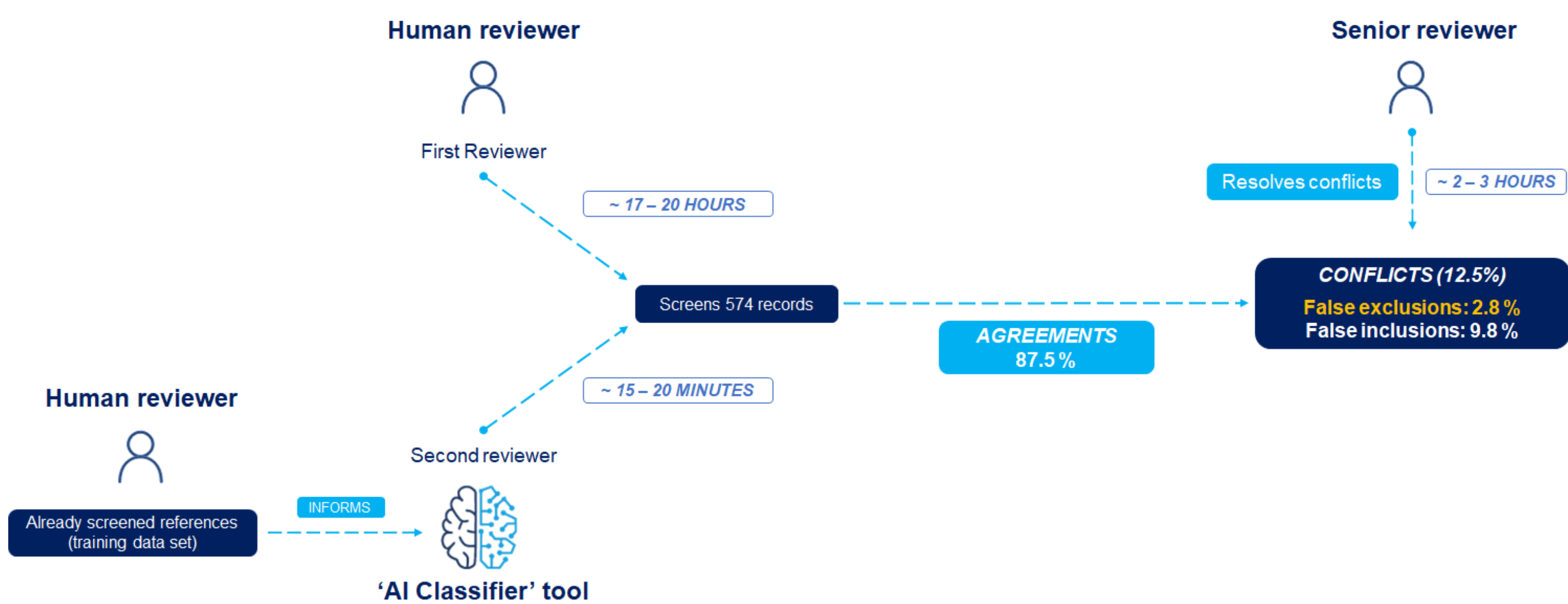


Introduction and objective

- Evidence synthesis is necessary to make informed decisions regarding patients and health policies. HEOR organizations employ literature reviews to summarize all available evidence for answering specific clinical question.
- Due to the exponential increase in publications, literature reviews are becoming increasingly difficult and time-consuming.
- The use of artificial intelligence and machine learning (AI/ML) in semi-automation of the conventional manual literature review process could expedite the process.
- In this study we explored the applicability of AI classifier tool of DistillerSR (Evidence Partners Inc. Ottawa, Canada) and validated its use in title-abstract screening.
- AI classifier is a tool that uses NLP (Natural Language Processing) to learn patterns from references’ titles and abstracts; and then compares those patterns with the human responses to build a model that can be applied to other references.

