

DEVELOPMENT OF ADVERSE OUTCOME NETWORK WITH THE EMPLOYMENT OF NLP: THE LIVER FIBROSIS PARADIGM

A. Karakoltzidis^{1,2}, I. S. Frydas^{1,2}, S. Karakitsios^{1,2}, D. A. Sarigiannis^{1,2,3,4*}

¹HERACLES Research Center on the Exposome and Health, Center for Interdisciplinary Research and Innovation, Aristotle University of Thessaloniki, Thessaloniki, Greece

²Environmental Engineering Laboratory, Department of Chemical Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

³University School of Advanced Study IUSS, Pavia, Italy

⁴National Hellenic Research Foundation, Athens, Greece

(*sarigiannis@auth.gr)

ABSTRACT

This research introduces an innovative methodology for developing Adverse Outcome Pathways (AOPs), employing text mining with a focus on liver fibrosis as the target adverse outcome. The approach integrates advanced transformer models for biomedical text mining, incorporating two distinct methodologies for topic modeling—specifically, Latent Dirichlet Allocation (LDA) and BERTopic. The use of these two algorithms enables cross-validation between the topics derived from each method, enriching the comprehension of intricate relationships within biological processes, and providing a systematic, data-driven approach to AOP development. The initial phase involved a literature review to identify keywords associated with liver fibrosis, including terms such as Cytokines, Ehr, and chemokines. A comprehensive search on PubMed and Web of Science yielded almost 196,000 abstracts, and after removing duplicates, 191,029 abstracts remained for analysis. These abstracts were then organized into a tabular format, tokenized, and cleaned by eliminating stop words and applying lemmatization. Subsequently, the results of topic modeling were represented in a network format, facilitating the identification of nodes and connections between significant factors linked to liver fibrosis. Examples include liver diseases such as viral hepatitis B and C, inflammation, exposure to environmental chemicals, oxidative stress, and cellular death. Notably, literature reports consistently highlight the co-occurrence of liver fibrosis and insulin resistance in patients. The identified topics serve as a foundation for refining the literature review, strategically constructing an AOP. The next phase involves gathering new articles based on keywords identified through topic modeling. These collected publications will undergo two distinct text mining methods: biological entity recognition and biological entity extraction. The primary objective is the systematic construction of an AOP for liver fibrosis through a bottom-up approach. Concurrently, articles categorized with significant topics relevant to the study will be incorporated into the subsequent analysis. It is crucial to emphasize that the subsequent step entails reapplying topic modeling to the newly collected articles, preserving the relevant ones, and progressing to a comprehensive full-text analysis.

KEYWORDS: NLP, Topic Modelling, Liver Fibrosis, Liver Diseases, Exposure

REFERENCES