Natural Language Processing for the COVID-19 Pandemic

Subjects: Health Care Sciences & Services | Computer Science, Artificial Intelligence
Contributor: Mohammed Ali Al-Garadi, Yuan-Chi Yang, Abeed Sarker

The COVID-19 pandemic is the most devastating public health crisis and has affected the lives of billions of people worldwide in unprecedented ways. Compared to pandemics of this scale in the past, societies are now equipped with advanced technologies that can mitigate the impacts of pandemics if utilized appropriately. However, opportunities are not fully utilized, particularly at the intersection of data science and health. Health-related big data and technological advances have the potential to significantly aid the fight against such pandemics, including the pandemic's ongoing and long-term impacts. Specifically, the field of natural language processing (NLP) has enormous potential at a time when vast amounts of text-based data are continuously generated from a multitude of sources, such as health/hospital systems, published medical literature, and social media. Effectively mitigating the impacts of the pandemic requires tackling challenges associated with the application and deployment of NLP systems.

Keywords: COVID-19 ; natural language processing ; health applications

1. Introduction

During a global health crisis such as the current COVID-19 pandemic, healthcare systems need practical solutions that can help provide effective care services and mitigate its impact on society. Outbreaks of novel diseases exert considerable pressure on public health and hospital systems [1][2]. Unlike past pandemics, however, the current one has occurred at a time when healthcare systems and public health agencies have access to large-scale data. Thus, the challenges posed by the crisis offer an opportunity to improve public health systems through the use of innovative technologies such as data-driven artificial intelligence (AI) [3]. One subset of AI technologies with enormous potential is natural language processing (NLP), particularly due to the large volumes of free-text data that are currently available and continuously generated through different channels, such as electronic health records (EHRs), published medical literature, and social media. The NLP of EHRs, for example, can help medical practitioners identify patterns in free-text clinical big data generated by COVID-19 patients, and/or discover the latent factors influencing their long-term outcomes [4]. The NLP of social media data may help address challenges associated with the COVID-19 *infodemic*, which refers to the massive spread of health disinformation and misinformation during the pandemic [5]. NLP applied to social media data related to COVID-19 may also help monitor people's mental health during the evolution of the pandemic, act as disease surveillance systems, and help to understand the psychological and sociological processes that can influence people to follow suggested health behaviors for the COVID-19 pandemic.

2. NLP for Electronic Health Records (EHRs)

The comprehensive adoption of EHRs in healthcare produces large real-world data that introduce new opportunities for critical clinical research. EHRs contain structured and unstructured data; the latter are typically referred to as clinical notes. As a significant volume of valuable clinical information is available in clinical notes, NLP techniques can be used for the real-time extraction of information from clinical free text. The utilization of EHRs for healthcare or scientific research requires data to be encoded and comparable [6]. In general, the role of NLP for this type of data is to convert unstructured data (i.e., free text data) into structured information that can be readily accessed and used. The key advantage of NLP applications for such data is that they enable the prompt utilization of extensive clinical data [2], allowing the use of EHRs for patients with novel diseases as soon as they are included in the system [7]. Although NLP application has been frequently recommended [8], such claims have not been tested in real time [7]. Thus, the present COVID-19 pandemic, with all of its challenges, can provide an opportunity to develop and implement real-time NLP models for EHRs with significant practical applications. The usefulness and applicability of NLP to clinical text [7] in response to emergencies have been evaluated with the main question of whether applying NLP models to unstructured textual information can yield clinically actionable knowledge. The outcomes indicate that NLP models can be developed rapidly to serve a novel

disease domain and extract valuable information [7]. When combined with structured data, the extracted knowledge is often able to increase the sample size satisfactorily to observe treatment effects that may not have been previously statistically detectable.

NLP models may serve as the main components of clinical AI systems that extract self-reported symptoms from individuals' audio or video recordings of clinic visits. A recording generally presents more informative facts about patient-reported symptoms compared to other sources. Recordings of clinic visits prepared at scale and combined with data from EHRs can enhance NLP models, thereby quickly creating patient-level clinical phenotypes of COVID-19 [9]. If clinical consultations are recorded and NLP models are effectively developed, benchtop virological findings can be better informed [9]. The potential role of NLP models to detect stroke during the COVID-19 pandemic from radiology reports has also been investigated [10]. The results demonstrated the potential of NLP approaches to automatically track acute or sub-acute ischemic stroke numbers for epidemiological studies. NLP models have also been developed to extract risk factors related to severe or non-severe COVID-19 from unstructured free text [11], and they showed promising results and the potential for real-time clinical applications.

