Leveraging Data Science to Combat COVID-19: A Comprehensive Review

Siddique Latif, Muhammad Usman, Sanaullah Manzoor, *Member, IEEE*, Waleed Iqbal, Junaid Qadir, *Senior Member, IEEE*, Gareth Tyson, Ignacio Castro, Adeel Razi, *Member, IEEE*, Maged N. Kamel Boulos, *Senior Member, IEEE*, Adrian Weller, and Jon Crowcroft, *Fellow, IEEE*

***Abstract*—COVID-19, an infectious disease caused by the SARS-CoV-2 virus, was declared a pandemic by the World Health Organisation (WHO) in March 2020. By mid-August 2020, more than 21 million people have tested positive worldwide. Infections have been growing rapidly and tremendous efforts are being made to fight the disease. In this paper, we attempt to systematise the various COVID-19 research activities leveraging data science, where we define data science broadly to encompass the various methods and tools—including those from artificial intelligence (AI), machine learning (ML), statistics, modeling, simulation, and data visualization—that can be used to store, process, and extract insights from data. In addition to reviewing the rapidly growing body of recent research, we survey public datasets and repositories that can be used for further work to track COVID-19 spread and mitigation strategies. As part of this, we present a bibliometric analysis of the papers produced in this short span of time. Finally, building on these insights, we highlight common challenges and pitfalls observed across the surveyed works. We also created a live resource repository at https://github.com/Data-Science-and-COVID-19/Leveraging-Data-Science-To-Combat-COVID-19-A-Comprehensive-Review that we intend to keep updated with the latest resources including new papers and datasets.**

***Impact Statement*—Data science, defined broadly, will play a central role in the global response to the COVID-19 pandemic. This paper facilitates the rapid engagement of data science and AI researchers with the breadth of the ongoing research work. In particular, we identify the major challenges involved, promising directions for further work, and important community resources. Given the interdisciplinary nature of the challenge, this review will help data scientists form collaborations across disciplines. We also elaborate the benefits of data science to strategists and policymakers and guide them in coming to grips with the challenges, opportunities, and pitfalls involved in using data science to combat the COVID-19 pandemic.**

***Index Terms*—Bibliometric analysis, COVID-19, data science, machine learning, medical image analysis, SARS-CoV-2, speech analysis, text mining.**

I. Introduction

The SARS-CoV-2 virus and the associated disease (designated as COVID-19) was first identified in Wuhan city (China) in December 2019 [1]–[3], and was declared a pandemic by the World Health Organisation (WHO) on 11 March 2020.[1] At the time of writing,[2] the Centre for Systems Science and Engineering at Johns Hopkins University reported 21,903,341 confirmed cases, 774,379 deaths, and 13,903,145 recovered.

Since December 2019, over 24,000 research papers from peer-reviewed journals and preprint sources are available [4]. Understanding this rapidly moving research landscape is particularly challenging since much of this literature has not been vetted through a peer-review process yet. This paper tries to overcome this challenge by presenting a detailed overview and survey of data science research related to COVID-19. It is intended as a community resource to facilitate accessibility to the large volume of data and papers published in recent months. We use the term "data science" as an umbrella term that encompasses all techniques and tools including

[1][Online]. Available: https://tinyurl.com/WHOPandemicAnnouncement

[2]10:30 am Saturday, 18 August 2020 Coordinated Universal Time (UTC).

Manuscript received April 27, 2020; revised July 7, 2020 and August 20, 2020; accepted August 26, 2020. Date of publication September 2, 2020; date of current version November 19, 2020. This paper was recommended for publication by Associate Editor Pablo Estevez upon evaluation of the reviewers' comments. The work of Adrian Weller was supported in part by David MacKay Newton research fellowship at Darwin College, The Alan Turing Institute under EPSRC Grants EP/N510129/1 and U/B/000074 and in part by Leverhulme Trust via CFI. The work of Adeel Razi was supported in part by the ARC (Refs: DE170100128 and DP200100757) and in part by NHMRC (Ref: APP1194910) and Wellcome Trust (Ref: 088130/Z/09/Z). The work of Ignacio Castro was supported by EPSRC under Grants EP/P025374/1 and EP/S033564/1. *(Corresponding author: Siddique Latif.)*

Siddique Latif is with the University of Southern Queensland, Springfield, Queensland 4300, Australia, and also with Distributed Sensing Systems Group, Data61, CSIRO, Pullenvale QLD 4069, Australia (e-mail: siddique.latif@usq.edu.au).

Muhammad Usman is with the Seoul National University, Seoul 08700, South Korea, and also with the Center for Artificial Intelligence in Medicine and Imaging, HealthHub Company Ltd., Seoul 06524, South Korea (e-mail: usman@healthhub.kr).

Sanaullah Manzoor and Junaid Qadir are with Information Technology University, Punjab 5400, Pakistan (e-mail: Sanaullah.manzoor@itu.edu.pk; junaid.qadir@itu.edu.pk).

Waleed Iqbal and Ignacio Castro are with the Queen Mary University of London, London E1 4NS, U.K. (e-mail: w.iqbal@qmul.ac.uk; i.castro@qmul.ac.uk).

Gareth Tyson is with the Queen Mary University of London, London E1 4NS, U.K., and also with Alan Turing Institute, London NW1 2DB, U.K. (e-mail: g.tyson@qmul.ac.uk).

Adeel Razi is with Turner Institute for Brain and Mental Health & Monash Biomedical Imaging, Monash University, Melbourne 3800, Australia (e-mail: adeel.razi@monash.edu).

Maged N. Kamel Boulos is with the School of Information Management, Sun Yat-sen University, Guangzhou 510006, China (e-mail: mnkboulos@ieee.org).

Adrian Weller is with the University of Cambridge, Cambridge CB2 1PZ, U.K., and also with the Alan Turing Institute, London NW1 2DB, U.K. (e-mail: aweller@turing.ac.uk).

Jon Crowcroft is with the University of Cambridge, Cambridge CB2 1TN, U.K., and also with the Alan Turing Institute, London NW1 2DB, U.K. (e-mail: jon.crowcroft@cl.cam.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAI.2020.3020521

TABLE I
ORGANISATION OF PAPER AND SUMMARY OF DIFFERENT SECTIONS

Sections	**Subsection**
(§**II**) **Datasets and Resources** This section provides information about numerous datasets related to COVID-19. It also gives information about ongoing data science competitions, and online resources.	(§II-A) COVID-19 Case Data
	(§II-B) COVID-19 Textual Data
	(§II-C) COVID-19 Biomedical Data
	(§II-F) Other Supportive Datasets
	(§II-E) COVID-19 Competition Datasets
(§**III**) **Applications of Data Science for COVID-19** This section highlights different use cases related to the application of data-driven methodologies for addressing COVID-19. It also discusses some examples of these use cases.	(§III-A) Risk Assessment and Patient Prioritisation
	(§III-B) Screening and Diagnosis
	(§III-C) Simulation and Modelling
	(§III-D) Contact Tracing
	(§III-E) Understanding Social Interventions
	(§III-F) Logistical Planning and Economic Interventions
	(§III-G) Automated Primary Care
	(§III-H) Supporting Drug Discovery and Treatment
(§**IV**) **Survey of Ongoing Research** This section surveys ongoing work across several types of data. It also provides brief summaries of outcomes and methodologies.	(§IV-A) Image Data Analysis
	(§IV-B) Textual Data Analysis
	(§IV-C) Voice Sound Data Analysis
	(§IV-D) Embedded Data Analysis
	(§IV-E) Pharmaceutical Research
(§**V**) **Bibliometric Analysis of COVID-19 Research** This section presents a bibliometric analysis of COVID-19 research.	(§V-A) Bibliometric Data Collection
	(§V-B) Peer-reviewed vs. Non-peer-reviewed publications
	(§V-C) Research Topics
	(§V-D) COVID-19 vs. Earlier Epidemics
(§**VI**) **Cross-Cutting Challenges** This section highlights challenges that researchers may face when performing data-driven research related to COVID-19.	(§VI-A) Data Limitations
	(§VI-B) Correctness of Results vs. Urgency
	(§VI-C) Security, Privacy, and Ethics
	(§VI-D) The Need For Multidisciplinary Collaboration
	(§VI-E) New Data Modalities
	(§VI-F) Solutions for the Developing World
(§**VII**) **Conclusions**	

artificial intelligence (AI), machine learning (ML), natural language processing (NLP), statistics, algorithms, modelling, simulation, and any other scientific methods that learn from structured and unstructured data. We recognise the importance of associated perspectives from the social sciences, ethics, history, and other humanities, but those areas are beyond the focus of this work.

In examining this growing landscape of data science research regarding COVID-19, we make the following five contributions. *First*, we present pressing research problems related to COVID-19, for which data scientists may be able to contribute. *Second*, we summarise publicly available COVID-19 datasets that are being used to drive research, and list how they could be utilised to address some of the aforementioned problems. *Third*, we survey some of the ongoing research in the area, highlighting the main topics covered. As our primary audience is computer scientists and engineers, we theme our discussion around the types of data analysis. *Fourth*, we broaden our analysis and present a bibliometric study. *Fifth*, bringing together our observations, we highlight some of the common challenges in this fast-moving space. We intentionally cast a wide net, covering research from several technical areas surrounding data science.

This paper builds upon recent reviews and perspective papers [5], [6] to help systematise existing resources and support the research community in building solutions to the COVID-19 pandemic. We have attempted to be comprehensive, however, in a rapidly-evolving field such as this, it is not possible to aim for exhaustiveness. Nonetheless, we hope that our work will provide a useful introduction to the field for researchers interested in this area.

The rest of this paper is organised as follows. In Section II, we present the details of available datasets and resources. In Section III, we present possible use cases where data science can help address COVID-19 challenges. In Section IV, we review contributions made by data scientists. In Section V we present a bibliometric analysis of the COVID-19 related papers. Next, we discuss common challenges facing this research in Section VI. Finally, Section VII concludes the paper. The organisation of the paper can be seen in Table I.

