ISSN 1870-4069

COVID-19 Pandemic: An Overview of Machine and Deep Learning Methods for Analysis of Digital Media Texts

Nouf Matar Alzahrani

Albaha University,
College of Computer Science and Information Technology,
Saudi Arabia

noufalzahrani@bu.edu.sa

Abstract. The global efforts to generate information during the COVID-19 outbreak are awe-inspiring. Governments are attempting to make sure people's health is safe and sound during this epidemic by automatically processing the giant amount of online text. This text helps governments to make appropriate policies promptly by understanding public opinion at a suitable time to avoid outrageous consequences. The social media platforms play a significant role in fostering healthy online public communities, particularly for the user to user interaction in such pandemic circumstances. However, manually processing and analyzing a huge amount of data is a troublesome task. So, the study aims to provide a comprehensive analysis of the methods used to automatically process and analyze the digital media text. Moreover, the study also sheds light on the traditional machine learning and deep learning algorithms used to monitor users' activity, attitude, and response to ample amounts of information on various social media platforms during the outbreak of COVID-19. The study concludes that each digital platform has been used for different goals, such as communication and thus user's response is different on each digital media platform.

Keywords: social media, machine learning, deep learning, natural language processing.

1 Introduction

The World Health Organization (WHO) officially announced the COVID-19 as the name of this new disease that poses a severe global threat[1]. A recent study [1] highlighted that worldwide hazards are unified, especially the case of the COVID-19 epidemic exhibits how the prevalence of information can be crucial in such pandemic situations.

The flow of a huge amount of information, in particular, fake news on the internet can significantly augment the epidemic process because it influences people and fragments the social response [2].

---

[1] WHO: Naming the coronavirus disease (COVID-19) and the virus that causes it.

For example, the American news channel is known as CNN[2] published an article showing the possibility of lock-down in Lombardy (a region in northern Italy). Since the news was published a few hours earlier than the official communication from the Italian Prime Minister, people started to travel from Lombardy to other regions. As a result, all the airports, train stations were overwhelmed before implementing the lockdown, and the government initiative was interrupted. It opens a room for many research questions, but the most important question is to investigate the sources where people receive information and make decisions. A recent study [3] highlighted that when people make a decision after consuming information, such decisions have an impact on their attitude. Finally, the COVID-19 epidemic clearly illustrates how social and traditional media influence people's decisions, in particular, and the complete society. Therefore, governments and other organizations are interested in designing automatic tools to understand public opinion in such catastrophic situations to provide people with the best health services and take proper steps to control the epidemic.

Easy access to the internet and communication technologies have provided people a significant opportunity to connect with each other and communicate regardless of the distance. The mechanism to monitor communication on social media platforms, especially when governments are imposing lock down restrictions due to the COVID-19 epidemic is demanding. The reason is that in such situations, information on social media platforms influences people's behavior and it can have a substantial impact on people's decisions. For example, information without proper evidence can easily disrupt government countermeasures especially when the situation is a severe threat to public health. Therefore, it requires automatic models to predict people's attitude towards epidemics such as COVID-19. These predictive models can be extremely fruitful to measure the people's responses after consuming traditional news steam and social media content [2, 4, 5].

People can have direct access to an ample amount of information on various platforms. These platforms include traditional news channels, multiple social media networks such as Facebook, Twitter, YouTube, Instagram. This direct access to information may contain rumors or unverified news. However, new techniques [6] have been proposed to target the specific audience based on the user's preference and response. Although the algorithms can mediate in finding specific audiences and to promote content, however, this may lead to elicit misleading information phenomena [6]. Thus, rumors or unvaried content can provoke public discussion on social media platforms [7, 8], and impact policy-making process, in particular, when the issues arouse controversy and help to develop social perceptions [9, 10].

Online social media platforms have an active wide range audience that provide the worldviews on various topics to its users [11]. These broader views often neglect the conflicting facts [12, 13] and therefore people establish groups that share the same narratives [14, 15]. As a result, when the degree of polarization is intense, it is more likely to disseminate false information [16, 17]. Previous research [18] showed that the factually incorrect news scatter with a higher velocity than the verified news. Nonetheless, the fact checking phenomena might be limited to a specific social media platform since the term "Fake News" may be misleading due to the face that the unpleasant news for a particular group can be annotated as a factually incorrect content

[2] CNN: Italy prohibits travel and cancels all public events in its northern region to contain coronavirus.