NLP approaches have also been shown to be useful for extracting signs or symptoms of COVID-19 from clinical free text [12]. Owing to the importance of such NLP tasks, datasets such as the COVID-19 Annotated Clinical Text (CACT) have been created [13]. CACT is a dataset with annotations for COVID-19 diagnoses, testing, and symptoms that are used for training NLP models to detect annotated COVID-19 entities. Such datasets and others have enabled the development of machine learning (ML)-oriented NLP models. For instance, using a combination of NLP and ML methods enables the prediction of potential ICU admissions from the EHRs of patients with COVID-19 [14]. Another study used hospital discharge summary notes to develop an NLP pipeline to categorize the discharge dispositions of such patients [15]. Within the Department of Veterans Affairs (VA), a study developed an NLP system to extract possible positive COVID-19 cases from clinical text [16]. Detecting positive cases from clinical notes can help reduce the number of patients that laboratory-based surveillance methods may miss, and therefore, are not counted in the overall number of cases. Since EHRs in the VA contain data from hospitals across the United States, such a model can be useful for surveillance at the national level.

From the aforementioned papers, it is evident that with recent advances the application of NLP techniques in clinical notes can reveal new insights into real-time self-reported symptoms extraction, predicting potential ICU admissions, and improving pandemic prediction. The valuable information from these real-world data can aid research, healthcare systems, and regulatory activities. However, the characteristics of clinical notes pose many challenges for the application of NLP techniques, such as varying data quality, the difficulty of accurately de-identifying notes to protect patients' privacy, and difficulties associated with interoperability.

3. NLP for Mental Health

During the COVID-19 pandemic, most governments around the globe implemented strict domestic quarantine policies to control the spread of the disease. Infringement on personal freedom, financial hardship, misinformation, and uncertainties about the new virus are among the significant stressors that have been reported to increase emotional distress and risks of psychiatric illnesses associated with COVID-19 [17]. The pandemic is associated with elevated levels of psychological distress which, in many cases, meet the threshold for clinical relevance. Thus, relieving the severe effects of COVID-19 on mental health has become a worldwide public health priority [18].

NLP models can promptly monitor public sentiments and emotions on a large scale [19][20]. The use of NLP techniques to understand the mental states of individuals through the analysis of their posts on social media platforms is increasing. This analysis of public commentaries, such as on Twitter, Reddit, and Facebook, can capture the users' concerns, emotions, and mental states in real-time. A recent study applied NLP techniques to COVID-19-related data on Reddit to understand individuals' mental health. The authors showed that NLP techniques have been helpful to reveal mental health complaints in real time, recognize vulnerable individuals, and detect rapidly rising mental health-related topics during COVID-19 [21]. The study shows that NLP techniques performed robustly in finding mental health complaints in real time, as well as identifying vulnerable groups and important mental health-related topics during the pandemic. As discovered by NLP techniques, several linguistic patterns of mental health status can serve as helpful indicators and clues for further investigation in clinical settings [21].

Another study that aims to provide a research resource for developing NLP models created the Emotion-Covid19-Tweet (EmoCT) dataset containing 1000 annotated English tweets used for NLP model training. In the dataset, English tweets are labeled as expressing *anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise*, and *trust* [22]. In a separate study, over 20 million COVID-19 tweets between January 28 and April 9, 2020 were used to examine the shift of public emotions

during the early phases of the disease outbreak [23]. Fears about the unavailability of COVID-19 tests and medical supplies gradually turned into common discussion topics. *Sadness* was expressed in discussions about losing friends and family members, whereas topics related to *joy* were found to contain words of appreciation for good health [23]. In a similar direction, another study applied NLP techniques to explore 47 million COVID-19-related comments extracted from Twitter, Facebook, and YouTube. The findings showed that a total of 34 negative topics appeared, out of which 15 were related to COVID-19, specifically focusing on health, psychosocial, and social issues from the population health perspective. Furthermore, 20 positive topics were found, which were commonly related to public awareness, inspiration, gratitude, online learning, charity, spiritual support, innovative research, and a better environment [24].

NLP techniques can help to analyze real-time social media posts to understand temporal mental health dynamics associated with changes in COVID-19 regulations (such as national lockdowns). For instance, the correlation between temporal mental health dynamics and COVID-19 events was investigated in a study [25], and the results empirically demonstrated an association between the populations' temporal mental health dynamics and national lockdowns. Such findings can be referenced as a second opinion during strategic decision making.