II. DATASETS AND RESOURCES

To enable research by the community, it is vital that datasets are made available. We start by surveying public datasets, some of which we summarise in Table II.

A. *COVID-19 Case Data*

The number of COVID-19 cases along with their geo-locations can help to track the growth of the pandemic and the geographical distribution of patients. Many countries are collecting and sharing infection information. One of the most used datasets is collated by John Hopkins University, which contains the daily number of positive cases, the number of cured patients and the mortality rates at a country and state/province level [20]. A further source of daily COVID-19 case data is available at Kaggle [25]. This dataset is annotated with other attributes such as patient demographics, case reporting date and location. Another epidemiological dataset, nCOV2019 [24], contains national and municipal health reports of COVID-19 patients. The key attributes are geo-location, date of confirmation, symptoms, and travel history. Similarly, the New York Times is compiling a state-wise dataset consisting of the number of positive cases and death count [26]. Whereas the above datasets are mostly based on statistics compiled by governmental administrations, other datasets are being collated using community surveys, requesting people to report infection rates among their social networks [30]. Common applications used with such data in the literature include data visualisation and predictive analytics [31].

A key limitation in these datasets is the divergence of testing regimes, which makes it challenging to compare results across countries [32]. It is estimated in one study[3] that the average detection rate of SARS-CoV-2 infections is just 6% worldwide. Similarly, variations in interventions, population densities and demographics have a major impact, as can be seen when contrasting, for example,

[3][Online]. Available: https://tinyurl.com/cov6percent

TABLE II
A LIST OF PROMINENT COVID-19 DATASETS

Dataset Name	Country/Region	Modality	Attributes	Ref.
BSTI Imaging Database	United Kingdom	CT scans data	Patient CT scans	[7]
COVID Chestxray Dataset	Italy	Chest X-ray scans and reports	X-Ray Image, date, patient, demographics, findings, location and survival information	[8]
COVID-CT-Dataset	All Countries	Chest CT-scans	Scans with associated labels	[9]
COVID-19 CT segmentation dataset	Italy	Lungs CT scans	JPG image scans with segmentation and label report	[9]
COVID-19 Russian CT-Scans Data	Russia	Lungs CT scans	Contains binary pixel masks of depicting the regions of interests	[10]
COVID-19 Community Mobility Reports	131 Countries	Mobility statistics with textual reports	Presence of people at grocery stores, pharmacies, recreational spots, parks, transit stations, workplaces, and residences	[11]
COVID-19 DATABASE	Italy	Radiology data	Xrays and demographics	[12]
COVID-19 Open Research Challenge	All Countries	Research articles dataset	Published date, author list, journal name, full text	[13]
Coronavirus Source Data	All Countries	Case statistics	Time series of confirmed daily COVID-19 cases for countries around the world	[14]
Coronavirus COVID19 Tweets	All Countries	Public Tweets on COVID-19	UserID, location, hashtags, tweet text	[15]
COVID-19 Korea Dataset	Korea	Case statistics	Patient routes, age, gender, diagnosed date	[16]
CHIME	All Countries	Case statistics	Daily number of susceptible, infected and recovered patient	[17]
Global research on COVID-19	All Countries	Database of research articles	Date, location, authors and journal	[18]
hCOV-19	All Countries	Genomic epidemiology	Genetic sequence and metadata	[19]
JHU CSSE COVID-19 Data	All Countries	Case statistics	Number of infections, number of cured patients, total mortality count, location	[20]
Kinsa Smart Thermometer Weather Map	USA	U.S. Health Weather Map	Temperature readings from internet-connected thermometers made by Kinsa Health.	[21]
LitCovid	All Countries	Dataset of research articles	Up-to-date research topics and geographic locations	[22], [23]
nCoV2019 Dataset	China, Japan, South Korea, Hong Kong, Taiwan, Thailand, Singapore	Epidemiological data	Patient demographics, case reporting date, location, brief history	[24]
Novel Corona-virus 2019 dataset	All Countries	Case statistics	Patient demographics, case reporting date, location, brief history	[25]
New York Times dataset	USA	State-wise cumulative cases	Date, state name, number of cases, death count	[26]
Public Corona-virus Twitter Dataset	All Countries	Tweet IDs	Twitter ID with location	[27]
RCSB Protein Data Bank	All Countries	Clinical and pathology	Genome sequences	[28]
RKI COVID19	Germany	Cases data	Number of infection cases	[29]

Japan vs. USA.[4] As such, regional prediction tasks are non-trivial, and we posit that temporal models such as Auto Regressive Integrated Moving Average (ARIMA) [33] and Long Short Term Memory (LSTM) [34] neural networks may be effective here.

B. COVID-19 Textual Data

The availability of rich textual data from various online sources can be used to understand the growth, nature and spread of COVID-19.

One prominent source is *social media*, for which datasets are already available covering COVID-19 discussions. There are open Twitter datasets covering Tweet IDs [27] and tweet text data [15]. These were gathered using Twitter's Streaming API to record tweets containing a series of related keywords, including "Coronavirus," "COVID-19," "N95," "Pandemic," etc. Another dataset of 2.2 millions tweets, alongside the code to collect more data is available [35]. This data could be used to monitor the spread of COVID-19, as well as people's reactions (e.g., to social distancing measures) using existing natural language processing techniques [36]–[38]. Sharma *et al.* [39] also built a public dashboard[5] summarising data across more than 5 million real-time tweets. There are also social media datasets that include image content: Zarei *et al.* [40] provide 5.3K Instagram posts related to COVID-19, including 18.5K text comments.

The wealth of *academic publications* in recent months is also creating a deluge of textual information. Information extraction from clinical studies is already being performed [41] using language processing models such as [42]. These bibliometric datasets can easily be collected from pre-print services such as arXiv, medRxiv, and biorXiv [43]–[45]. The White House has also released an open research articles dataset [13]. This dataset contains nearly 45,000 articles related to COVID-19, SAR-CoV-2 and other coronaviruses. These activities are mirrored across other organisations. For instance, in the US, The National Center for Biotechnology Information (NCBI) is providing up-to-date COVID-19 scientific literature [22], and WHO is compiling a database of recent research publications [18]. Closely related is the wealth of activity on Wikipedia, a community-driven encyclopedia, which already contains substantial information about COVID-19. The entirety of Wikipedia can be downloaded for offline analysis [46], and there are already pre-processed Wikipedia datasets focussing on COVID-19 available.[6]

C. COVID-19 Biomedical Data

Biomedical data can be used to support diagnosis, prognosis and treatment. A major source of data are physical medical reports (such as X-rays) or clinical pathology reports (genomic sequencing). As the current diagnosis and prognosis of COVID-19 often requires human interpretations, there is potential for applications of computer vision research, e.g., automated diagnosis from chest X-rays. Currently, there are some open-source COVID-19 X-ray scans such as the *COVIDx* dataset [47]. These can be used for training COVID-19 infection assessment and diagnosis models (exploiting known computer vision techniques [48]). Other X-ray

[4][Online]. Available: http://nrg.cs.ucl.ac.uk/mjh/covid19/index.html

[5][Online]. Available: https://usc-melady.github.io/COVID-19-Tweet-Analysis/

[6][Online]. Available: http://covid-data.wmflabs.org/

datasets for research are [8], [49], [50]. The latter contains date, patient, demographics, findings, location and survival information. However, there are some intrinsic challenges related to these X-ray datasets, such as the requirement of radiologists or clinicians for data labelling and annotation (before training models). As such, the datasets are still small, limiting the application of methods like convolutional neural networks.

Lung Computed Tomography (CT) scans can also be used for COVID-19 diagnosis and prognosis. Currently, there are datasets of lungs CT scans available. One of the datasets [9] covers 60 patients and comprises three class labels: ground glass, consolidation, and pleural effusion. The dataset is collected from 6, March to 13 March, 2020. A larger dataset of 288 CT scans is collected from 19 January to 25 March, 2020 [51]. The dataset has 275 CT scans of COVID-19 patients, which to the best of our knowledge, is the largest publicly available. Morozov *et al.* also provided COVID-19 lung CT scans from Russia containing binary pixel masks of CT-scans, depicting the regions of interest, collected between 1st March and 25th of April, 2020 [52]. The dataset is hosted at [10]. The National centre of UK is providing COVID-19 chest imaging dataset consisting of chest X-ray, CT and MR images of patients with suspected COVID-19 [53]. COVID-19 patient symptoms data are being collected by different healthcare organisations such as ELLIS Alicante Foundation [54], Leeds University Institute [55] and UK NHSx [56].

Besides the above physical scans, there are important genomic sequencing datasets available. The study of drug impact, protein-protein interactions and RNA structure in genomic data is an essential part of diagnosis test evaluations. Available datasets related to epidemiological and clinical data include RCSBdata [28] and GHDDI [57]. However, as COVID-19 has emerged very recently, these datasets are mostly incomplete or too small. For example, the biomedical datasets (see [51]) range from just a few up to 300 patients.

D. COVID-19 Datasets From Developing Countries

The impact and spread of COVID-19 in the developing world has become a matter of great concern. Several datasets have been gathered to study the nature of COVID-19 spread in developing countries such as Algeria, Nigeria, India, Pakistan, Kenya, Egypt, South Africa, and Latin America. Zhao *et al.* [58] studied COVID-19 spread, implications, prevention strategies, and control mechanism for African countries including Algeria, Nigeria, Senegal and Kenya, and South Africa. The dataset consists of confirmed, recovered, and death cases, taken from John Hopkins University.[7] COVID-19 cases data from India [25] and Pakistan [59] is hosted at Kaggle. These datasets provides city and state-wise COVID-19 cases, case reporting date, deaths, recoveries, patient demographics, and location.