[19]. The study aims to investigate the machine learning and deep learning methods used to process the digital media in the COVID-19 epidemics. A series of recent studies [10, 18, 20] have indicated that each platform has a different mechanism and provides different services to its users. For example, some platforms are used to share pictures, and stories. In contrast, some online platforms are used to keep people up to date by publishing latest news. To keep this in mind, in this research we presented an overview of the methods that are being used to automatically analyze and process the online text in the pandemic of COVID-19.

The remaining sections of the paper are structured as follows: applications of AI in digital media mining are demonstrated in section 2. In section 3, the methods used to process online text are provided in detail. Section 4 contains the analysis of various social media platforms. Furthermore, discussion and questions for further research are discussed in section 5 and 6 respectively. Finally, the paper concludes with the conclusion in Section 7.

2 Applications of AI in Mining Digital Media Texts

Social media platforms have successfully become an important medium of communication. These platforms are extremely helpful for people to communicate with each other, share their emotions in the form of posts and know up to date worldwide information. However, people are generating a giant amount of data each second on these platforms. For example, the World Economic[3] report that by 2020 it is expected to generate 44 zettabytes. So, it implies the number of stars[4] in the whole plant are 40 times less than bytes generated on the internet in 2020. Nonetheless, it requires automatic tools to process such amounts of data.

Covid-19[5] pandemic has restricted people at home and has physical distance to prevent a catastrophic situation. People have been using Social media platforms as a main communication tool and generate a substantial amount of data during the coronavirus epidemic. For example, a recent study [20] shows that, in every minute, there are approximately 3,3 million Facebook posts. In addition to this, the study [20] also reveals that, approximately 448,800 Tweeting posts are published on Twitter every minute.

It is evident to note that the flow of information can be seen on social media platforms within a few seconds after a crisis occurs. Nonetheless, manually processing to get useful insights from this huge amount of data is quite crucial due to a number of reasons. For example, each platform has a different flow and structure of information, such as Facebook Posts, Twitter Tweets, YouTube video, Instagram pictures. Moreover, the text used on these platforms is unstructured, unorganized, and heterogeneous (different languages text, acronyms, and misspellings). The multiple heterogeneous text features introduce a noise in the data that makes text different to process and understand to obtain fruitful insights and the conveyed message.

[3] https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/.

[4] https://www.visualcapitalist.com/how-much-data-is-generated-each-day/.

[5] https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

Table 1. Comparison of the data gendered in 2017 and 2020 in every minute [22].

Year	2017 (in millions)	2020 (in millions)
Facebook new users	360	400
Google searchers	3.8	4.2
Tweets constructed	448,800	480,000
Instagram images uploaded	66,000	60,000
YouTube videos viewed	4.2	4.7
Emails sent	150	200

Therefore, various applications of machine learning algorithms (supervised and unsupervised machine learning) came into being. These algorithms are significant in not only understanding the text and provide meaningful results, but also automate the entire processing.

3 Methods for Automatic Data Processing

In this section, we provide an overview of different methods used to automatically analyze the online text. These methods [50, 60] have been used to understand public sentiment, response and attitude towards COVID-19 [58, 59] epidemic.

3.1 Supervised Learning

In supervised machine learning, annotation (a process to label data) is required which is mostly done by humans. For example, the text on social media platforms about COVID-19 may contain some characteristics in it. So, the label of the text post can be real information or fake information in a binary class scenario. Furthermore, the supervised learning algorithms process the text in the form of features. The feature is defined as an important piece of information that can help algorithms to correctly classify the information according to its labeled class. There are multiple kinds of features, in particular, each task can be solved with unique features since there is no standard set of features that can help the algorithm to provide better results. In Natural Language processing tasks, these features can be factual based, emotional, sentiment based, word embeddings, lexical features. Moreover, some other features such as time, location, the source of information are also taken into account as features. Finally, the main goal of this learning is to train an algorithm so that it can correctly predict the label of the unknown data.