NLP approaches have also been applied to free-text notes from sources other than social media to assess mental health status. For example, research has analyzed the free text generated by college students through an application designed to help improve their mental health [26]. The study intended to understand the sentiments that students reveal on specific topics between pre- and post-COVID-19 periods. The findings disclosed that topics such as *Education* became remarkably less essential to students after the pandemic, whereas topics on *Health* became more imprinting and trending. Moreover, the students expressed more negative sentiments across all topics in post-COVID-19 discussions than before the pandemic [26].

The real-time monitoring of mental health during a pandemic is vital for public health agencies that strive to improve public awareness and reduce the negative impact of the pandemic on individuals' mental health. From the literature, it is evident that NLP techniques can be used in near real-time mental health surveillance systems that can track, at a large scale, trends in people's mental health statuses associated with news, guidelines, misinformation, and public health responses during distinct phases of the pandemic. However, the validity of observational social media research on mental health status is still a challenge, as discussed in previous research [27][28][29]. The challenges can introduce gaps that may limit the deployment of NLP techniques on social media data to predict mental health status in clinical and public health systems [29].

4. NLP for COVID-19 Question-Answering Systems

Since the beginning of the COVID-19 outbreak, academics and researchers have focused on investigating COVID-19 and publishing relevant discoveries. The resulting large amount of published knowledge causes information overload [30], making it challenging for clinicians, medical professionals, and general readers to stay up to date with actionable insights. Real-time answers to important questions such as how the virus is transmitted, effective strategies for prevention, and risk factors for infection are essential and updated in almost real-time. Moreover, significant evidence needs to be summarized accordingly and conveyed to the public in a timely manner. Therefore, real-time question-answering (QA) systems based on the scientific literature can effectively disseminate information during an urgent time such as the COVID-19 pandemic.

To provide a large number of researchers and the public access to scientific findings on COVID-19, the World Health Organization (WHO), European Commission, and scientific research publishers have made relevant publications open access [5][31].

For COVID-19 QA and automatic text summarization (ATS), the common datasets that are available to researchers are as follows:

I. COVID-19 Open Research Dataset (CORD-19) [31]: A recent initiative established by the Allen Institute for AI, which contains all COVID-19-related publications. The CORD-19 dataset is updated daily to include the latest relevant published papers from various databases (such as arXiv, bioRxiv, and medRxiv, Medline, and PubMed Central) [31][32]. CORD-19 has more than 160,000 articles, of which more than 70,000 are full text [5]. The motive behind releasing this dataset is "to mobilize researchers to apply for recent advances in NLP to produce new insights in support of the fight against this infectious disease" [31].

II. COVID-QA dataset [33]: This dataset was created from scientific articles related to COVID-19 and annotated by volunteer biomedical experts. COVID-QA contains 2019 questions-and-answer pairs.

III.COVID-QA dataset by [34]: This dataset contains 124 question-and-article pairs annotated from the CORD-19 dataset.

Manual summarization is expensive and impractical. In practice, a manual summarization or search for an answer is impractical in the presence of massive amounts of textual data. ATS and QA systems hold a promising and practical solution to extract insights from such massive textual data. Researchers responding to the urgent call for building such solutions have developed ATS and QA systems. One of the first QA systems built using the CORD-19 corpus is CovidQA [34], for which the authors evaluated transformer models and unsupervised (zero-shot) approaches. The transformer models were proven effective for domain-specific supervised learning settings but had limited usefulness for out-of-domain contexts [34]. The analysis of several transformer models showed that T5 for ranking [35][36] accomplished the highest effectiveness in recognizing sentences from documents that contained answers.

Another research article [37] discussed the development of a real-time neural QA and query-focused multi-document summarization system called CAiRE-COVID. The system initially starts with the most relevant documents related to the input user query from the CORD-19 dataset and highlights the text spans containing the potential answer. The main NLP models used for building the CaiRE-COVID system architecture are as follows: a combination of two QA models, HLTC-MRQA [38] and BioBERT [39], are employed to construct the neural QA model; BART [40] for abstractive summarization; and ALBERT [41] in extractive summarization block. BERT is also used with topic modeling through latent Dirichlet allocation (LDA) to extract articles related to domains and retrieve answers to COVID-19 questions [42]. A real-time QA system that uses both biomedical text mining and QA methods to answer COVID-19-related questions was developed and called COVIDASK [43]. The primary NLP model in this architecture is BioBERT [39]. In other related research efforts, QA examples were synthetically generated to optimize the system performance on closed domains [44]. Neural information retrieval and machine reading comprehension methods were combined. The proposed approach showed significant increases in the performance of end-to-end QA on the CORD-19 collection compared with a state-of-the-art open-domain QA baseline.

