

PREDICTION OF COVID - 19 EPIDEMIOLOGY USING THE DATA FROM SOCIAL MEDIA - A DESCRIPTIVE BIBLIOGRAPHIC ANALYSIS.

Niranjan Muralikrishnan

Article Info

*Corresponding Author:

Niranjan Muralikrishnan

Abstract

Background: The COVID - 19 pandemics has led to many kinds of basic and social research. This has served as a platform for the scientists to extensively mine the social media data and to identify and predict the potential disease hot-spots. **Methods:** Google scholar and Pubmed search were used to identify the articles which has used social media data to project the hot-spots. Search period window was one month and only open access / free to access articles were included for the study. Standardised search terms were used by a team of two researchers. **Results:** A total of 15 articles were selected and screened for the inclusion criteria. Then 12 articles which met the inclusion criteria were selected for this study. The highest read and cited articles were published from USA and China respectively. Even though Canada has been acknowledged as the country with highest social media usage the research with such data has to be given some impetus. **Conclusion:** Usage of social media data for predicting caseloads can significantly reduce the morbidity and mortality due to COVID - 19 which is relevant in these times of minimal digital divide around the globe.

Introduction:

The global emergency of COVID 19 or otherwise the SARS - CoV - 2 which had its genesis roots from the Republic of China was declared as a "Global Pandemic" by the World Health Organisation (WHO) because of its rapid spread and high case fatality. Deeper investigations on the disease-causing agent have revealed that the novel coronavirus is more than 90% identical to the coronavirus present in bats at the genome level.(1) It affects the host by causing severe local and systematic damage to the tissues through activating the immune response which causes a more inflammatory reaction within the human body and provides the impetus for the tissue damage.(2) This disease has been perceived to highly fatal because of the above mentioned characteristic and also the disease spread rate which is technically known as "Basic Reproduction Number" was also found to be around 2.0 to 2.5, which is roughly translated to disease can spread easily and rapidly to at least more than two healthy persons when they come into contact with one infected.(3)

To curtail this fast-spreading pandemic, policymakers, and people's elected representatives came with different Public Health measures such as school closures, work from home, the introduction of strict quarantine protocols, and lockdown. All these measures were enforced to cut the transmission line of the disease in turn reducing the caseloads.(4) But in contrast, the lockdown brought a huge magnitude of social isolation among the community which forced them to consume media through the world wide web all around the globe since it was the only platform available easily to socialize and to exchange ideologies.(5) A study done in Italy has revealed that there was a surge in social media usage among people who were 18 to 35 years of age to alleviate social isolation.(6) So this surge in usage of the internet has resulted in seeking information from social media more when compared with other types

PREDICTION OF COVID - 19 EPIDEMIOLOGY USING THE DATA FROM SOCIAL MEDIA - A DESCRIPTIVE BIBLIOGRAPHIC ANALYSIS.

of sources.(7) This acts as a stage for the people who are experiencing COVID 19 symptoms will showcase it on social media to an extent that active surveillance data and social media data were yielding similar results in predicting the epidemic successfully.(8) So this study tries to explore the prevailing culture of predicting the COVID 19 epidemiology through social media data on a global scale and analyses the need for such research.

Materials and Methods:

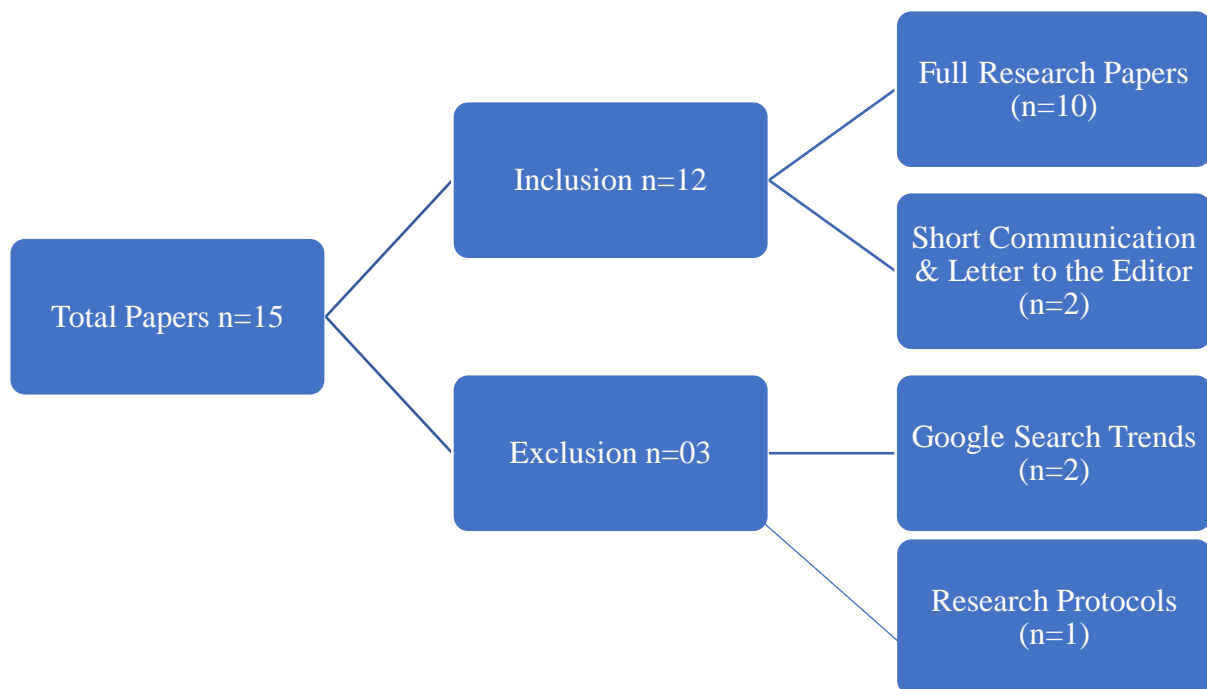
The Google Scholar and Pubmed search was used to find out the studies which fulfilled all the essential inclusion criteria. The inclusion criteria were those articles will be selected only if they have tried to project caseloads, pre-identifying disease hot-spots using the social media data alone. Only the original research and letter to the editor with any analysis were included in this study. Also, the open access articles were included in this research study due to their high potential of getting cited and read by many.

If the article is a research protocol report, predictions done using google search trends, news reports, other language research articles other than the English language were excluded from this study and further analysis. The search strategy has been given in Figure 1.

The search terms were standardized before searching for the publications relevant to this study's objective. The search terms are "Prediction of COVID 19 using Social Media Data," "Projection of COVID 19 using Social Media Data," "Identifying hot-spots of COVID using Social Media Data," and "Use of Social Media Data for predicting COVID." The search period window was from January 05 2021 to February 05, 2021. A team of 2 researchers started surveying in Scholar and Pubmed using the above-mentioned keywords. After looking in for the inclusion criteria curiously the articles were selected and steps were taken to identify the country of the corresponding author.

Descriptive statistics were used to analyze the country-wise frequency of articles published, type of articles got published, search engine wise frequency, and the number of articles got published was compared with the percentage of people using social media in various countries.

Figure 1:



PREDICTION OF COVID - 19 EPIDEMIOLOGY USING THE DATA FROM SOCIAL MEDIA - A DESCRIPTIVE BIBLIOGRAPHIC ANALYSIS.

Results:

Table 1: Bibliography Table for the selected Articles

S.no	Heading of the Research Article	Authored by	Name of the Journal	Year of Publication
1.	Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study.	Kaiyuan Sun and Jenny Chen	The Lancet Digital Health, Volume 2 and Issue 4.	April 2020.
2.	Retrospective analysis of the possibility of predicting the COVID 19 outbreak from internet and social media data, china, 2020.	Cuilian Li, Li Jia Chen, Xueyu Chen, Mingzhi Zhang, Chi Pui Pang, and Haoyu Chen.	Eurosurveillance, Volume 25 and Issue 10.	March 2020.
3.	Prediction of Number of Cases of 2019 Novel Coronavirus (COVID-19) Using Social Media Search Index	Lei Qin, Qiang Sun, Yidan Wang, Ke-Fei Wu, Mingchih Chen, Ben-Chang Shia and Szu-Yuan Wu.	International Journal of Environmental Research and Public Health, Volume 17 and Issue 7.	March 2020.
4.	Exploring Urban Spatial Features of COVID-19 Transmission in Wuhan Based on Social Media Data	Zhenghong Peng, Ru Wang, Lingbo Liu and Hao Wu.	International Journal of Geo-Information, Volume 9 and Issue 6.	June 2020.
5.	Social media WeChat infers the development trend of COVID-19	Yue Lu	Journal of Infection	2020
6.	Using Reports of Symptoms and Diagnoses on Social Media to Predict COVID-19 Case Counts in Mainland China: Observational Infoveillance Study	Shen C, Chen A, Luo C, Zhang J, Feng B, and Liao W	Journal of Medical Internet Research, Volume 22 and Issue 5.	May 2020
7.	Early warnings of COVID-19 outbreaks across Europe from social media.	Milena Lopreite, Pietro Panzarasa, Michelangelo Puliga & Massimo Riccaboni	Nature (Scientific reports) scientific reports 11 and Article number 2147.	January 2021
8.	JUE Insight: The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook.	Theresa Kuchler, Dominic Russel and Johannes Stroebel.	Journal of Urban Economics	January 2021
9.	Use of Twitter social media activity as a proxy for human mobility to predict the spatiotemporal spread of COVID-19 at global scale	Donal Bisanzio, Moritz U.G. Kraemer, Isaac I. Bogoch, Thomas Brewer, John S. Brownstein and Richard Reithinger.	Geospatial Health, Volume 15 and Issue 1.	June 2020
10	Data Mining and Content Analysis of the Chinese Social Media Platform Weibo During the Early COVID-19 Outbreak: Retrospective Observational Infoveillance Study	Jiawei Li, Qing Xu, Raphael Cuomo, Vidya Purushothaman, and Tim Mackey	JMIR Public Health Surveillance	April 2020
11	Identification of Risk Factors and Symptoms of COVID-19: Analysis of Biomedical Literature and Social Media Data	Jouhyun Jeon, Gaurav Baruah, Sarah Sarabadani and Adam Palanica.	Journal of Medical Internet Research, Volume 22 and Issue 10.	October 2020.
12	Geolocated Twitter social media data to describe the geographic spread of SARS-CoV-2	Donal Bisanzio DVM, Moritz U G Kraemer, Thomas Brewer, John S Brownstein and Richard Reithinger	Journal of Travel Medicine, Volume 27 and Issue 5.	July 2020

The above table reveals the research articles along with the authors with the journal name and year they got published.

PREDICTION OF COVID - 19 EPIDEMIOLOGY USING THE DATA FROM SOCIAL MEDIA - A DESCRIPTIVE BIBLIOGRAPHIC ANALYSIS.

Figure 2: Graph showing the number of articles got published along with respective countries.

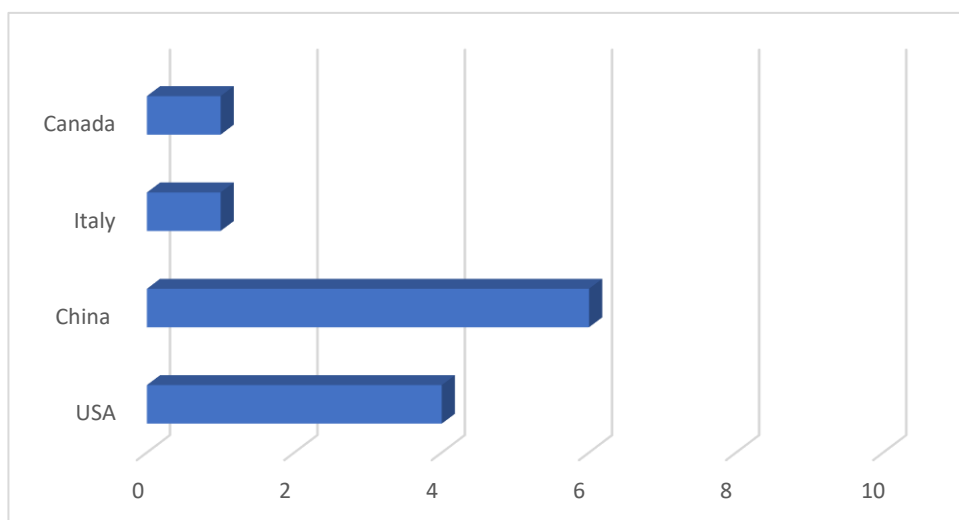
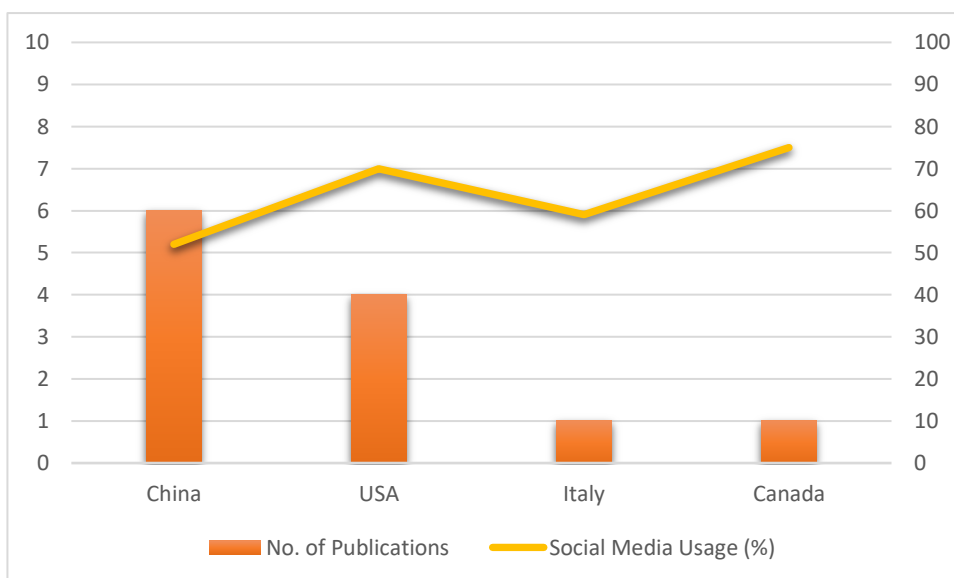


