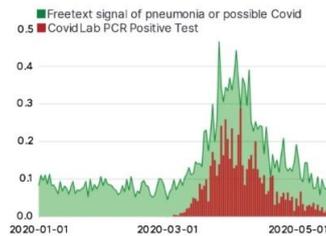

NLP 'Barometer' for Analysing Local Clinical Data



A vendor-neutral, open-source, natural language approach of EHR data is suggested by a team of U.K. researchers, to help providers plan for surges in patients, particularly in the COVID-19 context.

You might also like: The WHO has developed a number of surge planning tools, which help to visualise acute and intensive care capacity needs over time, identify the timing and severity of the peak of the outbreak, and plan human resources for health systems. [Learn more](#)

While natural language approach, mainly to social media content, has been previously used for epidemiologic forecasting, this proved to be susceptible to distortions and keyword spamming in the uncontrolled online environment. A new study (Teo et al. 2021) reports on the results from a different approach. Instead of analysing publicly available data, the researchers focussed on private health data lakes, i.e. unstructured, freetext data from EHR systems of two large U.K. hospitals.

The data from both systems were pooled in two separate data lakes using the CogStack platform and treated as “bags of words” to identify symptom keywords and phrases suggestive of COVID-19, such as ‘dry cough’, ‘pyrexia’, ‘fever’, ‘dyspnoea’, ‘anosmia’, etc. (to avoid the distortion due to the ‘hashtag effect’, the word ‘COVID’ was excluded from the index of signal). The results showed that these signals closely tracked the gold-standard data of tests of Covid-19 positivity (nasal swab PCR) in both hospitals, with up to four-day head-start.

Another finding indicated that as the pandemic developed and providers, through general media, became more aware of specific COVID-19 symptoms, such as anosmia, the incidence of relevant phrases increased, affected by recall bias. The authors point out that this phenomenon should be accounted for when deploying their approach.

The GogStack platform is [open-source](#) on Github and available for any healthcare organisation. The researchers note that its implementation is low-cost, flexible, EHR-vendor-neutral and does not interfere with clinical routines. Despite the scope of the current study being limited to closed health data lakes, the authors emphasise the scalability of their approach to regional and even national level for short-term patient surge forecasting.

Image credit: Teo et al. (2021)

Published on : Sat, 27 Feb 2021