Current QA systems, however, need further improvement to be used effectively during a pandemic. One of the primary challenges, mainly in the medical domain, is how to design QA systems that can respond with "*I do not know*" when a question is unanswerable or when an answer is uncertain. Moreover, while constructing QA systems, a follow-up question strategy to ask additional questions and information before providing the final answer, mainly when dealing with the complex question about COVID-19, is needed to avoid the ambiguity that may result in an inaccurate response [45]. QA systems should also include knowledge (e.g., common sense) beyond context-specific text and questions to which more accurate answers can be provided.

5. NLP for Knowledge Transfer

In response to the COVID-19 pandemic, universities and research centers conducted studies to understand the nature of the new virus, its transmission, risk factors, preventive steps, and measures to increase community awareness and prepare official guidelines. However, most of the published scientific reports and articles are in English, and translation of the scientific findings into several other languages is necessary to reach a larger population worldwide. NLP can play an important role to translate these findings and guidelines. For instance, NLP models were trained to offer multilingual translation support for general and biomedical domains [46]. A separate study constructed a multilingual dataset and then developed a model for cross-lingual intent detection to improve COVID-19 chatbots across the English, Spanish, French, and Spanglish languages [47]. Multilingual models have also been developed to understand people's sentiments about COVID-19 across various languages and countries [48].

6. Opportunities and Challenges for NLP Applications during the COVID-19 Pandemic

6.1. The Nature of a Pandemic

Pandemics are large-scale infectious disease outbreaks that can cause a critical upsurge in infection spread and mortality over a wide-ranging geographical region, leading to significant economic, social, and political disruptions [49][50]. The probability of pandemic occurrence has increased over the past century because of globalization, urbanization, changes in land use, and extensive exploitation of the natural environment [49][50]. Thus, improving the capability to respond to pandemics remains a challenge. COVID-19 is transmitted quite easily, with the average infected person spreading the disease to two or three others [51], and some recently emerging variants such as Delta and Omicron are even more infectious [52]. The rapid spread of COVID-19 necessitates the need for fast responses. However, developing NLP models that can efficiently support healthcare response systems still faces many obstacles. Most current successful NLP models

are trained on manually annotated data, which is time-consuming to create. Moreover, many annotated datasets, particularly those involving EHRs, are not publicly shared and are confined within the specific institution that is conducting the research. The lack of mechanisms for widespread data sharing presents challenges related to the generalization of implemented systems. Many systems that are developed remain effective only within the creating organization and typically underperform when applied to other healthcare settings. Creating frameworks that can enhance the data-annotation processes and enable widespread knowledge sharing can address such challenges and help develop NLP models that promptly meet the needs of people during the pandemic.

6.2. Characteristics of Health Misinformation

One potential application of NLP models is combatting the spread of health misinformation during the pandemic. However, misinformation is written in a manner that presents difficulties for the public to distinguish it from correct information [53]. Moreover, misinformation occurs as a distributed incident and usually spreads faster than the correct information [54] with dynamic modification to avoid automated detection [53][54].

The above challenges can be mitigated by designing NLP models that can speedily detect changes in public priorities, therefore, providing the necessary accurate information in a timely manner. Patterns and knowledge derived from social media can be used to guide targeted interventions [55]. Timely identification of the information discussed in subsets of populations can lead to more specific data campaigns and earlier public awareness of spreading misinformation [55].

6.3. Designing Clinically Applicable NLP Models

NLP models can be designed to extract actionable information by combining AI and clinical research [56]. On the one hand, the design of such systems must be clinically useful, and on the other hand, they must be implementable by NLP researchers who are typically not medical domain experts. An advantage of using NLP in healthcare is automation; clinicians cannot process data as rapidly as machines. Nevertheless, automated systems are trained and evaluated on selected databases that only contain information that may be specific to a targeted cohort or geolocation. If the databases do not represent the complete set of potential circumstances, then the automated systems can make incorrect decisions in cases that have never been examined [57].

The risk of inaccurate models is remarkably higher than that of a single doctor–patient interaction, yet the advantages of reducing cost, human errors, and inefficiencies in current healthcare systems are substantial [58]. One potential mechanism by which risks of AI or NLP-related errors can be mitigated is through the development of interpretable models. In this case, interpretability needs to focus on the medical practitioners who should be able to view the reasoning behind system decisions and decide if the system’s recommendations/decisions should be used. For critical clinical decisions, NLP researchers need to construct accurate but interpretable models that can identify the patterns that clinicians find interpretable, yet they should also be robust to make accurate decisions [57].

