Search engines and social media can forecast disease outbreaks

July 10 2023

Credit: Pixabay/CC0 Public Domain

Internet search engine queries and social media data can be early warning signals, creating a real-time surveillance system for disease forecasting, says a recent University of Waterloo study.

Using the example of COVID-19, researchers found there was an association between the disease's prevalence and search engine queries and social media posts. The study, "Digital Disease Surveillance for Emerging Infectious Diseases: An Early Warning System Using the Internet and Social Media Data for COVID-19 Forecasting in Canada," was published in the Ebook Series *Studies in Health Technology and Informatics*.

"The general public tends to use internet searches and social media for health information, and especially so during global epidemics," said Dr. Yang (Rena) Yang, a postdoctoral research fellow in the School of Public Health Sciences at Waterloo. "These behavior patterns can be used by public health authorities to develop a real-time surveillance system to flag when diseases are spiking or waning or respond quickly to emerging infectious diseases."

The team extracted symptom keywords from Google Trends and Twitter data in Canada from January to March 2020. These keywords included cough, runny nose, sore throat, shortness of breath, fever, headache, body ache, and fatigue on Google Trends. On Twitter, researchers looked at COVID-19-related hashtags, such as pneumonia, cough, fever, running nose and breath. They then cross-checked the information against COVID-19 data from the COVID-19 Canada Open Data Working Group.

The researchers found that search terms related to COVID-19 symptoms strongly correlated with daily COVID-19 cases with a time lag of between one and 13 days, suggesting that these tools can serve as early warning signals for digital disease surveillance in real time. The sophisticated machine learning model used for forecasting in this study performed better with Google Trends than with Twitter data.

Dr. Zahid Butt, lead investigator of the study and an assistant professor in the School of Public Health Sciences at Waterloo, noted there are challenges in modeling due to the noise from self-generated data, not to mention the ability to identify relevant keywords of an emerging infectious disease.

"Our future research will aim to systemically identify and organize pertinent symptom keywords for emerging diseases, even before they are commonly recognized or reported," Butt said. "These systems have the

potential to assist in epidemiological control and monitor public perceptions of the disease, as well as forecast trends in outbreaks. A multifaceted strategy that uses multiple data sources and multimodal modeling would help provide accurate and comprehensive emerging disease surveillance."

  More information: Yang Yang et al, Digital Disease Surveillance for Emerging Infectious Diseases: An Early Warning System Using the Internet and Social Media Data for COVID-19 Forecasting in Canada, *Studies in Health Technology and Informatics* (2023). DOI: 10.3233/SHTI230290

Provided by University of Waterloo