# **Precision Public Health: Application of AI to social media to characterize information dissemination on coronavirus disease 2019 (COVID-19)**

## **Background**

In early December 2019, patients with pneumonia of unknown origin were identified in Wuhan city, China. The causative agent was identified as a member of coronavirus family, initially named 2019 novel coronavirus (2019-nCoV), and associated disease as coronavirus disease 2019 (COVID-19). Initial cases of the disease presented with acute severe respiratory syndrome starting with fever and respiratory symptoms. However, a spectrum of illness from mild to severe disease has been described in the more recent reports. Emerging data suggests that about 15% of patients require critical care and case fatality rate is about 1-2%.

Sustained human-to-human transmission has been confirmed in China. Since its emergence, the virus has spread to 25 countries outside of China with limited human-to human spread in various settings.

On January 30, the World Health Organization declared COVID-19 a public health emergency of international concern. As of February 13, 2020, about 47,054 laboratory-confirmed cases have been documented globally, 99% of these are in China, mainly in Wuhan, Hubei province.

Since the initial reports of cluster of acute severe respiratory disease (COVID-19) and the potential for global spread, there has been widespread discussion and dissemination of information through social media. A lot of the information circulating on social media is false, misleading, and alarmist. The nature of some of the information and conversations on social media has also taken a discriminatory tone. This kind of information could not only hamper response efforts, but has also increased anxiety and panic about the risk of infection, risk of severe disease, prevention, and treatment. The feeling of anxiety in the population, in part can be assessed by increased demands and shortages for face masks and hand sanitizers even in countries such as Canada where only 7 cases have been confirmed to date.

Misinformation also affected Ebola outbreak. The identification, characterization, and tackling of misinformation on social media during the Ebola outbreak contributed to the successful response in West Africa. However, since the Ebola outbreak, the scale of misinformation has grown significantly. While the application of artificial intelligence (AI) techniques are relatively new to public health, they could contribute to rapid characterization and the tackling of misinformation through text analytics.

## **Aims**

We propose to (a) characterize misinformation on social media related to coronavirus, its origins, and its propagation, (b) develop a platform aimed at monitoring misinformation and (c) develop strategies and systems to tackle misinformation.

## **Approach and Plan**

We will assemble a team including public health, AI text analysis, communication, and planning experts to identify and refine needs and guide implementation team in their analysis.

We will collect tweets from twitter using various hashtags used along with COIVD-19 tweets. To assess changes in the information spread over time, we will collect tweets starting mid-November, about 3-4 weeks before any news related to COVID-19.

To characterize misinformation related to COVID-19, we will apply AI techniques to Twitter data. We will use previously established and novel approaches for text analysis to parse out topics, aspects, and sentiments (positive or negative emotions) attached to these aspects through aspect base sentiment analysis and topic modelling. The results of these analyses will be refined by manual review to refine algorithms and to provide the intended results in terms of identification and classification of information. In addition, we will use discourse parsing at the conversation level in order to understand underlying conversation structure to avenues for interventions. We will also track evolution of information content and varsity over time and across geographic areas.

Based on the above analysis, we will use twitter stream data and to apply algorithm to identify tweets spreading false information. This could be used to tackle information in real time.

## **Potential outcomes**

This project will characterize misinformation on social media, in particular on Twitter, with the aim to develop an automated system to characterize and identify the spread of false information on Twitter. The system could allow public health practitioners to develop counter messaging to alleviate anxiety and decrease the spread and impact of false information.

## **Team**

Naveed Janjua, MBBS, DrPH, Executive Director Data and Analytic Services, BCCDC, expertise in epidemiology, public health and data

Giuseppe Carenini, PhD, Professor, Computer Science Department, UBC & expert in natural language processing and text analytics

HyeJu Jang PhD, Postdoc Fellow BCCDC, UBC Computer sciences, expert in text analytics

Katie Fenn, Director Quality and Safety, BCCDC

Emily Remple**,** Knowledge Translation Lead BCCDC

## **Resources needed**

In order to reach our objectives, we will require resources for the following activities:

* Twitter data feed
* Computer programmer
* Subject matter expert (public health /communication) for data annotation
* Review of results