A number of studies have investigated various machine learning algorithms [23, 24, 25]. These algorithms such as Naive Bayes [26, 27], Support Vector Machine (SVM) [25,28], Random Forests [29], and Logistic Regression [30, 31] and Decision trees [32] have been used in various text classification application [57] especially in COVID-19 related text processing. To process social media especially on Twitter, recent studies used certain tags to analyze the twitter posting [33], in particular, the tag "breaking news" and "news" has been used to differentiate breaking news twitter posting [34]. Similarly, another study classified tweets using the word "cardiovascular disease" as a key feature [35].

In supervised learning, the algorithms are trained on labelled dataset. Substantial human efforts are required to annotate the dataset correctly. However, since automatic tools can be used to label the data, the risk of incorrect annotation is present which might lead an algorithm to make a wrong prediction. Labeling the dataset in crisis such as COVID-19 is a troublesome task [36]. Such situations require instant information to make new policy decisions for the government organizations. Furthermore, making use of a classifier trained on different data (not related to COVID-19) will not predict and provide required insights of the information related to COVID-19 since it did not learn the pattern of the COVID-19 related data. Therefore, this requires addressing the annotated data limitation by suggesting other methods. Finally, unsupervised learning techniques are introduced to meet the data labeling related challenges.

3.2 Unsupervised Learning

Another machine learning method known as unsupervised methods that require unlabeled data for training [37]. These algorithms are powerful to provide valuable information about the data using clustering techniques. Clustering techniques are significant when seekers are ignorant about what information can be obtained from data. This is the particular case when clustering techniques have been used in the COVID-19 text processing to understand the situation globally [38]. It significantly not only reduces the human efforts to label the data, but also provides data insights (by grouping similar data together) promptly that can be helpful to control any outrageous situation such as COVID-19. Similarly, a recent study used spatio-temporal distribution to spots potentially COVID-19 affected areas using unsupervised learning [39, 54].

Another clustering technique known as soft clusters have been used [40] for predicting COVID-19 patients. This method permits items (in this case patients) to join other clusters with variant degrees simultaneously. In addition to this, the approach has been applied [42] in predicting COVID-19 outbreak by finding the tweets similarity using different features like word selection, length of the tweets, time and location.

3.3 Deep Learning

Artificial intelligence has substantially advanced the techniques to perform a number of tasks automatically [53]. The deep learning (DL) methods [51, 53, 55] are based on human brain structure. The emergence of DL has provided a nudge to the whole research community in solving challenging tasks that require huge computation power and resources.

The deep learning methods are based on various hidden layers (deep layers) and each layer contains a number of neurons. Each neuron in the deep layer is activated and passes the information to the next neuron. This process continues until the final is provided. These deep layers are extremely helpful in identifying the hidden complex structures in large datasets.

Another important aspect of deep learning method is extremely good in finding the suitable features automatically. These methods are given raw data as an input according to the required data type. For example, in Natural language processing, word embeddings of the data are given as an input to the deep learning models instead of raw data. In contrast, traditional machine learning techniques, we need to perform careful feature analysis on raw data to investigate what features contain more information and are valuable for the algorithm to correctly make the prediction.

There are several deep learning methods such as Convolutional neural networks, recurrent neural networks, which have been used in predicting medical symptoms [52, 56], particularly, the COVID-19 outbreak [42,43]. Deep convolutional nets performed extremely well in processing images related data, however, LSTM which is another type of recurrent nets have been used in classifying the sentiments related to COVID-19 [44] in online discussions. The method [44] used sequential data analysis on text to identify public sentiment on coronavirus diseases. Another study [45] used various neural networks to highlight COVID-19 patients who had pneumonia on their chest. A recent study [46] is exclusively focused on analyzing digital media text on social media platforms, checking the facts of the information.

Deep learning methods require a substantial amount of data for training. Covid-19 generated a substantial amount of data related to different diseases, death and the public opinion. For example, a study [47] used BERT technique to process Twitter posting using unlabeled data. Furthermore, similar studies [48, 49] used deep methods in the authorship attribution task. Finally, it is important to mention that deep learning methods have a variety of other applications to recognize hidden structure using images, audios and video related data and provide extremely good results to automatically perform multiple tasks.

4 Analysis of Digital Media Texts

Social media platforms have become an important part of our society because these platforms provide technologies to communicate with each other, engage with worldwide information [2, 5]. The current study investigates the importance of social media analysis in catastrophic situations such as COVID-19 pandemic. This is extremely difficult to understand a giant amount of text automatically that could lead to better understanding the critical situation and helpful for the policy makers to take appropriate measures and design policies in such situations.