Table 2: Article Metrics table along with the Country of Origin

S.no	Article name	Citations	Fully read	text
China				
1.	Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study.	133	964	
2.	Retrospective analysis of the possibility of predicting the COVID 19 outbreak from internet and social media data, china, 2020.	37	9000	
3.	Prediction of Number of Cases of 2019 Novel Coronavirus (COVID-19) Using Social Media Search Index	54	5000	
4.	Exploring Urban Spatial Features of COVID-19 Transmission in Wuhan Based on Social Media Data	19	1600	
5.	Social media WeChat infers the development trend of COVID-19	13	NA	
6.	Using Reports of Symptoms and Diagnoses on Social Media to Predict COVID-19 Case Counts in Mainland China: Observational Infoveillance Study	25	NA	
United States of America				
7.	JUE Insight: The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook.	96		
8.	Use of Twitter social media activity as a proxy for human mobility to predict the spatiotemporal spread of COVID-19 at global scale	13	86	
9.	Data Mining and Content Analysis of the Chinese Social Media Platform Weibo During the Early COVID-19 Outbreak: Retrospective Observational Infoveillance Study	40	12000	
10.	Geolocated Twitter social media data to describe the geographic spread of SARS-CoV-2	2	5	
Italy				
11.	Early warnings of COVID-19 outbreaks across Europe from social media.	1	774	
Canada				
12.	Identification of Risk Factors and Symptoms of COVID-19: Analysis of Biomedical Literature and Social Media Data	1	44	

The above table depicts the countries and titles of the research articles it produced along with the citations score and those numbers who read the article fully. This data throws some light on the nature of visibility and access to the full article. Gaining full access to any research papers by the readers will improve its dissemination among the scientific community. The highest read article was from the USA and the highest cited article was from China.

Figure 4: Bar chart comparing the number of publications and social media usage according to the country.



The above table expresses the correlation between the number of articles published by the countries with its social media usage or Internet penetration rate. Despite the higher social media usage, Canada has come up with 1 article only.

Discussion:

Among the worst-hit countries by COVID - 19, China also faced hardships apart from the fact that it was the country of origin. China has published more research papers in esteemed international journals in other languages also apart from English.(9) The authors who worked in this field of predicting the cases with the help of social media data were found to be higher in China since they produce more research papers than any other country in the world. Any scientific article will gain potential if it gets cited by other researchers to support or contradict with their findings or other empirical shreds of evidence. The average citation score was 36.16, this average reveals that the manuscripts published with the ideology of predicting the hot spots and caseloads with the help of social media data were getting cited 36 times on an average after the date it got published. Regarding the usage of social media among the countries which published manuscripts and internet penetration rate is negatively correlated. Canada, one of the top countries which have a significantly higher internet penetration rate has produced just one paper on predicting caseloads using social media data.(10) Finally, China has published six research papers followed by four research papers by the USA. Then each from Italy and Canada on this prediction of COVID - 19 cases and identifying hot spots using the social media data.

How Social Media Data helped in identifying the COVID - 19 hotspots?