E. COVID-19 Competition Datasets

To promote research in this area, there are several recent open data science competitions established on Kaggle (summarised in Table III). These are mostly based on the previously discussed data. For instance, the White House in coalition with some leading research groups (e.g., Kaggle and SGS Digicomply) has opened a new challenge using the earlier mentioned dataset of 45,000 research articles [13]. For this, there a few questions posed; for example, "*What do we know about virus genetics, origin, and evolution?*" For each task, there is an associated prize of $1000.

TABLE III
COVID-19 RELATED KAGGLE COMPETITIONS

Challenges	Aims
Answer 9 key questions	This challenge asks data scientists to understand COVID-19 faster by exploring 47,000 scholarly articles about COVID-19 and related coronaviruses.
COVID19 Global Forecasting	This challenge asks data scientists to predict the number of cases and fatalities by city between April 9 and May 7, 2020.
UNCOVER COVID-19	This challenge asks data scientists to use exploratory analysis to answer research questions that help support frontline responders.

The Roche Data Science Coalition (RDSC) also established the challenge "UNCOVER COVID-19" [60]. RDSC has rolled-out a multi-modal dataset collected from 20 sources and has posed questions prepared by front-line healthcare experts, medical staff, WHO and governmental policymakers. This dataset is mainly collected from John Hopkins, the WHO, New York Times and the World Bank. It includes local and national COVID-19 cases, geo-spatial data and social distancing polices. Participants are required to design solutions to address questions like "*which populations are at risk of contracting COVID-19?*" and "*which populations have contracted COVID-19 and require ventilators?*" Finally, the White House Office of Science and Technology Policy (OSTP) has opened a weekly challenge to predict the number of COVID-19 cases and fatalities at particular locations around the world [61]. Competitors are also required to unveil the factors associated with COVID-19 transmission rate and are asked to forecast the number of COVID-19 cases and deaths.

For those wishing to engage in these competitions, there are several helpful tools and guideline blogs available. These resources provide support for data pre-processing, visualisations, and the implementation of different frameworks. We provide a list in Table IV.

F. Other Supportive Datasets

As part of monitoring secondary factors related to COVID-19 and the surrounding interventions, there are several other relevant datasets. For example, air quality index statistics can be used as an indirect measure of social distancing polices, i.e., if movements are restricted there will be fewer vehicles (and pollution). For example, a recent study showed that the air quality of six populous world cities has improved between February and March 2020 due to the measures to combat COVID-19 [63]. The data is publicly available [64] as well as the related COVID-19 case data [20]. Mobility trace data [65] can also serve a similar purpose—a collection of such logs is available here [66]. Note that mobility datasets have already been re-purposed: Google has released community mobility reports for public health officials in 131 countries [11]. These reports are compiled using Google Maps and describe how busy places such as grocery stores, transit stations, and workplaces are. In a recent not-yet-peer-reviewed study [67], an indirect COVID-19 spread correlation is reported with wastewater samples. The wastewater sample data consists of 23 raw and 8 treated samples which is collected from three major wastewater treatment plants in France during 5, March to 7, April, 2020. It was found that all raw samples, and 6 out of 8 treated samples, tested positive for $SARS-CoV2$.

[7] [Online]. Available: https://coronavirus.jhu.edu/

TABLE IV
PROMINENT COVID-19 COMMUNITY RESOURCES

Resources	Details
AI against COVID-19	This webpage contains information related to recent papers, projects, and datasets for COVID-19.
AitsLab-Corona	This is an NLP toolbox and related text processing resources for SARS-CoV-2 and COVID-19 NLP research.
CORD-19	Allen Institute for AI with partners has prepared the COVID-19 Open Research Dataset (CORD-19), a free resource of over 52,000 scholarly articles
Amazon AWS	Amazon AWS public data lake for COVID-19 data analysis
CDC, USA	Centers for Disease Control and Prevention (CDC) COVID-19 research articles downloadable database.
ChemML [62]	ChemML is a machine learning and informatics program that support and advance the data-driven research in the domain of chemical and materials.
COVID-19 Graphs	This repository provides the tools to visualise the different statistics of COVID-19 using case data.
COVID-19 Data Portal	European Molecular Biology Lab-European Bioinformatics Institute (EMBL-EBI) and partners have set up this portal to bring together relevant datasets for biomedical data.
JHU's CSSE	Coronavirus COVID-19 Resource Page by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU).
NIH NLM LitCovid	LitCovid is a curated literature hub for tracking up-to-date scientific information about COVID-19. It provides central access to more than 3558 relevant articles in PubMed.
MATLAB resource	MATLAB based tutorial on deep learning based techniques for detecting COVID-19 using chest radiographs (in MATLAB).
MONTREAL.AI	Open source code and tools to model different aspects of COVID-19.
Partnership on AI	Page on AI- and Technology- Related COVID-19 Efforts
Vector Institute	This is a webpage that provides information about various resources and research tools for COVID-19.
WHO resource	This is a webpage of the WHO which contains updated details on the global research on COVID-19.
Telehealth Toolbox	This toolbox is providing an online treatment and telemedicine platform to combat COVID-19.
Zenodo	This is a webpage that provides information about various resources and research tools for COVID-19.
Openi	This is an open access biomedical image search engine that contains COVID-19 X-rays and CT-scan datasets.

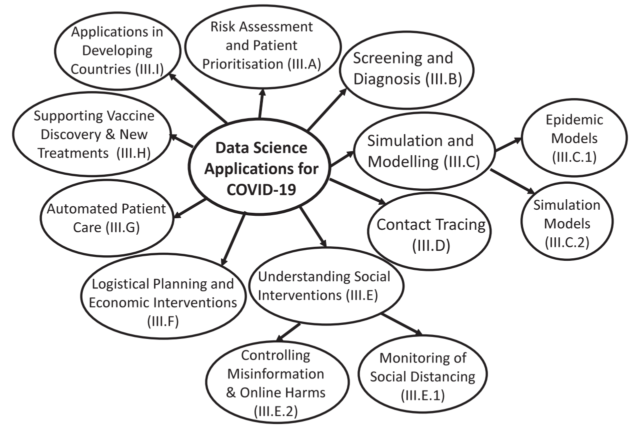


Fig. 1. Summary of data science applications and subsection structure.

III. DATA SCIENCE APPLICATIONS FOR COVID-19

Data science is a broad term covering topics such as Machine Learning (ML), statistical learning and time-series modelling. In this section, we summarise some of the key research use cases to which data scientists may be able to contribute. Figure 1 summarises the applications for COVID-19.

A. Risk Assessment and Patient Prioritisation

Healthcare systems around the world are facing unprecedented pressures on their resources (e.g., availability of intensive care beds, respirators). This creates the need to rapidly assess and manage patient risk, while allocating resources appropriately. In periods of peak load, this must be done rapidly and accurately, creating a substantial challenge for healthcare professionals who may not even have access to historical patient data. Various studies have already proposed algorithmic risk assessments of diseases such as cancer [68], diabetes [69], and cardiac-related diseases [70] with Artificial Neural Networks (ANNs). Due to diverse symptoms and disease trajectories, researching technologies for data-driven risk assessment and management in individual COVID-19 patients would be useful. For instance, traits like age, gender, or health state can be utilised to provide an estimate of mortality risk. This is particularly important when resources are limited, e.g., for patient prioritisation when Intensive Care Unit (ICU) resources are insufficient.

B. Screening and Diagnosis

A major issue facing countries with growing COVID-19 infection rates is the lack of proper screening and diagnosis facilities. This further complicates capacity management as well as social distancing measures, since those with mild symptoms are often unaware they carry the disease. A key use case is to develop remote computational diagnosis tools. Some already exist, which could be expanded, e.g., Babylon is a mobile app that provides medical advice via questioning. Other solutions could rely on valuable data from wearable or other monitoring devices. For instance, COVID-19 Sounds is a mobile app collecting audio of breathing symptoms to help perform diagnosis.[8] Oura wearable rings are tracking heart rate, and body temperature while you are sleeping or performing physical activity to track early viral symptoms.[9] We posit that such research will be particularly useful in developing countries that have a shortage of healthcare facilities [71]. Automated tools can further be developed to facilitate screening in larger groups of people (e.g., at airports), e.g., using computer vision based thermal imaging to detect fever [72].

C. Simulation and Modelling

Currie *et al.* [73] provide a detailed review of how models and simulations can help reduce the impact of COVID-19. For example, accurate epidemiological models are indispensable for planning and decision making. Here we discuss some of the potential modelling and simulation use cases.

1) Epidemic Models: Epidemic models are used to predict the macroscopic behaviour of an infectious disease. A key use case is developing and parameterising such models. For example, in epidemiology, *compartmental models* are widely used [74]. In these models, populations are divided into compartments and the flow of people among compartments is modelled using (ordinary) differential equations. For example, COVID-19's spread has recently been modelled using the SEIR model [75], [76], which models the flow of people between four states (or compartments): susceptible (S), exposed (E), infected (I), and recovered (R).

Generative models represent another broad class of models which proceed by generating consequences from causes (using hidden states and parameters). An example generative model is based upon ensemble or population dynamics that generate outcomes (new cases of COVID-19 over time) [76]. Such approaches can capture the effects of interventions (e.g., social distancing) and differences among populations (e.g., herd immunity) to predict what

[8][Online]. Available: http://www.covid-19-sounds.org/
[9][Online]. Available: https://ouraring.com

might happen in different circumstances in a single region [77]. Using (Bayesian) hierarchical modelling, one can combine several of these (epidemic) models to create a (pandemic) model of viral spread among regions [78]. For interested readers, websites offering COVID-19 forecasting have emerged,[10] each using a different model (although they should be treated with caution due to the uncertainty of such predictions [79] [80]).