6.4. Synergic Implementation and Deployment

NLP systems can be most beneficial when incorporated into healthcare and public health systems. Digital health data (EHRs, scientific research findings, health information in social media) can be combined and processed in the NLP systems that benefit from each data source to provide recommendations on the individual and population levels. In the future, healthcare systems that can link clinical notes across different institutions must be developed to provide clinicians with tools to automate tasks and extract useful information. The NLP of scientific research can provide clinicians with timely and accurate updates, and social media can be used for outreach, crowdsourcing information, surveillance, and fighting misinformation. Ideally, such an NLP system can work on various data sources but still serve the ultimate goal of decreasing the consequences of outbreaks in society.

6.5. Sampling Bias on Social Media

Social media is a crucial data source to understand the impact of COVID-19 on subsets of populations. However, conducting social media-based studies, such as on mental health, can introduce sampling bias. Social media users are more likely to be younger and technologically savvy, resulting in biased samples. However, the wealth and diversity of accessible content make social media attractive as a data source [59]. Additionally, according to PEW research [60], the adoption of social media is growing among older populations, which means that in the future, it will be better representative of populations.

6.6. Data Analysis Challenge

NLP methods for studying health behavior, conducting pandemic surveillance, and monitoring mental health status at large scales can provide more comprehensive findings and insights than traditional approaches. The main objective is to translate the textual content into insightful statistical numbers (e.g., numbers of positive/negative posts, the intensity of positivity/negativity or emotion in a post, or a number of self-reported COVID-19 cases). However, researchers tend to aggregate statistical numbers to make them more manageable and perform overall descriptive analysis. How this aggregation of numerical findings is accomplished can compromise the final findings and may provide incorrect interpretations [27]. For example, when aggregating the number of positive or negative sentiments to study the sentiment changes during the COVID-19 phase, the number of positive or negative posts may give weight to active users' sentiments in the final inference, which in turn may lead to a biased conclusion toward these sentiments, rather than a conclusion derived from the overall population.