Graphical presentation



Methodology

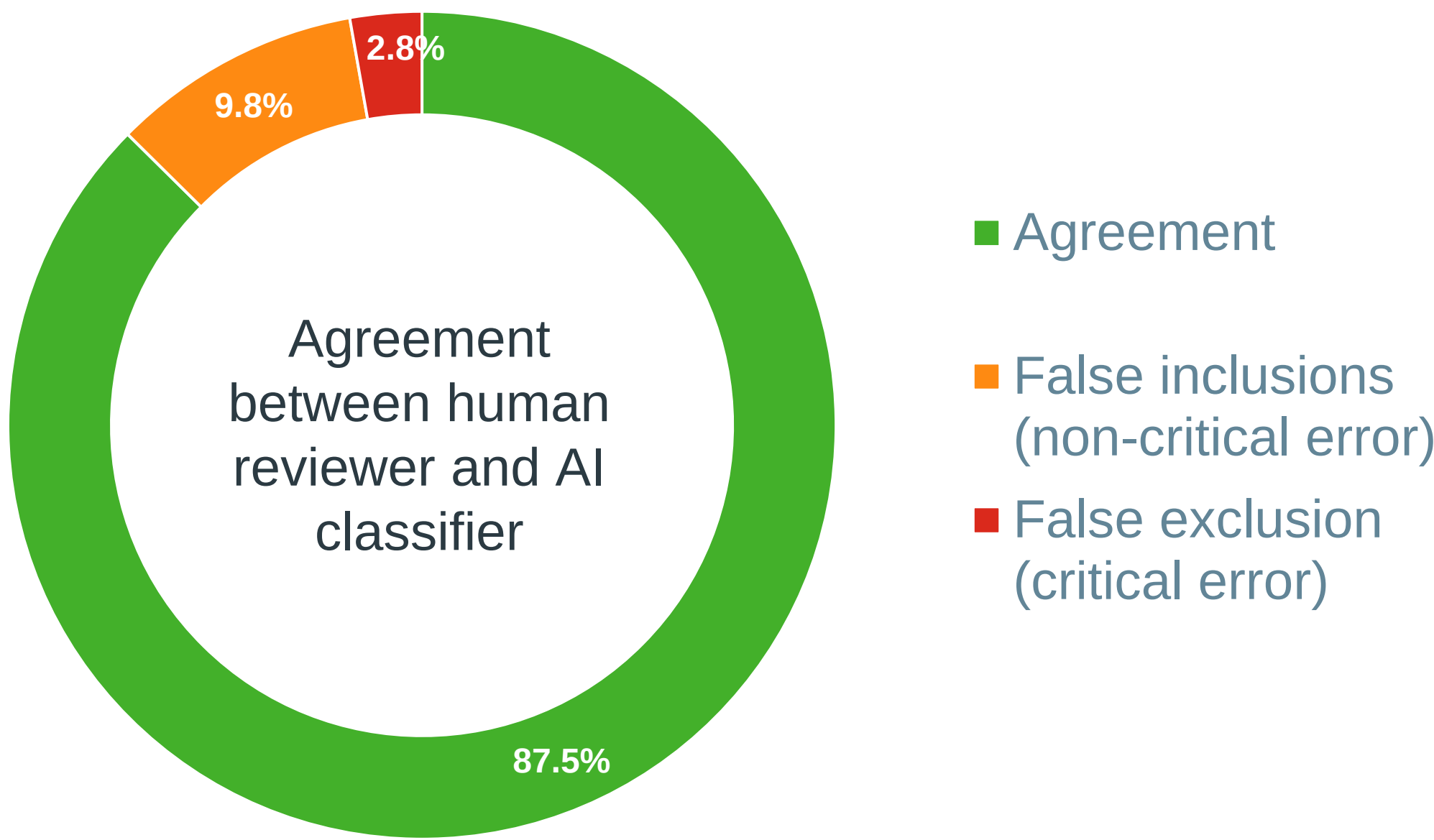
- In the double-review screening of a systematic review, the AI classifier tool was utilized as a second reviewer.
- A custom AI classifier was created and trained on set of references that were already screened by a human reviewer.
- The trained AI classifier was then utilized to classify unreviewed references as include or exclude.
- The new set of unreviewed references identified from database searches were reviewed by trained AI classifier.
- The AI classifier’s responses were compared to those of human reviewer.
- Any discrepancies and conflicts between human and AI classifier’s responses were resolved by an independent reviewer.

References:

1. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ open. 2017;7(2):e012545.
2. Wang Z, Nayfeh T, Tetzlaff J, O’Blenis P, Murad MH. Error rates of human reviewers during abstract screening in systematic reviews. PloS one. 2020;15(1):e0227742.
3. Wallace BC, Trikalinos TA, Lau J, Brodley C, Schmid CH. Semi-automated screening of biomedical citations for systematic reviews. BMC Bioinformatics. 2010;11(1):55

Results

- The custom AI classifier was trained and validated on the same training set of already reviewed references to get statistical insights that included a Balanced Accuracy Score of 0.75 (+/- 0.04), Recall Score of 0.86 (+/- 0.09), and F1 Score of 0.71 (+/- 0.08), where 0 represented lowest accuracy and 1 represented highest accuracy.
- The trained and validated classifier reviewed 574 references. Responses consisted of 16% includes and 84% excludes.
- On comparison with human responses, 87.5% of the AI classifiers’ responses tallied with human reviewer’s responses and 12.5% conflicts were reported.
- The conflicts were resolved by an independent reviewer in an estimated 2-3 hours and identified 2.8% false exclusions (considered as critical error) and 9.8% false inclusions (considered as non-critical error) by AI classifier.



- The total effort of manual double-screening of the 574 records was estimated to be 34-40 hours and utilizing AI classifier reduced the effort to 20-23 hours.
- Utilization of AI classifier can potentially save up to 50-60% of human hours, leading to significant productivity gains.

Limitations

- AI Classifier does not provide reason for exclusion while reviewing references which is typically required in reporting of systematic review as per the PRISMA guidelines.

Conclusion

- The AI classifier tool can be utilized for target literature reviews and scoping reviews, resulting in considerable reduction human effort.
- The limitations and few quality issues do not allow complete replacement of human efforts in literature reviews conducted for regulatory submission and reimbursement.