4.1 Analysis of Social Media Texts during Covid-19

Social Media platforms are different in nature and have different characteristics. It means that each platform has a different data organization, data handling and postings flow. This is why it is extremely important to understand how the data is published in these platforms. Most of the platforms, such as Facebook, Twitter accept a variety of data types like pictures, text, videos. The pictures, video types data is beyond the scope of this study. The automatic algorithms divide sentences into words to understand the text published on social media platforms. This text can be twitter postings, facebook's stories, comments etc.

In addition to this, such types of data pose new issues such as incorrect word spellings, confusing abbreviations that may lead to a completely different meaning. It is extremely important to mention that the performance of supervised learning algorithms is highly data quality dependent. The better the data quality is, the good results can be obtained through supervised learning approaches. Therefore, the correctness of a giant amount of data is a troublesome task and this is why traditional supervised methods are inefficient to provide better results in real time prediction tasks. Similarly, unsupervised approaches [39, 40] use clustering techniques to find word similarities of correct word and its different misspellings. However, in the case of predicting COVID-19, due to its dependency on rigid metrics for similarity that influence the clusters, these methods are also not a good option to deal with catastrophic situations.

Deep learning methods, particularly, predicting the COVID-19 outbreak [42, 43] provides better results by deducting high-level abstractions (e.g. meaning) from lower levels (e.g. letters or substring), and these methods are good in representing words similarities and its various spellings variations. In addition to this, these methods are really good in the case of COVID-19 when the training set is not substantial [48, 49].

5 Discussion

It is evident that recent machine learning and deep learning methods have proven great potential in combating the COVID-19 pandemic. These techniques have been extremely helpful for governments and organizations to process big data related to COVID-19 and design appropriate policies like lockdown, social distancing, finding challenges people face on a daily basis to improve the medical health services. However, a number of challenges are still unaddressed and require the attention of responsible institutions. Although a substantial number of researches have been conducted in creating datasets related to Covid-19 [33, 61, 62, 63, 64, 65], a critical challenge arises from the lack of standard datasets.

As discussed in previous sections, various algorithms have been used to process the data, in particular, for the virus detection. For example, recent studies [63, 64] suggested different algorithms for virus detection. These algorithms are tested on different datasets, the study [63] achieved 82.9% accuracy while the study [64] demonstrated that their algorithms outperformed by getting 98.27% accuracy in detecting virus. Nevertheless, it is crucial to come to a conclusion to decide which algorithm performed other algorithms to detect COVID-19 virus since both datasets are different in size.

Standard datasets can be extremely fruitful to combat COVID-19 prompt and effectively. Substantial researchers used online data, concatenated with some other online available datasets to check their proposed algorithm performance. However, the standard dataset collection issue can be easily addressed if multiple organizations, government, tech firms, and various main health organizations (e.g., WHO and CDC) collectively assemble big datasets. This collaboration will foster not only identifying the solutions prompt, but also help in providing unique sources for the high-quality research work. To keep this in mind, the study [65] proposed a CORD-19 dataset that is created by multiple organizations.

These organizations are Georgetown's Center for Security, Microsoft Research, National Institutes of Health, and many other companies like Allen Institute for AI, and Chan Zuckerberg Initiative.

6 Questions for Further Research

It is prominent by a number of studies that AI and big data technologies play a significant role in combating the COVID19 pandemic. However, after reviewing the state-of-the-art literature, several research questions arose that motivated us to work on future research. For example, in the future research, we will investigate not only how AI analytics and big data can help us in studying the COVID-19 virus protein structure, but also understand how the performance in terms of accuracy and reliability of the data analytics tools can be improved to develop COVID-19 vaccine. This research will be fruitful in terms of economic and scientific perspectives.

