Curation at Scale: Computational approaches to support curation in LitCovid

Qingyu Chen*, Alexis Allot*, Zhiyong Lu†

*Equal contributions. †Corresponding author. Email: zhiyong.lu@nih.gov; Tel: 301-594-7089

We recently developed LitCovid [1], a daily curated literature hub, to help scientists, healthcare professionals, and the general public stay informed with the latest research on COVID-19 and SARS-COV-2. Since its public release, LitCovid has been accessed over 10 million times by its users from ~200 countries for various kinds of information needs, such as evidence synthesis [2], drug repurposing [3], and guidelines for diagnosis [4]. Initially, all data collection and literature curation were done manually with little machine assistance. However, as the outbreak evolved, the number of publications grew rapidly, at a rate of ~10,000 articles per month, making manual curation increasingly challenging. In response, we developed and evaluated several artificial intelligence (AI) and Machine Learning (ML) algorithms to assist manual curation, and improve the efficiency. In this work, we will describe how cutting-edge AI/ML research is used in different stages of our pipeline, from accurately classifying and categorizing relevant articles to extracting entity names from free text. We will also describe the remaining challenges for future research, such as the use of LitCovid's data to extract informative key terms, create COVID-19 knowledge graphs and assist drug repurposing [5].

Acknowledgement: This research is supported by the NIH Intramural Research Program, National Library of Medicine.

[1] Q. Chen, A. Allot, and Z. Lu, "Keep up with the latest coronavirus research," *Nature*, 2020.

[2] M.-L. Bonte, L. Bourgeois, G. Lenaerts, J. Verbakel, and C. Matheï, "Best Evidence Topic Report 23."

[3] S. Chakraborti, and N. Srinivasan, "Drug Repurposing Approach Targeted Against Main Protease of SARS-CoV-2 Exploiting 'Neighbourhood Behaviour'in 3D Protein Structural Space and 2D Chemical Space of Small Molecules," *Preprint. https://doi. org/10.26434/chemrxiv,* vol. 12057846, pp. v1, 2020.

[4] A. Bhimraj, R. L. Morgan, A. H. Shumaker, V. Lavergne, L. Baden, V. C.-C. Cheng, K. M. Edwards, R. Gandhi, W. J. Muller, and J. C. O'Horo, "Infectious Diseases Society of America guidelines on the treatment and management of patients with COVID-19," *Clinical Infectious Diseases*, 2020.

[5] L. Yeganova, R. Islamaj, Q. Chen, R. Leaman, A. Allot, C.-H. Wei, D. C. Comeau, W. Kim, Y. Peng, and W. J. Wilbur, "Navigating the landscape of COVID-19 research through literature analysis: A bird's eye view," *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2020.