Parameterising the above models requires up-to-date information on the virus spread. Thus, an important use case is finding ways to better capture such data. For instance, this could be done by processing social media information to identify people who are likely to have been infected, or even analysing ambulance call out data [81]. Another beneficial use case would be to develop ways to more accurately evaluate "*what-if*" scenarios with these models [79]. As an example, the initial policy of the UK government (of adopting almost no social isolation measures) was later changed based on results from an extended SEIR model from Imperial College London [82]. This model projected that without interventions there would be up to half a million fatalities, highlighting the importance of accurate predictions. A comprehensive review focused on modelling infectious disease dynamics in the complex landscape of global health can be seen at [83]. It is also worth mentioning that there are several national level modelling efforts underway. For instance, in the UK, The Royal Society has established the Rapid Assistance in Modelling the Pandemic (RAMP)[11] initiative with a focus on mechanistic modelling of disease spread and outcomes. They have also established the Data Evaluation and Learning for Viral Epidemic (DELVE)[12] which focuses on data-driven and inferential modelling of COVID-19.

2) Simulation Models: Simulation models have broad applicability and can be used in a variety of settings [73], including *decisions that affect disease transmission*—e.g., decisions related to quarantine and social distancing strategies; *decisions regarding resource management*—e.g., decisions related to capacity of in-patient hospital beds, critical care units, staffing, and resource allocation within and across regions; and *decisions about care*—e.g., deciding thresholds for admission and discharge of patients and minimising the impact on other patients. In particular, pandemics generate a large number of questions all of which cannot be answered by epidemiological models alone. A key use case is integrating a diversity of models into simulations that can be used to answer diverse questions. This might range from understanding disease spread to predicting the consumption of medical supplies for hospital management. By considering the range of model outputs, an additional benefit is that an estimate of uncertainty can be produced, which may help policy makers gauge expected benefits against risks. For interested readers, [73] provides an overview of the use of various simulation models, including those based on system dynamics [84], agent based models, discrete event simulations, and hybrid simulations.

D. Contact Tracing

Most countries reacted to the early stages of COVID-19 with containment measures. This typically involves rapidly identifying infected individuals, followed by quarantine and contact tracing. Countries, such as South Korea, conducted rigorous testing campaigns, which allowed other potentially infected contacts to be quickly quarantined. This approach has been seemingly successful in containing the outbreak [85]. A valuable use case can therefore be facilitating more rapid and comprehensive contact tracing at scale [86]. Smartphone contact sensing, online surveys and automated diagnosis have all been proposed to rapidly identify exposure [87]. For example, there are ongoing efforts to survey general populations via social media to learn of symptoms within individuals' social networks [30]. Even prior to COVID-19, FluPhone [88] used Bluetooth communications to identify contacts between people, and BlueDot monitored outbreaks of infectious diseases to alert governments, hospitals, and businesses [89].

If data from contact tracing is augmented with personal information such as geolocation, health characteristics and test results, there is the potential to continually update probabilistic estimates of the inferred states of individuals. This can also be used to test the sensitivity of different types of test, the patterns of disease progress for individuals and populations, as well as understanding how immunity declines over time. While this information could be very helpful, the benefits will have to be weighed against concerns about loss of individual privacy (see Section VI-C). As such, there has been extensive debate surrounding the design of contact tracing apps, primarily related to user privacy [90], [91]. For example, some apps have followed a centralised model whereby contacts are computed on the server. In contrast, others are decentralised, performing computation on the end device and therefore preventing a central point from recording contacts. See Section IV-D for further details.

E. Understanding Social Interventions

Governments have taken steps to manage social interactions as part of their response to COVID-19. We highlight two main use cases of relevance.

1) Monitoring of Social Distancing: Many governments have implemented social distancing strategies to mitigate the spread of COVID-19. This is a non-pharmaceutical intervention that reduces human contact within the population [92] and therefore constrains the spread of COVID-19 [93]. Data science can support contact tracing for the monitoring of social distancing, for instance by extracting social media data and using language processing techniques [94], [95]. These analyses could also help in keeping record of interactions to be used as individuals develop symptoms. Furthermore, these could be used for general population tracking to understand compliance with social distancing. This could then be complemented with other datasets (e.g., cellular trace data or air pollution monitoring [63]) to better understand human mobility patterns in the context of social distancing. Similar to the previous case, these solutions present complex trade offs with regards to privacy (see Section VI-C).

2) Controlling Misinformation & Online Harms: The spread of misinformation can undermine public health strategies [96] and has potentially dangerous consequences [97], [98]. For example, online rumours accusing 5G deployments of causing COVID-19 led to mobile phone masts being attacked in the UK [99]. Wikipedia maintains an up-to-date list of misinformation surrounding COVID-19 [100]. This confirms the spread of a number of dangerous forms of misinformation, e.g., that vinegar is more effective than hand sanitiser against COVID-19. Naturally, users who believe such misinformation could proceed to undermine public health. One important use case would therefore be to develop classifiers and

[10]For example: (1) *COVID-19 worldwide peak forecasting method* ([Online]. Available: https://www.people.vcu.edu/~tndinh/covid19/en/) and (2) *COVID-19 forecasting* ([Online]. Available: http://epidemicforecasting.org/)

[11][Online]. Available: https://royalsociety.org/topics-policy/health-and-wellbeing/ramp/

[12][Online]. Available: https://tinyurl.com/y99c33dc

techniques to stem this flow. For example, Pennycook *et al.* [101] are testing simple interventions to reduce the spread of COVID-19 misinformation. An infodemic observatory analysing digital responses in online social media to COVID-19 has been created by CoMuNe lab at Fondazione Bruno Kessler (FBK) institute in Italy, and is available online.[13] The observatory uses Twitter data to quantify collective sentiment, social bot pollution, and news reliability and displays this visually.

F. Logistical Planning and Economic Interventions

COVID-19 has created serious challenges for healthcare supply chains and provisioning. This includes personal protective equipment such as masks and gowns, alongside intensive care equipment like test kits, beds, and ventilators. There is a history of applying machine learning to logistical planning, e.g., by Amazon Fulfilment.[14] A simple use case would be to apply data science techniques to help supply chain management for healthcare provisioning. This can also be used to preemptively allocate resources, e.g., researchers from the University of Cambridge are using depersonalised data (like lab results and hospitalisation details) to predict the need for ventilation equipment.[15] This use case is critical for ensuring appropriate equipment is available on time.

Social distancing measures are also having a major impact on the global economy [102], [103]. As organisations emerge from economic hibernation they will be challenged to return to normal levels of service and operation given disruptions to their supply chains and workforce. Data scientists might be able to assist in identifying problems limiting recovery. For instance, governments can use data science to identify optimal economic interventions at a high level of granularity. Companies can use data science to detect unusual patterns of behaviour in the market or in their own customer base.

G. Automated Patient Care

The pandemic has triggered a shortage of healthcare workers. To alleviate this, automated primary care tools, such as remote chatbots and expert systems, could be developed. Such systems can help people in providing information about the outbreak, symptoms, precautionary measures, etc. For instance, an interactive chatbot by the WHO and Rakuten Viber aims to provide accurate information about COVID-19 to people in multiple languages [104]. Automated healthcare methods could also be utilised to help monitor the conditions of COVID-19 patients in emergency care [105].

Another use case would be to gather and collate observational data to monitor the efficacy of treatments for certain patient types, enabling decision support for better personalised patient treatment. For example, the DeCOVID project at the Alan Turing Institute is attempting to use clinical data to identify factors and generate insights that can lead to more effective clinical strategies. Similarly, physical and psychological (self) recordings could be used to augment personal plans. Due to the need to rapidly discharge patients from hospitals, further monitoring could continue remotely with the help of various existing remote care devices. For instance, AliveCor [106] Kardia Mobile 6L device can assist healthcare professionals in remotely managing COVID-19 patients by measuring QTc (heart rate corrected interval) through a six-lead personal ECG [107]. CLEW's TeleICU solution is another remote care solution that can help identify respiratory deterioration [108]. FreeStyle Libre [109] is an app that connects people with their doctor remotely. Home pulse oximetry [110] can also help decrease COVID-related mortality by performing detection remotely. Researchers at Northwestern University and Shirley Ryan AbilityLab in Chicago have developed a wearable device and are creating a set of algorithms tailored to catch early signs and symptoms associated with COVID-19.[16] Data scientists could contribute to these remote care solutions. The USA Department of Health and Human Services (HHS) has recently announced an expansion of the coverage of Medicare for tele-health visits in a bid to better manage the COVID-19 outbreak by [111]. Similarly, the COVID-19 Telehealth Program in the USA [112] announced a $200 million funding in response to COVID-19 for provision of virtual health services by healthcare practitioners to patients at their homes or remote locations [113].

H. Supporting Vaccine Discovery and New Treatments

The international effort to discover or re-purpose drug treatments and vaccines can also benefit from extensive data science work predating COVID-19 [114]. For example, computational methods can reduce the time spent on examining data, predicting protein structures and genomes [115], [116]. It can also assist in identifying eligible patients for clinical trials [117], which is often a time-consuming and costly part of drug development. There is also substantial scope for applying advanced methods to managing trials, such as applying Bayesian clinical trials to adapt treatments based on information that accrues during the trial [118]. This may be critical in expediting the delivery of drug treatments, and we argue this is another area where data scientists can contribute. The field of *network medicine*, which applies techniques and insights from network science to medicine, is also being actively pursued for the purpose of developing and validating computational tools that can help identify drug repurposing opportunities [119].

I. Applications in Developing Countries

The above applications have primarily targeted developed countries. However, there are some unique challenges in developing countries where data science could help. For instance, healthcare systems in developing countries is struggling to provide health services to their populations. This naturally becomes more critical during a pandemic situation, as there is a shortfall of healthcare professionals in developing countries. To tackle the current situation, developing countries need to strengthen their capacity in terms of screening COVID-19 cases and health facilities and physicians need to be equipped for management of identified cases. In all these cases, data can be utilised for streamlining procedures. For instance, mass screening can be performed by utilising mobile devices, automated and remote care can be provided to the people living in rural areas, country-wide (targeted) public awareness campaigns using text messages can be carried out on personal hygiene and social distancing, and contact tracing of positive cases and hot spot identification can be performed to control and spread minimisation.