References

1. Legido-Quigley, H.; Asgari, N.; Teo, Y.Y.; Leung, G.M.; Oshitani, H.; Fukuda, K.; Cook, A.R.; Hsu, L.Y.; Shibuya, K.; Heymann, D. Are high-performing health systems resilient against the COVID-19 epidemic? Lancet 2020, 395, 848–850.
2. El Bcheraoui, C.; Weishaar, H.; Pozo-Martin, F.; Hanefeld, J. Assessing COVID-19 through the lens of health systems' preparedness: Time for a change. Glob. Health 2020, 16, 112.
3. Budd, J.; Miller, B.S.; Manning, E.M.; Lampos, V.; Zhuang, M.; Edelstein, M.; Rees, G.; Emery, V.C.; Stevens, M.M.; Keegan, N. Digital technologies in the public-health response to COVID-19. Nat. Med. 2020, 26, 1183–1192.
4. Venkatakrishnan, A.; Pawlowski, C.; Zemmour, D.; Hughes, T.; Anand, A.; Berner, G.; Kayal, N.; Puranik, A.; Conrad, I.; Bade, S. Mapping each pre-existing condition's association to short-term and long-term COVID-19 complications. Npj Digit. Med. 2021, 4, 117.
5. Zarocostas, J. How to fight an infodemic. Lancet 2020, 395, 676.
6. Ohno-Machado, L. Realizing the full potential of electronic health records: The role of natural language processing. J. Am. Med. Inform. Assoc. 2011, 18, 539.
7. Neuraz, A.; Lerner, I.; Digan, W.; Paris, N.; Tsopra, R.; Rogier, A.; Baudoin, D.; Cohen, K.B.; Burgun, A.; Garcelon, N. Natural language processing for rapid response to emergent diseases: Case study of calcium channel blockers and hypertension in the COVID-19 pandemic. J. Med. Internet Res. 2020, 22, e20773.
8. Elkin, P.L.; Froehling, D.A.; Wahner-Roedler, D.L.; Brown, S.H.; Bailey, K.R. Comparison of natural language processing biosurveillance methods for identifying influenza from encounter notes. Ann. Intern. Med. 2012, 156, 11–18.
9. Barr, P.J.; Ryan, J.; Jacobson, N.C. Precision Assessment of COVID-19 Phenotypes Using Large-Scale Clinic Visit Audio Recordings: Harnessing the Power of Patient Voice. J. Med. Internet Res. 2021, 23, e20545.
10. Li, M.; Lang, M.; Deng, F.; Chang, K.; Buch, K.; Rincon, S.; Mehan, W.; Leslie-Mazwi, T.; Kalpathy-Cramer, J. Analysis of stroke detection during the COVID-19 pandemic using natural language processing of radiology reports. Am. J. Neuroradiol. 2021, 42, 429–434.
11. Schoening, V.; Liakoni, E.; Drewe, J.; Hammann, F. Automatic identification of risk factors for SARS-CoV-2 positivity and severe clinical outcomes of COVID-19 using Data Mining and Natural Language Processing. medRxiv 2021.
12. Wang, J.; Abu-el-Rub, N.; Gray, J.; Pham, H.A.; Zhou, Y.; Manion, F.J.; Liu, M.; Song, X.; Xu, H.; Rouhizadeh, M. COVID-19 SignSym: A fast adaptation of a general clinical NLP tool to identify and normalize COVID-19 signs and symptoms to OMOP common data model. J. Am. Med. Inform. Assoc. 2021, 28, 1275–1283.
13. Lybarger, K.; Ostendorf, M.; Thompson, M.; Yetisgen, M. Extracting COVID-19 diagnoses and symptoms from clinical text: A new annotated corpus and neural event extraction framework. J. Biomed. Inform. 2021, 117, 103761.
14. Izquierdo, J.L.; Ancochea, J.; Soriano, J.B.; Group, S.C.-R. Clinical characteristics and prognostic factors for intensive care unit admission of patients With COVID-19: Retrospective study using machine learning and natural language processing. J. Med. Internet Res. 2020, 22, e21801.
15. Fernandes, M.; Sun, H.; Jain, A.; Alabsi, H.S.; Brenner, L.N.; Ye, E.; Ge, W.; Collens, S.I.; Leone, M.J.; Das, S. Classification of the Disposition of Patients Hospitalized with COVID-19: Reading Discharge Summaries Using Natural Language Processing. JMIR Med. Inform. 2021, 9, e25457.
16. Chapman, A.B.; Peterson, K.S.; Turano, A.; Box, T.L.; Wallace, K.S.; Jones, M. A Natural Language Processing System for National COVID-19 Surveillance in the US Department of Veterans Affairs. Openreview 2020, 7, 1–7.

