Update on the LiveSTART™ Artificial Intelligence (AI) Systematic Literature Review (SLR) Tool – Can It Aid in Limited Data Extraction for LiveRef™?

Poster #MSR8

Rozee (Junhan) Liu[1], Reza Jafar[2], Lee Ann Girard [3], Kristian Thorlund[4], Anna Forsythe[4]

[1]Cytel, Inc., Toronto, ON, CA; [2]Cytel, Inc. Vancouver, BC, CA; [3]Cytel, Inc. Montreal, QC, CA; Cytel, Inc. Waltham, MA, US

Background

- Systematic literature reviews (SLR) are labor intensive and time consuming; however, they are required for submissions to regulatory and health technology assessment (HTA) agencies.
- Numerous references inform a global value dossier (GVD), including data from SLRs and grey literature, as well as information from hand searches.
- The application of LiveSTART™ in title abstract review stage of an SLR has been previously published.[1] LiveSTART™ has the potential to be used for LiveRef™ limited extraction to increase the efficiency of a GVD update.

Objective

- The objectives of this research were to:
  - Validate the LiveRef™ machine learning (ML)-assisted limited extraction accuracy
  - Evaluate the change in efficiency in the case of a GVD update

Methods

Annotated dataset preparation for ML training

- Data were extracted manually by two independent, experienced analysts.
- Seventy-eight datasets containing extractions from 6,377 congress abstracts and curated library references were used to train (4,782) and validate (1,595) an artificial intelligence (AI) model in LiveSTART™ for data extraction.

ML prediction structure

- Objective data including indication, population, country, study category, study design, treatment products, sample size, data source, and reported variables were prepared.
- Subjective interpretation as quantitative and qualitative summaries was also included.

Accuracy and Hamming score

- The accuracy and Hamming scores were calculated.
- Accuracy is a measure of concordance between prediction and actual results for attributes that are either predicted correctly or not.
- Hamming score is used for predictions that include multiple components; therefore, a percentage of prediction compared with the total accurate results are weighted.

Results

- LiveRef™ is a continuously updated, web-based, library of indication-specific publications reporting data on epidemiology, disease burden, treatment practices, and comparative effectiveness (Figure 1).

Figure 1. LiveRef GVD Library

- The LiveRef™ auto-extraction template is shown in Figure 2, where only three attributes were input for the ML algorithm: the title, the abstract, and the authors and their affiliations (optional). References that were entered into the template included relevant data from Embase, MEDLINE, and Cochrane databases, scientific congresses, trial registries, regulatory and HTA websites.
- An example of the LiveRef™ auto-extraction output (formatted) is shown in Figure 3.

Figure 2. LiveRef™ auto-extraction template

	A	B	C
1	Title	Abstract	Authors and Affiliations
2			
3			
4			
5			
6			
7			

Please copy the title and abstract columns from your scraping file. If the Authors and Affiliations information is also available in your scraped file, please copy it here, if not, you can leave the column empty.
These three columns are mandatory but you can include other columns in your dataset.

Figure 3. LiveRef™ auto-extraction output

- Using 1,595 human-annotated references, LiveRef™ auto-extraction accuracy (either predicted correctly, or not) and Hamming score (weighting the percentage of accurately predicted components that are correct) are presented in Table 1.

Table 1. LiveRef™ auto-extraction accuracy and Hamming score

Outcomes predicted by ML algorithm	Accuracy	Hamming score
Indication	0.86	**0.90**
Category of evidence	0.78	**0.84**
Study type/sub-category	0.39	**0.53**
Products	0.37	**0.65**
Sample size	**0.76**	NA
Data variables evaluated	0.17	**0.50**
Population	0.59	**0.64**

Abbreviations: ML, machine learning; NA, not applicable

- The LiveRef™ auto-extraction tool extracted data from 1,595 abstracts in approximately 30 minutes.
- A manual GVD update with a hypothetical 1,595 references will take approximately 265 hours.

LiveRef™ GVD update

264.5 hours (99.8%)

Reduction in time compared with manual extraction

With the addition of the LiveSTART™ title abstract review tool, a total of **3.29 weeks** could be saved in a GVD update project.

Discussion

- Even with the weighted accuracy using Hamming scores, LiveRef™ auto-extraction's ML algorithm is still performing at an average level for study type/sub-category, products, and data variables evaluated.
- This low accordance is likely because 1) the limited number of datasets only based on congress abstracts, resulting in a relatively homogenous collection for training; and 2) the expected outcomes for the above-mentioned attributes are not normalized. For example, sub-category should be predicted based a pre-defined limited list for ML to choose from, and not on an unlimited number of categories.
- The next step in LiveRef™ auto-extraction would be data normalization.

Conclusion

- With the combination of machine-assisted title and abstract review, and data extraction, LiveSTART™ AI could potentially yield comparable accuracy with considerable time-savings in an SLR project.
- However, this method will require the endorsement of regulatory and HTA agencies before its full potential can be achieved.

References

1. Liu J, Jafar R, Girard LA, Thorlund K, Forsythe A. (2022) Can Artificial Intelligence (AI) Replace a Human Reviewer in Systematic Literature Review (SLR)? Validation of the LiveSTART™ Tool. Value In Health 25(S12): S364. ISPOR Europe, Vienna, Austria. November 2022. https://www.ispor.org/heor-resources/presentations-database/presentation/euro2022-3566/121569

Presented at ISPOR US | Boston, MA | May 7–10, 2023

Contact information: www.cytel.com