IV. Survey of Ongoing Data Science Related COVID-19 Research

The above provides an overview of public datasets. Next, we detail some of the ongoing research in this space. We theme this

[13] COVID19 Infodemics Observatory. [Online]. Available: https://covid19obs.fbk.eu/

[14] [Online]. Available: https://services.amazon.co.uk/services/fulfilment-by-amazon/features-benefits.html

[15] [Online]. Available: https://tinyurl.com/CambridgeCenterAIMedicineCOVID

[16] [Online]. Available: https://tinyurl.com/WearableCOVID19MonitoringDev

section around the above datasets and summarise key studies in Table V.

A. Image Data Analysis

Various studies [120]–[122] have used computer vision algorithms to speed up the process of disease detection across several imaging modalities with some studies demonstrating that image analysis techniques have the potential to outperform expert radiologists [123], [124]. To diagnose COVID-19, two medical imaging modalities (CT and X-ray) have been experimented with [125], which we discuss below.

1) Computed Tomography (CT) Scans: Recent studies have found that radiologists can diagnose COVID-19 using Chest CT scans with lower false positive rates [126], [127] than other imaging modalities such as X-ray and Ultrasound scans. Thus, many deep learning (DL) techniques related to CT scans have been proposed to expedite the diagnosis process. Wang *et al.* [128] utilise DL methods to detect radiographical changes in COVID-19 patients. They evaluate the proposed model on the CT scans of pathogen-confirmed COVID-19 cases and show that DL can extract radiological features suitable for COVID-19 diagnosis. Xiaowei *et al.* [129] present a method for the automatic screening of COVID-19 in pulmonary CT scans using a 3D DL model with location-attention. They achieve promising accuracy to identify COVID-19 infected patients scans from other well-known infections. Chen *et al.* [130] exploit the UNet++ architecture [131] to detect suspicious lesions on CT scans. They trained their model on 289 scans and test on 600 scans. They achieve 100% accuracy in identifying the suspicious areas in CT scans of COVID-19 patients. Ophir *et al.* [132] employ 2D and 3D convolutional neural networks (CNNs) to calculate the Corona score (which represents the evolution of the disease in the lungs). They estimate the presence of the virus in each slice of CT scan with a 2D CNN and detect other lung diseases (i.e., lung nodule) by using a 3D CNN. Similarly in [133], a neural network (COVNet), is developed to extract visual features from volumetric chest CT exams for the detection of COVID-19. The study suggests that DL-based models can accurately detect COVID-19 and differentiate it from community acquired pneumonia and other lung diseases.

2) X-ray Scans: There has also been work on processing X-rays scans. Although less sensitive than CT scans, they are less invasive, have a lower ionising radiation dose, and are more portable. Following the IR(ME)R 17 guidelines, ionising radiation dose should be kept *As Low As Reasonably Achievable* (ALARA) whilst still producing an image of diagnostic quality. [134]. Ezz *et al.* [135] propose a DL-based framework (COVIDX-Net) to automatically diagnose COVID-19 in X-ray images. COVIDX-Net includes seven different CNN models, such as VGG19 [136] and Google MobileNet [137]. The models can classify the patient status as either COVID-19 negative or positive. However, due to a lack of data, the technique is validated on only 50 X-ray images, among which 25 were of confirmed corona patients. Linda *et al.* [47] introduce another DL-based solution tailored for the detection of COVID-19 cases from chest X-ray images. They also develop a dataset named COVIDx and leverage it to train a deep CNN. In [138], three different CNN-based models (i.e., ResNet-50, Inception and Inception-ResNet) are employed to detect COVID-19 in X-rays of pneumonia infected patients. The results show that the pre-trained ResNet-50 model [139] performs well, achieving 98% accuracy. Similarly, Farooq *et al.* [140] provide the steps to fine-tune a pre-trained ResNet-50 [139] architecture to improve model performance for detecting COVID-19 related abnormalities (called COVID-ResNet). Prabira *et al.* [141] use DL to extract meaningful features from chest X-rays, and then trained a support vector machine to detect infected patients. We also briefly note that several companies have released commercial solutions, some of which are freely available, e.g., Lunit [142] CXR solution for COVID-19 and VUNO Med [143] solution for chest CT and X-ray scans. Such solutions help to expedite the initial screening of COVID-19.

B. Textual Data Analysis

Researchers are currently utilising text mining to explore different aspects of COVID-19, mainly from social media and bibliometric data. To assist in this, Kazemi *et al.* [144] have developed a toolbox for processing textual COVID-19 data. This toolbox comprises English dictionaries related to the disease, virus, symptoms and protein/gene terms.

In terms of social media research, Lopez *et al.* [145] explore the discourse around the COVID-19 pandemic and government policies. They use Twitter data from different countries in multiple languages and identify the popular responses to the pandemic using text mining. Similarly, Saire and Navarro [146] use text mining on Twitter data to show the epidemiological impact of COVID-19 on press publications in Bogota, Colombia. Intuitively, they find that the number of tweets is positively correlated with the number of infected people in the city. Schild *et al.* [147] inspect Twitter and 4Chan to measure sinophobic behaviour driven by the pandemic. Cinelli *et al.* [148] analyse Twitter, Instagram, YouTube, Reddit and Gab data on COVID-19. They find different volumes of misinformation on each platform. Singh *et al.* [149] also monitor the (mis)information flow across 2.7M tweets, and correlate it with infection rates to find that misinformation and myths are discussed, but at lower volume than other conversations. For those seeking easy access to this information, the FBK institute is collecting COVID-19 related tweets to visualise the presence of bots and misinformation.[17]

In terms of bibliometric analysis, Li *et al.* [150] analyse research publications on other coronaviruses (e.g, SARS, MERS). This is used to build a network-based drug re-purposing platform to identify drugs for the treatment of COVID-2019. Using module detection and drug prioritisation algorithms, authors identify 24 disease-related human pathways, five modules and suggest 78 drugs to re-purpose. The rapid growth in COVID-19 literature further led Hossain *et al.* [151] to perform a bibliometric analysis of COVID-19 studies. They review relationships, citations and keywords.

Finally, there is work processing text data from patient records. Roquette *et al.* [152] train a deep neural network to forecast patient admission rates using unstructured text data available for triage. There are also other studies that utilise text data mining techniques to explore the important aspect of current situation.

C. Voice Sound Data Analysis

The most common symptoms of COVID-19 are linked to pneumonia, with the main mortality risk being cardiovascular and chronic respiratory diseases. Hence, audio analysis is a potential means for lightweight diagnosis. There is work performing diagnosis with respiratory and lung sound analysis [153], using low-cost