7 Conclusion

The study aims to provide an overview of the state-of-the-art methods used to process and analyze the digital media text in the battle against the COVID-19 pandemic. In the beginning, the study provided an introduction about the outbreak of the COVID-19 disease. Subsequently, the applications of AI in mining digital media text and effective methods used to combat the COVID-19 disease. The study also reviewed and highlighted the limitations of the traditional machine learning models that are unable to provide good performance to track and predict people's response and attitude towards COVID-19 epidemic. On the other hand, a comprehensive analysis of multiple deep learning methods is provided that can be helpful to mitigate the issue and address the catastrophic situation effectively. In addition to this, the study also shedded light on the challenges that require special attention to successfully fight against this havoc. Finally, the study highlighted questions for future research and concluded that the deep learning models such as BERT outperforms other methods in mining digital text related to COVID-19 pandemic.

References

1. Quattrociocchi, W.: Part 2-social and political challenges: 2.1 western democracy in crisis? In: Global Risk Report World Economic Forum (2017)
2. Kim, L., Fast, S.M., Markuzon, N.: Incorporating media data into a model of infectious disease transmission. PloS One, 14(2) (2019)
3. Sharot, T., Sunstein, C.R.: How people decide what they want to know. Nature Human Behaviour, pp. 1–6 (2020)
4. Shaman, J., Karspeck, A., Yang, W., Tamerius, J., Lipsitch, M.: Real-time influenza forecasts during the 2012–2013 season. Nature Communications, 4(1), pp. 1–10 (2013)
5. Viboud, C., Vespignani, A.: The future of influenza forecasts. Proceedings of the National Academy of Sciences, 116(8), pp. 2802–2804 (2019)

6. Kulshrestha, J., Eslami, M., Messias, J., Saptarshi, B.Z., Ghosh., Krishna, P. Gummadi., & Karahalios, K.: Quantifying search bias: Investigating sources of bias for political searches in social media. In: Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing, pages 417–432 (2017)
7. Starnini, M., Frasca, M., Baronchelli, A.: Emergence of metapopulations and echo chambers in mobile agents. Scientific Reports, 6, pp. 31834 (2016)
8. Schmidt, A., Zollo, F., Scala, A., Betsch, C., Quattrociocchi, W.: Polarization of the vaccination debate on facebook. Vaccine, 36(25), pp. 3606–3612 (2018)
9. Schmidt, A.L., Zollo, F., Del Vicario, M., Bessi, A., Scala, A., Caldarelli, G., Stanley, H.G., Quattrociocchi, W.: Anatomy of news consumption on facebook. Proceedings of the National Academy of Sciences, 114(12), pp. 3035–3039 (2017)
10. Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H.E., Quattrociocchi, W.: The spreading of misinformation online. Proceedings of the National Academy of Sciences, 113(3), pp. 554–559 (2016)
11. Bessi, A., Coletto, M., Davidescu, G.A., Scala, A., Caldarelli, G., Quattrociocchi, W.: Science vs conspiracy: Collective narratives in the age of misinformation. PloS One, 10(2) (2015)
12. Zollo, F., Bessi, A., Del Vicario, M., Scala, A., Caldarelli, G., Shekhtman, L., Havlin, S., Quattrociocchi, W.: Debunking in a world of tribes. PloS One, 12(7) (2017)
13. Baronchelli, A.: The emergence of consensus: a primer. Royal Society Open Science, 5(2), pp. 172189 (2018)
14. Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., Quattrociocchi, W.: Echo chambers: Emotional contagion and group polarization on facebook. Scientific reports, 6, pp. 37825 (2016)
15. Bail, C.A., Argyle, L.P., Brown, T.W., Bumpus, J.P., Chen, H., Hunzaker, M.B.F., Lee, J., Mann, M., Merhout, F., Volfovsky, A.: Exposure to opposing views on social media can increase political polarization. Proceedings of the National Academy of Sciences, 115(37), pp. 9216–9221 (2019)
16. Del Vicario, M., Quattrociocchi, W., Scala, A., Zollo, F.: Polarization and fake news: Early warning of potential misinformation targets. ACM Transactions on the Web (TWEB), 13(2), pp. 1–22 (2019)
17. Wardle, C., Derakhshan, H.: Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe report, 27 (2017)
18. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science, 359(6380), pp. 1146–1151 (2018)
19. Ruths, D.: The misinformation machine. Science, 363(6425), pp. 348–348 (2019)
20. Bovet, A., Makse, H.A.: Influence of fake news in twitter during the us presidential election. Nature Communications, 10(1), pp. 1–14 (2019)
21. Nauhwar, A.: Digital Data: How digital data is going to become a gold for all industries in upcoming year. Iconic Research and Engineering Journals (2020)
22. NodeGraph: How much data is on the internet? The Big Data Facts Update 2020. https://www.nodegraph.se/how-much-data-is-on-the-internet/ (2020)
23. Kotsiantis, S.B., Zaharakis, I.D., Pintelas, P.E.: Machine learning: a review of classification and combining techniques. Artificial Intelligence Review, 26(3), 159–190 (2006)
24. Amjad, M., Sidorov, G., Zhila, A.: Data augmentation using machine translation for fake news detection in the Urdu language. In: Proceedings of the 12th Language Resources and Evaluation Conference, European Language Resources Association, pp. 2537–2542 (2020)
25. Amjad, M., Sidorov, G., Zhila, A., Gomez-Adorno, H., Voronkov, I., Gelbukh, A.: Bend the truth: 2541 a benchmark dataset for fake news detection in Urdu language and its evaluation. Journal of Intelligent & Fuzzy Systems, In Press (2020)
26. Rish, I.: An empirical study of the naive Bayes classifier. In: IJCAI 2001 Workshop On Empirical Methods In Artificial Intelligence, 3(22), pp. 41–46 (2001)