17. Pfefferbaum, B.; North, C.S. Mental health and the COVID-19 pandemic. N. Engl. J. Med. 2020, 383, 510–512.
18. Xiong, J.; Lipsitz, O.; Nasri, F.; Lui, L.M.; Gill, H.; Phan, L.; Chen-Li, D.; Iacobucci, M.; Ho, R.; Majeed, A. Impact of COVID-19 pandemic on mental health in the general population: A systematic review. J. Affect. Disord. 2020, 277, 55–64.
19. Calvo, R.A.; Milne, D.N.; Hussain, M.S.; Christensen, H. Natural language processing in mental health applications using non-clinical texts. Nat. Lang. Eng. 2017, 23, 649–685.
20. Abd Rahman, R.; Omar, K.; Noah, S.A.M.; Danuri, M.S.N.M.; Al-Garadi, M.A. Application of machine learning methods in mental health detection: A systematic review. IEEE Access 2020, 8, 183952–183964.
21. Low, D.M.; Rumker, L.; Talkar, T.; Torous, J.; Cecchi, G.; Ghosh, S.S. Natural Language Processing Reveals Vulnerable Mental Health Support Groups and Heightened Health Anxiety on Reddit During COVID-19: Observational Study. J. Med. Internet Res. 2020, 22, e22635.
22. Li, I.; Li, Y.; Li, T.; Alvarez-Napagao, S.; Garcia-Gasulla, D.; Suzumura, T. What Are We Depressed About When We Talk About COVID-19: Mental Health Analysis on Tweets Using Natural Language Processing. In Artificial Intelligence XXXVII, Proceedings of the 40th SGAI International Conference on Artificial Intelligence, AI 2020, Cambridge, UK, 15–17 December 2020; Lecture Notes in Computer Science; Bramer, M., Ellis, R., Eds.; Springer: Cham, Switzerland; Volume 12498.
23. Lwin, M.O.; Lu, J.; Sheldenkar, A.; Schulz, P.J.; Shin, W.; Gupta, R.; Yang, Y. Global sentiments surrounding the COVID-19 pandemic on Twitter: Analysis of Twitter trends. JMIR Public Health Surveill. 2020, 6, e19447.
24. Oyebode, O.; Ndulue, C.; Adib, A.; Mulchandani, D.; Suruliraj, B.; Orji, F.A.; Chambers, C.; Meier, S.; Orji, R. Health, Psychosocial, and Social issues emanating from COVID-19 pandemic based on Social Media Comments using Text Mining and Thematic Analysis. JMIR Med. Inform. 2021, 9, e22734.
25. Islam, M.M.; Karray, F.; Alhajj, R.; Zeng, J. A review on deep learning techniques for the diagnosis of novel coronavirus (COVID-19). IEEE Access 2021, 9, 30551–30572.
26. Sharma, R.; Pagadala, S.D.; Bharti, P.; Chellappan, S.; Schmidt, T.; Goyal, R. Assessing COVID-19 Impacts on College Students via Automated Processing of Free-form Text. arXiv 2020, arXiv:2012.09369.
27. Olteanu, A.; Castillo, C.; Diaz, F.; Kıcıman, E. Social data: Biases, methodological pitfalls, and ethical boundaries. Front. Big Data 2019, 2, 13.
28. Howison, J.; Wiggins, A.; Crowston, K. Validity issues in the use of social network analysis with digital trace data. J. Assoc. Inf. Syst. 2011, 12, 2.
29. Chancellor, S.; De Choudhury, M. Methods in predictive techniques for mental health status on social media: A critical review. Npj Digit. Med. 2020, 3, 43.
30. Rathore, F.A.; Farooq, F. Information overload and infodemic in the COVID-19 pandemic. J. Pak. Med. Assoc. 2020, 70, 162–165.
31. Colavizza, G.; Costas, R.; Traag, V.A.; Van Eck, N.J.; Van Leeuwen, T.; Waltman, L. A scientometric overview of CORD-19. PLoS ONE 2021, 16, e0244839.
32. Wang, L.L.; Lo, K.; Chandrasekhar, Y.; Reas, R.; Yang, J.; Eide, D.; Funk, K.; Kinney, R.; Liu, Z.; Merrill, W. Cord-19: The COVID-19 open research dataset. arXiv 2020, arXiv:2004.10706v2.
33. Möller, T.; Reina, A.; Jayakumar, R.; Pietsch, M. COVID-QA: A Question Answering Dataset for COVID-19. In Proceedings of the ACL 2020 Workshop on Natural Language Processing for COVID-19 (NLP-COVID), Seattle, DC, USA, 9 July 2020.
34. Tang, R.; Nogueira, R.; Zhang, E.; Gupta, N.; Cam, P.; Cho, K.; Lin, J. Rapidly bootstrapping a question answering dataset for COVID-19. arXiv 2020, arXiv:2004.11339.
35. Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the limits of transfer learning with a unified text-to-text transformer. arXiv 2019, arXiv:1910.10683.
36. Nogueira, R.; Jiang, Z.; Lin, J. Document ranking with a pretrained sequence-to-sequence model. arXiv 2020, arXiv:2003.06713.
37. Ching, T.; Himmelstein, D.S.; Beaulieu-Jones, B.K.; Kalinin, A.A.; Do, B.T.; Way, G.P.; Ferrero, E.; Agapow, P.-M.; Zietz, M.; Hoffman, M.M. Opportunities and obstacles for deep learning in biology and medicine. J. R. Soc. Interface 2018, 15, 20170387.
38. Su, D.; Xu, Y.; Winata, G.I.; Xu, P.; Kim, H.; Liu, Z.; Fung, P. Generalizing Question Answering System with Pre-Trained Language Model Fine-Tuning. In Proceedings of the 2nd Workshop on Machine Reading for Question Answering, Hong Kong, China, 4 November 2019; pp. 203–211.