[17] [Online]. Available: https://covid19obs.fbk.eu/

TABLE V
SUMMARY OF DATA SCIENCE WORK RELATED TO COVID-19. PAPERS ARE CATEGORISED BASED ON THE DATASET USED

Authors	Area	Modality/Data Type	Technique	Methodology
Wang et al. [128]		Chest CT scans	InceptionNet on ROIs	InceptionNet is used to detect the anomalies related to COVID-19 infection in lungs CT scan.
Xu et al. [129]		Chest CT scans	3D CNNs	3-D CNN models used to classify the COVID-19 infected regions in CT scans.
Chen et al. [130]		Chest CT scans	UNet++	UNet++ architecture has been used to identify the suspicious areas in CT scans.
Gozes et al. [132]		Chest CT scans	2D + 3D CNNs	2D and 3D CNNs models have been simultaneously employed to quantify the infection in the lungs of COVID19 patients.
Lin et al. [133]		Chest CT scans	CNN	COVNet; CNN-based model is developed to detect COVID-19 in chest CT scans.
Shan et al. [182]		Chest CT scans	DNN	DL-based segmentation system is developed to quantify infected ROIs in lung CT scans.
Zhang et al. [183]		Chest CT scans	DenseNet	Used DenseNet-like architecture and optimised it for classification task to detect COVID-19 infection.
Wang et al. [184]		Chest CT scans	Pre-training + DNN	Pre-trained DNN has been used to improve detection of COVID-19 in lungs scans.
Mucahid et al. [185]		Chest CT scans	Features + SVM	GLCM, LDP, GLRLM, GLSZM, and DWT algorithms are used as feature extraction and SVM for classification.
Zhao et al. [51]		Chest CT scans	CNN	Developed a public dataset and employed CNN for COVID-19 detection on chest CT scans.
Gozes et al. [186]		Chest CT scans	U-Net + ResNet	Used UNet for lung segmentation, ResNet for 2D slice classification and fine grain localisation for detection of infected regions in lungs.
Asnaui et al. [187]	Image Analysis	Chest X-rays & CT images	Fine tuning + CNNs	Various CNN-based models used for binary classification in COVID-19 detection on pneumonia affected X-ray and CT images.
Ezz et al. [135]		Chest X-rays	CNN-based models	Introduced COVIDX-Net, which includes seven different CNN models for classification of COVID-19 infected X-rays.
Linda et al. [47]		Chest X-rays	ResNet	An open source solution (COVID-Net) utilized ResNet to detect COVID-19.
Narin et al. [138]		Chest X-rays	ResNet50, InceptionV3 and InceptionResNetV2	Different CNN-based models are used to detect COVID-19 pneumonia infected patients chest X-rays.
Prabira et al. [141]		Chest X-rays	DNN + SVM	Used DNN to extract meaningful information from X-rays and SVM for classification of corona affected X-rays.
Farooq et al. [140]		Chest X-rays	Fine-tuning + ResNet	Devised multi-stage fine-tuning scheme to improve performance and training time.
Abbas et al. [188]		Chest X-rays	Transfer learning (TL) + CNN	Employed TL and used previously developed CNN, called Decompose, Transfer, and Compose (DeTraC).
Rajaraman et al. [189]		Chest X-rays	Weakly labeled data augmentation	An image argumentation technique has been proposed to create the chest X-ray images for fine-tuing of pre-trained models.
Yujin et al. [190]		Chest X-rays	CNN + Patch-wise	Used limited data and employed patch-wise data argumentation for detection COVID-19
Alqudah et al. [191]		Chest X-rays	CNN, SVM, and Random Forest	Applied various ML techniques for classification of COVID-19 infected X-rays.
Goshal et al. [192]		Chest X-rays	Bayesian CNN	Investigated the significance of dropping weights BCNN.
Fatima et al. [193]		Chest X-rays	CNN	Trained CNN for COVID-10 detection in X-rays.
Xin et al. [91]		Chest X-rays	DenseNet	Used DenseNet Architecture [194] for COVID-19 detection in X-rays.
Karim et al. [195]		Chest X-rays	DNN	Used neural ensemble method for classification and provided human-interpretable explanations of the predictions of COVID-19.
Ioannis et al. [196]		Chest X-rays	TL + CNN	TL is used for extracting patterns from common bacterial pneumonia patients X-rays using CNN to detect COVID-19.
Jahanbin et al. [197]		Twitter data	Evolutionary algorithm	Fuzzy rule-based evolutionary algorithm was used to timely detect outbreaks of the COVID-19 by using Twitter data.
Zhao et al. [198]	Text data Mining	Sina Microblog hot search list	Content mining algorithms	This work investigates the public's response at the beginning (December 31, 2019, to February 20, 2020) of the COVID-19 epidemic in China.
Li et al. [199]		Weibo data	SVM, Naïve Bayes (NB), RF	Weibo data was used to characterise the propagation of situational information in social media during COVID-19.
Schild et al. [147]		Twitter & 4Chan data	word2vec	Authors look at rise of COVID-19 related sinophobic abuse on Twitter and 4Chan.
Prabhakar et al. [200]		Twitter data	Topic modelling	In this work, the information flow on twitter during COVID-19 pandemic was studies using topic modelling.
Stephany et al. [201]		Risk reports data	Multiple text mining algorithms	Text mining methods are used to identify industry-specific risk assessments related to COVID-19 in real-time.
Zhavoronkov et al. [57]		Crystal structure, homology modelling, and co-crystallised, ligands	Generative models	Generative models were used to generate the molecules for the 3C-like protease that can act as potential inhibitors for SARS-CoV-2.
Hofmarcher et al. [202]		Drug-discovery databases	DNNs	Used ChemAI [203], [204], a DNN trained on millions of data pints across 3.2 millions of molecules, for screening favourable inhibitors from the ZINC database [205] for SARS-CoV-2.
Beck et al. [206]	Pharmaceutical Research	SMILES strings, amino acid sequences	Deep learning model	Authors utilise a pre-trained drug-target interaction model to predict commercially available antiviral drugs for COVID-19.
Kim et al. [207]		SMILES strings, amino acid sequences	AI-based prediction	A binding affinity prediction platform is used to detect available FDA approved drugs that can block SARS-CoV-2 from entering cells.
Richardson et al. [208]		Biomedical data	AI-driven knowledge graph	Authors use BenevolentAI to search for approved drugs that can block the viral infection process.
Stebbing et al. [209]		Biomedical data	AI-driven knowledge graph	This study examines approved antiviral and anti-inflammatory treatments for COVID-19.
Vijil et al. [210]		SMILES	Generative models	Design drug candidates specific to a given target protein sequence. They release around 3000 COVID-19 drug candidates.

smartphones [154]. High mortality risk groups can also be continuously monitored using speech analysis [155]. The patterns of coughs [156], [157], sneezing [156], throat clearing and swallowing [158] can all be analysed using speech processing. At present, COVID-19 speech data has limited availability, although the potential benefits are highlighted in [153]. Thus, mobile apps like COVID-19 Sounds attempt to collect large audio datasets. [159] presents an app called AI4COVID-19 for the diagnosis of COVID-19. It requires a 2 second cough sample and provides the preliminary diagnosis within a minute. This work confirms the feasibility of COVID-19 detection using cough samples with promising results.

D. *Embedded Sensor Data Analysis*

Embedded sensor data is being used for remote patient care and diagnosis [160]. This can include mobility data, physiological vital signs, blood glucose, body temperature, and various other movement-related signals. In [161], the authors develop a system utilising real-time information, including demographic data, mobility data, disease-related data, and user-generated information from social media. The proposed system, called α-Satellite, can provide hierarchical community-level risk assessment that can inform the development of strategies against the COVID-19 pandemic. Google has also been using location data from smartphones to show people's movement during the pandemic [162]. Another study [163] presents the design of a low-cost framework for the detection of COVID-19 using smartphone sensors. They propose the use of the mobile phones of radiologists for virus detection. They highlight that the proposed framework is more reliable as it uses multi-readings from different sensing devices that can capture symptoms related to the disease.

Another recent study [86] concluded that COVID-19's "spread is too fast to be contained by manual contact tracing". To address this, disease tracking apps [88] use contact/location sensor data. The simplest ones aim to understand the spread of the disease, particularly mild cases that are not routinely lab tested. For example, the COVID Symptom Tracker app[18] and COVID Near You[19] service. Others, like Hong Kong's StayHomeSafe and Poland's Home Quarantine app [164], try to monitor if people obey quarantine rules (via geofencing). More advanced solutions can notify users if they have come into contact with somebody infected. Examples include China's Close Contact Detector app [165], China's complementary QR health code system [166]), Singapore's TraceTogether [167] app, and Israel's HaMagen [168] app.

We note that one critical challenge in the above apps is protecting user privacy [169], [170]. For instance, uploading contact data for server-side computation could create a nation-wide database of social relationships, particularly in countries where usage is mandatory. To address this, Decentralised Privacy-Preserving Proximity Tracing (DP-3T) [171] was proposed. This is a mobile app that offers privacy-preserving alerts for people who may have recently been in contact with an infected person. TraceSecure [172] supports similar features based on homomorphic encryption, whereas [173] offers privacy guarantees via private set intersection. Apple and Google have announced a partnership to develop their own privacy-preserving contact tracing specifications based on Bluetooth.[20]

[18][Online]. Available: https://covid.joinzoe.com/

[19][Online]. Available: https://www.covidnearyou.org/

[20][Online]. Available: https://www.apple.com/covid19/contacttracing/

E. *Pharmaceutical Research*

There is work to support the search for COVID-19 pharmaceuticals. This has received substantial attention in recent months in an attempt to build models to explore the 3D structure of SARS-CoV-2. In [174], the authors use the AlphaFold model to predict the structures of six proteins related to SARS-CoV-2. AlphaFold [175] is a DL model based on a dilated ResNet architecture [139], which predicts the distance and the distribution of angles between amino acid residing on protein structure. In [176], the authors use a DNN-based model for de novo design of new small molecules capable of inhibiting the chymotrypsin-like (3CL) protease—the protein targets for corona-viruses. Based on the results they were able to identify 31 potential compounds as candidates for testing against SARS-CoV-2. Studies also attempt to improve the RT-PCR test by utilising ML and novel genome technologies. Metsky *et al.* [177] employ CRISPR to develop assay designs for the detection of 67 respiratory viruses.

As well as the above, studies have utilised ML models to speed up drug development. Hu *et al.* [178] exploit a multi-task DNN for the prediction of potential inhibitors against SARS-CoV-2. They aim to identify existing drugs that can be re-purposed. Zhang *et al.* [179] perform DL-based drug screening against 4 chemical compound databases and tripeptides for SARS-CoV-2. They provide a list of potential inhibitors that can help facilitate drug development for COVID-19. Tang *et al.* [180] propose the use of reinforcement learning (RL) models to predict potential lead compounds targeting SARS-CoV-2. Similarly, in [181] the authors propose an antiviral discovery approach using deep RL.

Finally, pharmaceutical interventions must go through clinical trials before deployment. Accelerated clearance pathways for COVID-19 studies have been established by several regulators [211]. As of March 24, 2020, 536 relevant clinical trials were registered. A major barrier though is recruiting suitable patients. Data-driven solutions are available to rapidly identify eligible participants [117], [212] and data collection platforms already exist to monitor symptoms remotely [213].

V. Bibliometric Analysis of COVID-19 Research

COVID-19 has been accompanied by a surge in the number of related academic publications. We next explore how the academic community has reacted to this urgency by conducting a bibliometric analysis of the academic publications related to COVID-19.

A. *Bibliometric Data Collection*

There are many data repositories which contain COVID-19 research articles, both peer-reviewed [22], [214], [215] and non-peer reviewed [43]–[45]. We gather data from pre-print archives and from the Scopus database. Each entry includes title, authors, journal, publication date, etc. Our dataset covers papers on COVID-19 from all of the mentioned sources till June 19, 2020. We extracted these papers from the corpus of papers using keyword matching on titles and abstract of the paper. We use "COVID-19", "COVID", "CoronaVirus", "Corona Virus", "Pandemic", "Epidemic", and "SARS-CoV-2" as candidate keywords. Finally we did a manual check to confirm that extracted papers do not include any unrelated papers. In total, the dataset covers 11590 publications, of which 6461 are pre-prints and 5129 are from peer reviewed journals.

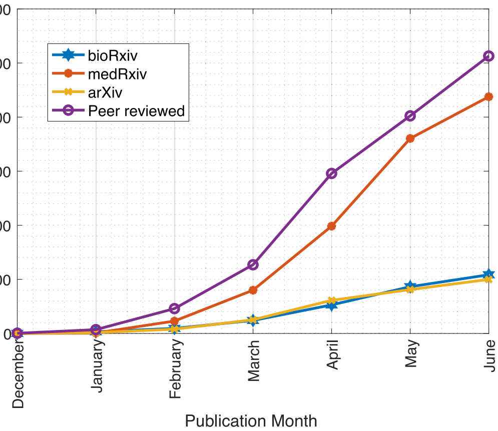


Fig. 2. Cumulative distribution of publications per month on COVID-19 (data gathered till June 19, 2020).

B. Peer-Reviewed vs. Non-Peer-Reviewed Publications

The pandemic has resulted in the rapid production of academic material, much of which is yet to go through the peer review process due to the urgency of dissemination.

