
Proceedings Paper:
Alharbi, E. and Stevenson, R. (2025) Assessing the impact of emerging relevant research on existing systematic review outcomes. In: Proceedings of HealTAC 2025. HealTAC 2025: 8th Healthcare Text Analytics Conference, Glasgow, United Kingdom, 2025-06-16 - 2025-06-18. HealTAC.

Assessing the Impact of Emerging Research on Systematic Reviews

Ebrahim Alharbi and Mark Stevenson

School of Computer Science, University of Sheffield, Sheffield, United Kingdom

1 Introduction

Systematic Reviews play a crucial role in synthesising and analysing scientific research to inform clinical decisions [1], but must be kept up to date to reflect the latest available evidence in order to avoid suboptimal clinical decisions or potential harm to patients [2]. Previous work found that there is no evidence that systematic reviews are updated when new relevant publications become available [3] and as many as 23% of reviews may be out of date within two years of their publication [4]. Automatic tools to assess the impact of new research on existing reviews can assist in deciding whether they need to be updated.

This study introduces the novel task of determining whether a publication strengthens or weakens the outcomes of an existing review. This information could help to assess whether a review remains consistent with current evidence or needs to be updated.

2 Data and Methods

The Cochrane Collaboration, an international network that conducts and maintains systematic reviews in the medical field [5], provided access to the Cochrane Library of Systematic Reviews[1]. This resource was used to create a dataset consisting of systematic reviews together with publications that appear in their updated versions and information about whether those publications strengthen or weaken the review's outcomes.

The following information was extracted from reviews for which an updated version is available: title, abstract, outcomes and associated meta-analyses of effect size. The updated version of the review was then examined to identify new publications that were not available when the original review was produced. For each of these, the following information was also extracted: title, abstract, publication year and estimate of the effect size. The effect size reported in each new publication is then compared against the original review's meta-analysis for each outcome to determine whether the information it contains is consistent within the original review's meta-analysis and each outcome-publication pair labelled accordingly. If the effect size reported in the new publication for a particular outcome falls within the confidence interval for the same outcome in the original review, that publication is considered to strengthen that outcome; otherwise,

[1] https://www.cochranelibrary.com/

it is assumed to weaken it. The resulting dataset contains 17,679 review outcome-publication pairs, each labelled as either strengthen or weaken.

A set of classifiers was then developed to automatically determine whether new relevant publications supported or weakened an existing review outcome. Multiple approaches were compared: traditional machine learning models (i.e., Naive Bayes, logistic regression, and SVM) using text from the review and new publication weighted using Term Frequency-Inverse Document Frequency (TF-IDF) scores. Additionally, transformer-based models (including BERT, BioBERT, PubMedBERT, BlueBERT, SciBERT, and DeBERTa v3 base) were fined-tuning using consistent hyperparameters (learning_rate=2e-5, batch_size=16, and epochs=10). A simple baseline approach of choosing the majority class (Strengthen) was also implemented. Approaches were evaluated using standard metrics: accuracy, Precision (P), Recall (R), and F1 score macro averaged across classes.

3 Results

Results are shown in Table 1. All approaches outperform the baseline. However, the accuracy for some models, particularly those based on traditional machine learning, is not much higher than achieved by simply choosing the majority class. The transformer-based approaches outperform those based on traditional machine learning for all metrics, with SciBERT producing the best overall performance.

	Accuracy	**Precision**	**Recall**	**F1**
Majority class	0.5787	0.2894	0.5000	0.3666
Naive Bayes	0.6313	0.6223	0.6219	0.6221
Logistic Regression	0.6609	0.6514	0.6466	0.6479
SVM	0.6581	0.6486	0.6447	0.6458
BERT-base-uncased	0.7010	0.6935	0.6900	0.6913
BioBERT-v1.1	0.7024	0.6958	0.6971	0.6963
PubMedBERT-base-uncased	0.6945	0.6884	0.6905	0.6891
DeBERTa-v3-base	0.6764	0.6693	0.6702	0.6697
BlueBERT-PubMed	0.6931	0.6857	0.6852	0.6854
SciBERT-uncased	**0.7120**	**0.7052**	**0.7053**	**0.7052**

Table 1: Model Performance reported as accuracy and (macro-averaged) Precision/Recall/F1.

4 Conclusion and Future Work

This research introduces a novel task for automatically determining the impact of new relevant publications on existing systematic reviews. We described the methodology used to create a dataset and the development of classifiers. Transformer-based models, particularly SciBERT, provided the best results.

Future work should explore whether the performance of these models is accurate enough for them to be useful to inform review update decisions. In addition, decoder models, such as GPT-4 and Claude 3, could also be applied to this task.

Study context

The data used in this research were sourced from the original and updated versions of systematic review files provided under license by the Cochrane Collaboration. We gratefully acknowledge their support.

References

[1] Knezevic NN, Manchikanti L, Hirsch JA. Principles of Evidence-Based Medicine. In: Essentials of Interventional Techniques in Managing Chronic Pain. Springer; 2024. p. 101-18.

[2] Chandler J, Cumpston M, Li T, Page MJ, Welch V. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken: Wiley. 2019;4.

[3] Bashir R, Surian D, Dunn AG. Time-to-update of Systematic Reviews Relative to the Availability of New Evidence. Systematic Reviews. 2018;7:1-8.

[4] Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How Quickly do Systematic Reviews Go Out of Date? A Survival Analysis. Annals of Internal Medicine. 2007;147(4):224-33.

[5] Cumpston M, Flemyng E. Chapter IV: Updating a Review. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions; 2023. .