27. Amjad, M., Voronkov, I., Saenko, A., Gelbukh, A.: Comparison of text classification methods using deep learning neural networks. In: Proceedings of the 20th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing) (2019)
28. Jain, M., Narayan, S., Balaji, P., Bhowmick, A., Muthu, R.K.: Speech emotion recognition using support vector machine (2020)
29. Shi, F., Xia, L., Shan, F., Wu, D., Wei, Y., Yuan, H., Shen, D.: Large-scale screening of COVID-19 from community acquired pneumonia using infection size-aware classification (2020)
30. Chen, C., Yan, J.T., Zhou, N., Zhao, J.P., Wang, D.W.: Analysis of myocardial injury in patients with COVID-19 and association between concomitant cardiovascular diseases and severity of COVID-19. Zhonghua Xin Xue Guan Bing Za Zhi, 48 (2020)
31. Zhong, B.L., Luo, W., Li, H.M., Zhang, Q.Q., Liu, X.G., Li, W.T., Li, Y.: Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. International Journal of Biological Sciences, 16(10), pp. 1745 (2020)
32. Forrester, J.D., Nassar, A.K., Maggio, P.M., Hawn, M.T.: Precautions for operating room team members during the COVID-19 pandemic. Journal of the American College of Surgeons (2020)
33. Chen, E., Lerman, K., Ferrara, E.: Covid-19: The first public coronavirus twitter dataset (2020)
34. Park, H.W., Park, S., Chong, M.: Conversations and medical news frames on twitter: infodemiological study on COVID-19 in South Korea. Journal of Medical Internet Research, 22(5), e18897 (2020)
35. Clerkin, K.J., Fried, J.A., Raikhelkar, J., Sayer, G., Griffin, J.M., Masoumi, A., Schwartz, A.: Coronavirus disease 2019 (COVID-19) and cardiovascular disease. Circulation (2020)
36. Roda, W.C., Varughese, M.B., Han, D., Li, M.Y.: Why is it difficult to accurately predict the COVID-19 epidemic?. Infectious Disease Modelling (2020)
37. Chapelle, O., Scholkopf, B., Zien, A.: Semi-supervised learning. IEEE Transactions on Neural Networks, 20(3), pp. 542–542 (2009)
38. Pung, R., Chiew, C.J., Young, B.E., Chin, S., Chen, M.I., Clapham, H.E., Low, M.: Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures. The Lancet (2020)
39. Hisada, S., Murayama, T., Tsubouchi, K., Fujita, S., Yada, S., Wakamiya, S., Aramaki, E.: Syndromic surveillance using search query logs and user location information from smartphones against COVID-19 clusters in Japan (2020)
40. Hu, Z., Ge, Q., Jin, L., Xiong, M.: Artificial intelligence forecasting of COVID-19 in China (2020)
41. Jahanbin, K., Rahmanian, V.: Using twitter and web news mining to predict COVID-19 outbreak. Asian Pacific Journal of Tropical Medicine, 13 (2020)
42. Ferrara, E.: COVID-19 on twitter: bots, conspiracies, and social media activism (2020)
43. Alqudah, A.M., Qazan, S., Alqudah, A.: Automated systems for detection of COVID-19 using chest x-ray images and lightweight convolutional neural networks (2020)
44. Jelodar, H., Wang, Y., Orji, R., Huang, H.: Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach (2020)
45. Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., Cao, K.: Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. Radiology (2020)
46. Alam, F., Shaar, S., Nikolov, A., Mubarak, H., Martino, G.D.S., Abdelali, A., Nakov, P.: Fighting the COVID-19 Infodemic: modeling the perspective of journalists, fact-checkers, social media platforms, policy makers, and the society (2020)