Figure 2 presents the cumulative number of COVID-19 related papers published since December, 2019 including non-peer-reviewed COVID-19 literature. We see that the number of papers has increased dramatically since the beginning of January. To date, non-peer-reviewed articles are the most numerous (bioRxiv, medRxiv and arXiv combined), whereas peer reviewed articles are also increasing. By far the most active outlet is medRxiv, which has published 67% of all non-peer reviewed papers in our dataset. Figure 3 complements the above analysis by presenting the geo-distribution of both groups of publications. A major part of COVID-19 research has been contributed by USA. China holds the second position in terms of research contributions.

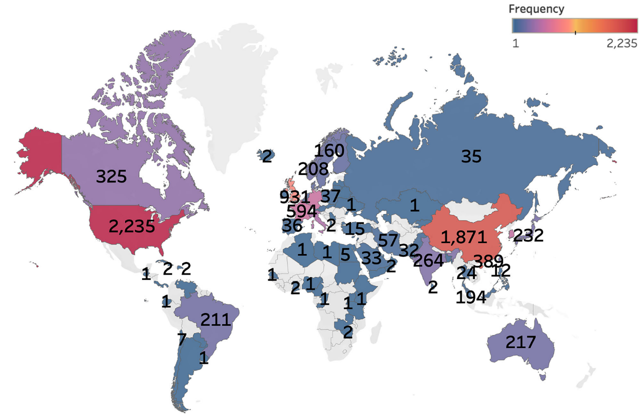
C. Research Topics

We next use topic modelling to identify core sub-topics within the publications. For this, we use Latent Dirichlet Allocation (LDA) [216]. This algorithm extracts and clusters abstract topics that exist within the papers. We have tagged these papers manually based on their title and abstract. Note that we split the results into peer reviewed vs. pre-print publications.

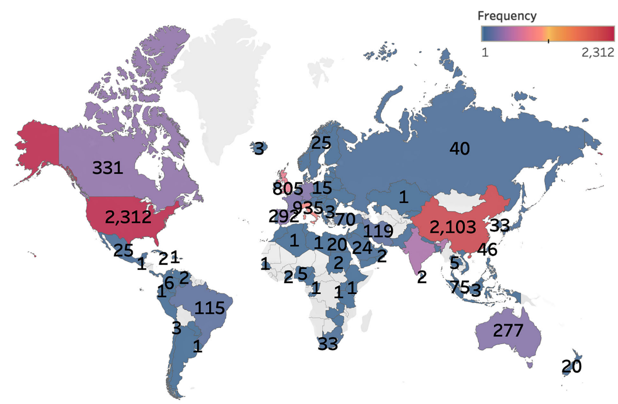
Table VI, shows the list of topics observed in data science related COVID-19 papers. These topics show that data science research on COVID-19 is being carried out using various techniques and algorithms. Noteworthy algorithms and techniques include multidimensional kernel estimation, Bayesian learning, and deep learning based epidemic forecasting with synthetic information (TDFESI). We hope that these results will be useful to the community in identifying key topics receiving coverage.

D. COVID-19 vs. Earlier Epidemics

We conclude our bibliometric analysis by briefly comparing the rate of publication for COVID-19 research vs. prior epidemics. For this, we select Ebola and SARS-CoV-1. Figure 4 presents a time series for the first 3 years of peer reviewed publications. Note that the X-range differs and, naturally, we only have data since December 2019 for COVID-19.



(a) Peer reviewed papers



(b) Pre-print (non-peer-reviewed) papers

Fig. 3. Publication count of different countries on COVID-19 (data gathered till June 19, 2020).

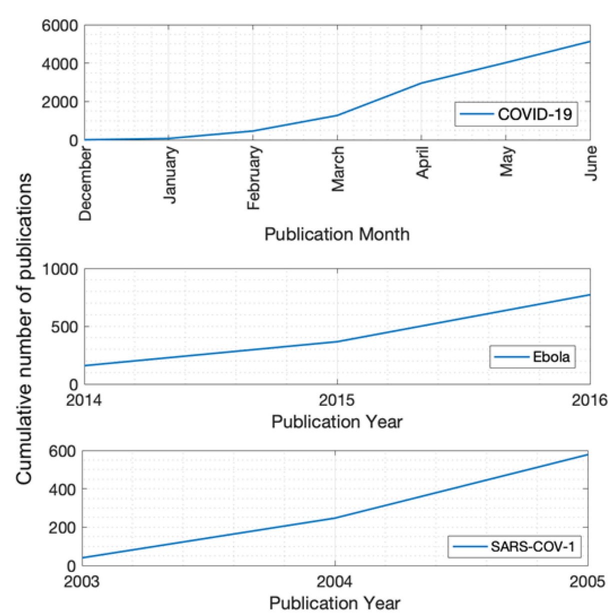


Fig. 4. Cumulative publication rates for peer-reviewed publications in COVID-19, SARS-CoV-1, Ebola (data gathered till June 19, 2020). Note the different X-ranges.

TABLE VI
TOP TOPICS DISCUSSED IN COVID-19 *DATA SCIENCE BASED* RESEARCH PAPERS

Topic No.	Extracted Topics
1	(model, number, use, countries, passengers, access, china, reduction, outbreak, result)
2	(dimensions, kernel, complex, structure, spectral, time, network, distance, base, infection-link)
3	(learn, image, covid-19, detect, dataset, feature, patient, predict, risk, death)
4	(sample, network, estimate, image, detrace, mean, transfer, covid-19, x-ray, medics)
5	(epidemic, risk, data, detect, method, health, influence, outbreak, measure, covid-19)
6	(model, graph, number, mixture, rate, predict, infect, algorithm, china, covid-19)
7	(sepsis, learn, feature, clinic, severe, treatment, differ, disease, auroc, automate)
8	(forecast, data, epidemic, high-resolute, tdefsi, method, ili, disease, mds, perform)
9	(data, world, period, trend, death, register, pandemic, epidemic, model, covid-19)
10	(crime, virus, sars-cov-2, genotype, isolate, mutate, global, genome, public, policies)

(b) Top topics in *peer-reviewed data science based* COVID-19 papers

Topic No.	Extracted Topics
1	(estimate, number, outbreak, model, epidemic, method, data, rate, dynamics, coronavirus)
2	(infect, estimate, death, risk, disease, quarantine, asymptomatic, coronavirus, intervene, individual)
3	(number, case, infect, model, data, epidemic, patient, control, peak, forecast)
4	(report, forecast, use, cumulative, predict, growth, data, outbreak, transmission, improve)
5	(coronavirus, quarantine, countries, data, suspect, measure, effect, ratio, intervention, transmission)
6	(outbreak, coronavirus, period, transmission, peak, predict, reproduction, mean, intervention)
7	(case, cities, model, number, outbreak, fit, dynamics, prevent, trend, predict)
8	(outside, travel, cause, viral, range, detect, phase, pneumonia, incubate, quarantine)
9	(case, estimate, epidemic, global, export, forecast, risk, incident, reproduction, severe)
10	(control, outbreak, trace, isolate, transmission, symptom, prevent, model, onset, strategies)

We see that COVID-19 literature is growing faster than any prior epidemic. There have been more peer-reviewed publications ($\sim$5.1K) in around 6 months for COVID-19 than there were in 3 years for SARS-COV-1 and Ebola. Furthermore, as noted earlier, there are even more pre-prints being released which means that COVID-19 has rapidly overtaken other epidemics in terms of academic attention. Of course, this is driven in-part by the wider geographic coverage of COVID-19, impacting numerous highly research active countries (e.g., China, USA, UK, Germany).

VI. CHALLENGES IN DATA SCIENCE RELATED COVID-19 RESEARCH

In this section, we highlight some of the most important data science challenges. We specifically focus on cross-cutting challenges that impact all previously discussed use cases.

A. Data Limitations

Data science systems typically learn and improve as more data is gathered over time. Ideally, the data should be of high fidelity and voluminous. For many of the above use cases, extensive labelled datasets are not yet available, e.g., for speech analysis. Although there are a few publicly available datasets for medical images and textual analysis, these datasets are small compared to the requirements of deep learning models. For example, in the case of biomedical data, sample sizes range from a few up to 60 patients (see [9]). The scarcity of measured data is frequently due to the distributed nature of many data sources. For example, electronic healthcare records are often segregated on a national, regional, or even per-hospital level. A key challenge is therefore federating these sources, and overcoming practical differences across each source, e.g., in terms of schemas. Thus, better and more automated approaches to data munging, data wrangling etc. may be critical in attaining fast, reliable and robust outcomes. Common standards and international collaboration will help.

Beyond these challenges regarding data availability, there are also major challenges within the data itself. The time-critical nature of this research is causing hurdles in developing certain types of high-quality dataset. For instance, by the time social media data is collected, curated and annotated it can become obsolete. Due to this, COVID-19 datasets and their causal interpretations often contain poorly quantified biases [32]. For example, daily infection rates in Japan exhibit few similarities to those in the Italy. Training models on unrepresentative datasets will lead to poor (and even dangerous) outcomes. Whereas techniques such as transfer learning could allow models to be specialised with regional characteristics, the fast-moving nature of the problem can make it difficult to perform informed model selection and parameterisation. A key challenge is devising analytical approaches that can work with these data limitations.

B. Correctness of Results vs. Urgency

There is a clear need for rapid results, yet the methods surveyed in this paper are largely based on statistical learning using (rapidly produced) datasets. In a recent systematic review of prediction models for diagnosis and prognosis of COVID-19, Wynants *et al.* [217] report that all 31 reviewed prediction models have a high risk of bias (due to non-representative selection of control patients and model overfitting). The reported models are therefore error susceptible. This is an inherent risk in all scientific work but, given the fast-moving nature of the situation, errors can have severe consequences. It should further be remembered that the outcomes of research may impact healthcare policy. For example, predictions may be used by governments to decide the extent of social distancing. Yet political actors are often less well placed to understand the nuance of scientific studies. We therefore posit that a key challenge is balancing exigency vs. the need for well-evidenced and reproducible results that can inform policy.