39. Lee, J.; Yoon, W.; Kim, S.; Kim, D.; Kim, S.; So, C.H.; Kang, J. BioBERT: A pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 2020, 36, 1234–1240.
40. Lewis, M.; Liu, Y.; Goyal, N.; Ghazvininejad, M.; Mohamed, A.; Levy, O.; Stoyanov, V.; Zettlemoyer, L. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv 2019, arXiv:1910.13461.
41. Lan, Z.; Chen, M.; Goodman, S.; Gimpel, K.; Sharma, P.; Soricut, R. Albert: A lite bert for self-supervised learning of language representations. arXiv 2019, arXiv:1909.11942.
42. Venkataram, H.S.; Mattmann, C.A.; Penberthy, S. TopiQAL: Topic-aware Question Answering using Scalable Domain-specific Supercomputers. Proceedings of 2020 IEEE/ACM Fourth Workshop on Deep Learning on Supercomputers (DLS), Atlanta, GA, USA, 11 November 2020; pp. 48–55.
43. Lee, J.; Yi, S.S.; Jeong, M.; Sung, M.; Yoon, W.; Choi, Y.; Ko, M.; Kang, J. Answering questions on COVID-19 in real-time. arXiv 2020, arXiv:2006.15830.
44. Reddy, R.G.; Iyer, B.; Sultan, M.A.; Zhang, R.; Sil, A.; Castelli, V.; Florian, R.; Roukos, S. End-to-End QA on COVID-19: Domain Adaptation with Synthetic Training. arXiv 2020, arXiv:2012.01414.
45. Zhu, F.; Lei, W.; Wang, C.; Zheng, J.; Poria, S.; Chua, T.-S. Retrieving and reading: A comprehensive survey on open-domain question answering. arXiv 2021, arXiv:2101.00774.
46. Bérard, A.; Kim, Z.M.; Nikoulina, V.; Park, E.L.; Gallé, M. A Multilingual Neural Machine Translation Model for Biomedical Data. arXiv 2020, arXiv:2008.02878.
47. Arora, A.; Shrivastava, A.; Mohit, M.; Lecanda, L.S.-M.; Aly, A. Cross-lingual Transfer Learning for Intent Detection of COVID-19 Utterances. Openreview 2020, 1–8.
48. Kruspe, A.; Häberle, M.; Kuhn, I.; Zhu, X.X. Cross-language sentiment analysis of European Twitter messages duringthe COVID-19 pandemic. arXiv 2020, arXiv:2008.12172.
49. Madhav, N.; Oppenheim, B.; Gallivan, M.; Mulembakani, P.; Rubin, E.; Wolfe, N. Pandemics: Risks, Impacts, and Mitigation. In Disease Control Priorities: Improving Health and Reducing Poverty, 3rd ed.; The International Bank for Reconstruction and Development/The World Bank: Washington, DC, USA, 2017.
50. Jones, K.E.; Patel, N.G.; Levy, M.A.; Storeygard, A.; Balk, D.; Gittleman, J.L.; Daszak, P. Global trends in emerging infectious diseases. Nature 2008, 451, 990–993.
51. Gates, B. Responding to COVID-19—A once-in-a-century pandemic? N. Engl. J. Med. 2020, 382, 1677–1679.
52. CDC. Delta Variant: What We Know About the Science. Cent. Dis. Control. Prev. 2021.
53. de Oliveira, N.R.; Pisa, P.S.; Lopez, M.A.; de Medeiros, D.S.V.; Mattos, D.M. Identifying Fake News on Social Networks Based on Natural Language Processing: Trends and Challenges. Information 2021, 12, 38.
54. Southwell, B.G.; Niederdeppe, J.; Cappella, J.N.; Gaysynsky, A.; Kelley, D.E.; Oh, A.; Peterson, E.B.; Chou, W.-Y.S. Misinformation as a misunderstood challenge to public health. Am. J. Prev. Med. 2019, 57, 282–285.
55. Stokes, D.C.; Andy, A.; Guntuku, S.C.; Ungar, L.H.; Merchant, R.M. Public priorities and concerns regarding COVID-19 in an online discussion forum: Longitudinal topic modeling. J. Gen. Intern. Med. 2020, 35, 2244–2247.
56. Wu, J.T.; Dernoncourt, F.; Gehrmann, S.; Tyler, P.D.; Moseley, E.T.; Carlson, E.T.; Grant, D.W.; Li, Y.; Welt, J.; Celi, L.A. Behind the scenes: A medical natural language processing project. Int. J. Med. Inform. 2018, 112, 68–73.
57. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Mach. Intell. 2019, 1, 206–215.
58. Topol, E.J. High-performance medicine: The convergence of human and artificial intelligence. Nat. Med. 2019, 25, 44–56.
59. Tabak, T.; Purver, M. Temporal Mental Health Dynamics on Social Media. arXiv 2020, arXiv:2008.13121.
60. Auxier, B.; Anderson, M. Social Media Use in 2021. Pew Research Center. 2021. Available online: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2021/04/PI_2021.04.07_Social-Media-Use_FINAL.pdf (accessed on 1 October 2022).