47. Müller, M., Salathé, M., Kummervold, P.E.: COVID-Twitter-BERT: A natural language processing model to analyse COVID-19 content on twitter (2020)
48. Posadas-Durán, J.P., Gómez-Adorno, H., Sidorov, G., Batyrshin, I., Pinto, D.: Application of the distributed document representation in the authorship attribution task for small corpora. Soft Computing 21(3), pp. 627–639 (2017)
49. Gómez-Adorno, H., Posadas-Durán, J.P., Sidorov, G., Pinto, D.: Document embeddings learned on various types of n-grams for cross-topic authorship attribution. Computing 100 (7), pp. 741–756 (2018)
50. Bengio, Y., LeCun, Y.: Scaling learning algorithms towards AI. Large-Scale Kernel Machines, 34(5), pp. 1–41 (2007)
51. Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., Lew, M.S.: Deep learning for visual understanding: A review. Neurocomputing, 187, pp. 27–48 (2016)
52. Bejnordi, B.E., Veta, M., Van Diest, P.J., Van Ginneken, B., Karssemeijer, N., Litjens, G., Geessink, O.: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. Jama, 318(22), pp. 2199–2210 (2017)
53. Bengio, Y., Goodfellow, I., Courville, A.: Deep learning. 1, MIT press (2017)
54. Bengio, Y.: Deep learning of representations for unsupervised and transfer learning. In: Proceedings of ICML Workshop On Unsupervised and Transfer Learning, pp. 17–36 (2017)
55. Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M.P., Iyengar, S.S.: A survey on deep learning: Algorithms, techniques, and applications. ACM Computing Surveys (CSUR), 51(5), pp. 1–36 (2018)
56. Sun, W., Zheng, B., Qian, W.: Computer aided lung cancer diagnosis with deep learning algorithms. In: Medical imaging 2016: computer-aided diagnosis. 9785, pp. 97850Z, International Society for Optics and Photonics (2016)
57. Zhang, N., Zhang, R., Yao, H., Xu, H., Duan, M., Xie, T., Zhou, F.: Severity Detection For the Coronavirus Disease 2019 (COVID-19) Patients Using a Machine Learning Model Based on the Blood and Urine Tests. (2020)
58. Jiang, X., Coffee, M., Bari, A., Wang, J., Jiang, X., Shi, J., He, G.: Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. Computers, Materials & Continua, 63(1), pp. 537–551 (2020)
59. Albahri, A.S., Hamid, R.A.: Role of biological data mining and machine learning techniques in detecting and diagnosing the novel coronavirus (COVID-19): a systematic review. Journal of Medical Systems, 44(7) (2020)
60. Pham, Q.V., Nguyen, D.C., Hwang, W.J., Pathirana, P.N.: Artificial intelligence (AI) and big data for coronavirus (COVID-19) pandemic: A survey on the state-of-the-arts (2020)
61. Fong, S.J., Li, G., Dey, N., Crespo, R.G., Herrera-Viedma, E.: Finding an accurate early forecasting model from small dataset: A case of 2019-ncov novel coronavirus outbreak (2020)
62. Gao, Z., Yada, S., Wakamiya, S., Aramaki, E.: Naist covid: multilingual COVID-19 twitter and weibo dataset (2020)
63. Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., Xu, B.: A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19), MedRxiv (2020)
64. Ozkaya, U., Ozturk, S., Barstugan, M.: Coronavirus (COVID-19) classification using deep features fusion and ranking technique (2020)
65. Kaggle: COVID-19 open research dataset challenge (CORD-19): an AI challenge with AI2, CZI, MSR, Georgetown, NIH & The White House. www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge (2020)