Traditionally there are many ways available for the Public Health authorities for implementing stringent monitoring mechanisms and name a few like active surveillance, lab-based screening activities, mass level diagnostic campaigns, etc. now the recent development in the surveillance of epidemics is conducting the community level surveillance through people’s social media usage.(11) This became a blessing in disguise for the health authorities and policymakers to identify the hot spots of COVID - 19 well in advance even before it happened. The commonest methodology followed for the prediction of COVID cases in a specified area was through monitoring the people’s messages, posts, tweets, etc with pre-identified keywords and mapping it with the use of GIS software. When there is more exchange of those keywords happening through social media then the chance for that area to get affected or possibly turn into a hot spot was found to be higher.(12) Along with this to get more refined results and for us to have a deeper understanding some researchers have gone one step further in the analysis by controlling some key variables like distance, the population density of that respective area taken for the study, and categorizing that community according to the income levels and results were obtained.

PREDICTION OF COVID - 19 EPIDEMIOLOGY USING THE DATA FROM SOCIAL MEDIA - A DESCRIPTIVE BIBLIOGRAPHIC ANALYSIS.

To do this type of analysis we require social mapping to know about the social proximity (rather than physical proximity) among the regions, states, or countries. This social proximity is being determined with the help of a measure called "Social Connectedness Index" and then correlate this index with friend links of the area under the study so that areas that are densely linked will have more chance of spreading the disease due to physical contact.(13)

Conclusion:

The COVID - 19 pandemic has raised some serious public health concerns and has put many countries' health systems to test. Due to the serious spread of the disease, an avenue for research and adding new knowledge to the scientific community has been created. Since many countries are now having a minimal digital divide usage of social media has picked up a pace in recent years. So, it's a great opportunity to know and predict the caseloads and identifying potential COVID - 19 hotspots through social media data for the scientific community. This will help in containing the disease spread and thereby reducing the infection and mortality rates due to COVID - 19. Its is recommended that following the output of this bibliography study a meta - analysis can be done and can report the findings which will throw more light on the findings which will be having a good scientific rigour.

References:

1. Velavan TP, Meyer CG. The COVID-19 epidemic. *Trop Med Int Health*. 2020 Mar;25(3):278-80.
2. Cao X. COVID-19: immunopathology and its implications for therapy. *Nat Rev Immunol*. 2020 May;20(5):269-70.
3. Dashraath P, Wong JLJ, Lim MXK, Lim LM, Li S, Biswas A, et al. Coronavirus disease 2019 (COVID-19) pandemic and pregnancy. *Am J Obstet Gynecol*. 2020 Jun 1;222(6):521-31.
4. Atalan A. Is the lockdown important to prevent the COVID-19 pandemic? Effects on psychology, environment and economy-perspective. *Ann Med Surg*. 2020 Aug 1;56:38-42.
5. Lemenager T, Neissner M, Koopmann A, Reinhard I, Georgiadou E, Müller A, et al. COVID-19 Lockdown Restrictions and Online Media Consumption in Germany. *Int J Environ Res Public Health*. 2021 Jan;18(1):14.
6. Cellini N, Canale N, Mioni G, Costa S. Changes in sleep pattern, sense of time and digital media use during COVID-19 lockdown in Italy. *J Sleep Res*. 2020;29(4):e13074.
7. Ali KF, Whitebridge S, Jamal MH, Alsafy M, Atkin SL. Perceptions, Knowledge, and Behaviors Related to COVID-19 Among Social Media Users: Cross-Sectional Study. *J Med Internet Res*. 2020;22(9):e19913.
8. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. *Eurosurveillance*. 2020 Mar 12;25(10):2000199.
9. Al-Zaman MdS. A bibliometric and co-occurrence analysis of COVID-19-related literature published between December 2019 to June 2020. *Sci Ed*. 2021 Feb 20;8:57-63.
10. Poushter J, Bishop C, Chwe H. Social Media Use Continues to Rise in Developing Countries but Plateaus Across Developed Ones. :46.
11. Lopreite M, Panzarasa P, Puliga M, Riccaboni M. Early warnings of COVID-19 outbreaks across Europe from social media. *Sci Rep*. 2021 Jan 25;11(1):2147.
12. Peng Z, Wang R, Liu L, Wu H. Exploring Urban Spatial Features of COVID-19 Transmission in Wuhan Based on Social Media Data. *ISPRS Int J Geo-Inf*. 2020 Jun;9(6):402.
13. Kuchler T, Russel D, Stroebel J. The Geographic Spread of COVID-19 Correlates with the Structure of Social Networks as Measured by Facebook [Internet]. Cambridge, MA: National Bureau of Economic Research; 2020 Apr [cited 2021 Apr 23] p. w26990. Report No.: w26990. Available from: <http://www.nber.org/papers/w26990.pdf>