Due to the above, another clear challenge is finding ways to capture and represent the uncertainty of conclusions produced within the flurry of research. Bayesian methods can be used to capture uncertainty, although we have seen limited quantification of uncertainty in studies so far [218]. To ensure the correctness of data analysis, researchers must also describe their goals and process, and facilitate reproducible conclusions, e.g., sharing code, data and documentation. This, again, can create challenges as such requirements are balanced against the need for urgency. Another potential avenue is 'Explainable AI' [219], which can be used to provide context to results. That said, it is not clear if this will protect against problems such as unintentional bias [220] or even adversarial scenarios [221].

C. Security, Privacy, and Ethics

Most of the works that we discussed imply the *sharing* and/or *use* of potentially sensitive data. Devising solutions that exhibit good results but also protect privacy and adhere to high ethical standards is a key challenge. We argue that this could be vital for encouraging uptake among populations, particularly as infrastructure setup may persist beyond the pandemic [222]. There are already substantial efforts to build privacy-preserving medical analytics. For example, MedCo [223] uses homomorphic encryption to allow sites to federate datasets with privacy guarantees. Drynx [224] supports privacy-conscious statistical analysis on distributed datasets. This links closely into the availability of data (see §VI-A), as often data can only be shared when robust privacy guarantees are in place.

Broadly speaking, there is some consensus as outlined in Floridi *et al.* [225] on the five main "AI ethics principles": (1) *beneficence*, (2) *non-maleficence*, (3) *autonomy*, (4) *justice*, and (5) *explicability*. However, in the situation imposed by COVID-19, decisions may need to choose a trade-off between these AI ethics virtues [226], [227]. For example, to what extent does the current situation warrant the prioritisation of "public health" and "beneficence" over "individual privacy" and "autonomy". And even if this is warranted in the short-term, how can we ensure that these compromises do not become permanent and it is possible to roll back these trade-offs in the future as the situation changes. Other difficult questions include the issue of allocation of scarce resources and the trade-offs involved therein. As highlighted in the Call for Action presented in March 2020 by a coalition of experts on data governance [228], there is also a need for data sharing between the public and private sectors to ensure that data is used for "beneficence". In effect, the failure to share data in such contexts may be considered maleficence since withholding critical data may block an opportunity to bring potential benefit. That said, good governance mechanisms with suitable regulations should be in place to oversee ethical use of data as much as possible.

Privacy may also become particularly challenging when considering the roll-out of interventions (e.g., targeted social distancing measures), as the intervention itself may expose sensitive information [229]. This, for example, may apply to contact tracing apps, which strive to notify users when they have been in contact with an infected person. Although privacy-preserving implementations exist (e.g., DP-3T, TraceSecure), notifications may still allow users to guess who the infected person is (see [230] for a discussion of security issues in tracing apps).

To move ahead, simple measures can be adopted to help ensure ethical data science research. For example, data collected should be transparent (the users should be informed about what data is being collected) and stewarded with a limited purpose (even when it is anonymised) and governed with ethical oversight and appropriate safeguards (e.g., with time limits and sunset provisions). Interested readers are referred to several comprehensive data ethics resources [225], [231]–[235], to a recent report from the TUM Institute for Ethics in Artificial Intelligence [226], and the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems [227] on the ethical challenges involved in using AI for managing the COVID-19 outbreak.

D. The Need for Multidisciplinary Collaborations

Our understanding of COVID-19's long-term impact is preliminary. Contributing serious insights will require a mix of domain expertise from multiple fields, and there is already a push for better international collaboration and tracking of COVID-19 [236]. For example, the use of black-box models might yield a superficially practical solution, but could be useless without the involvement of (international) medical and biotechnology expert interpretations. This will further have implications for licensing technologies and engendering uptake (as healthcare professionals are unlikely to engage with technologies developed without medical expertise). Rapidly bringing together cohorts of complementary expertise is therefore important. This also brings many further challenges, e.g., ensuring a team's interpretation of things like ethics, benefits and risks are coherent.

E. New Data Modalities

The data science community has limited exposure to certain modalities of data that may prove critical in combating COVID-19. A natural challenge is rapidly adapting existing techniques to reflect these new data types. For example, whereas the community has substantial expertise in computer vision tasks, there is less experience in processing ultrasound scans. Yet these have shown good results that are similar to chest CT scans and superior to standard chest radiography for the evaluation of pneumonia and/or acute respiratory distress syndrome (ARDS) in corona patients [237], [238]. They also have the benefit of greater ease of use, absence of radiation, and low cost. Despite these advantages, to the best of our knowledge, no study has yet explored the potential of automatically detecting COVID-19 infections via ultrasound scans. Similarly, magnetic resonance imaging (MRI) is considered the safest imaging modality as it is a non-invasive and non-ionising technique, which provides a high resolution image and excellent soft tissue contrast [121]. Some studies like [239] have described the significance of MRI in fighting against COVID-19 infections. Yet the modality remained under-explored by the computer vision community due to a lack of sufficient training data. Thus, a challenge is to rapidly develop a well-annotated dataset of such medical imaging modalities.

F. Solutions for the Developing World

The COVID-19 pandemic poses unique challenges to populations that have limited access to healthcare (e.g. in developing countries), particularly as such people are disproportionately affected by limited access to information [240]. A key challenge is developing technologies that are designed so that they are globally inclusive. This requires considering how such technologies will impact different communities, and exploring how they could be deployed in both rural and economically deprived regions [241]–[243], as well as how they might be misused in certain contexts. This subsumes several practical challenges that naturally vary based on the specific use case. For example, if building a mobile app for contact tracing, it should be low cost and require limited resources; it should be designed with limited network connectivity in-mind; it should also support multiple languages and be accessible to illiterate users or those with disabilities. We emphasise that ensuring wide accessibility of technological solutions is critical for addressing this *global* pandemic.

VII. Conclusion

Data scientists have been active in addressing the emerging challenges related to COVID-19. This paper has been written to make available a summary of ongoing work for the wider community. We have attempted to make five broad contributions. We first summarised publicly available datasets for use by researchers. This is intended as a community resource to shorten the time taken to discover relevant data. We then presented relevant use cases of data science, which have the potential to help in the pandemic. This is by no means a comprehensive list and we expect the set to expand in the coming months. Following this, we surveyed some of the ongoing research in this area. As the paper is mainly intended for a computer science and engineering audience, we themed our analysis around the different types of datasets available. Following this, we broadened our analysis and presented a bibliometric study of thousands of publications in recent months. Finally, we highlighted some of the common challenges we observed as part of our systematic review, e.g., availability of data and privacy concerns. We also note that many of the systems discussed in this paper are not operational yet. In view of this, we intend to keep updating our live resource repository with new information.

References

[1] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[2] J. T. Wu, K. Leung, and G. M. Leung, "Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in Wuhan, China: A modelling study," *Lancet*, vol. 395, no. 10225, pp. 689–697, 2020.

[3] N. Zhu *et al.*, "A novel coronavirus from patients with pneumonia in China, 2019," *New England J. Medicine*, vol. 382, pp. 727–733, 2020.

[4] K. Hao, *Over 24,000 Coronavirus Research Papers are now Available in one Place*, published on 16-March-2020; accessed on: 6-April-2020." [Online]. Available: https://tinyurl.com/MITTECHREV24000papers

[5] J. Bullock *et al.*, "Mapping the landscape of artificial intelligence applications against COVID-19," 2020, *arXiv:2003.11336*.

[6] M. van der Schaar *et al.*, "*How Artificial Intelligence Machine Learn. Can Help Healthcare Syst. Respond to COVID-19*, published on 27-Mar.-2020; accessed on: 1-Apr.-2020."

[7] SIRM, "Covid-19-BSTI Imaging Database," 2020. [Online]. Available: https://www.bsti.org.uk/training-and-education/covid-19-bsti-imaging-da% tabase/

[8] J. P. Cohen, P. Morrison, and L. Dao, "COVID-19 image data collection," 2020, *arXiv 2003.11597*. [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset

[9] MegSeg, "COVID-19 CT segmentation dataset," Mar. 2020. [Online]. Available: http://medicalsegmentation.com/covid19/

[10] "Artificial intelligence in radiology." [Online]. Available: https://mosmed.ai/datasets/covid19_1110, Accessed on: Sep. 9, 2020.

[11] G. Inc., "COVID-19 community mobility reports," Mar. 2020. [Online]. Available: https://www.google.com/covid19/mobility/

[12] SIRM, "COVID-19 DATABASE," 2020. [Online]. Available: https://www.sirm.org/category/senza-categoria/covid-19/

[13] Allen-Institute, "CORD-19 research challenge," Mar. 2020. [Online]. Available: https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challeng% e

[14] ECDC, "European Centre for Disease Prevention and Control (ECDC)," 2020. [Online]. Available: https://ourworldindata.org/coronavirus-source-data

[15] Smith, "Coronavirus (covid19) tweets," Mar. 2020. [Online]. Available: www.kaggle.com/smid80/coronavirus-covid19-tweets

[16] "COVID-19 Korea Dataset with Patient Routes," 2020. [Online]. Available: https://github.com/ThisIsIsaac/Data-Science-for-COVID-19

[17] CHIME, "COVID-19 Hospital Impact Model for Epidemics," 2020. [Online]. Available: https://github.com/CodeForPhilly/chime

[18] WHO, "Global research on novel coronavirus-2019," 2020. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-% research-on-novel-coronavirus-2019-ncov

[19] GISAID, "Genomic epidemiology of hCoV-19," 2020. [Online]. Available: https://www.gisaid.org/epiflu-applications/next-hcov-19-app/

[20] CSSEGISandData, "CSSEGISandData/COVID-19," Mar. 2020. [Online]. Available: https://github.com/CSSEGISandData/COVID-19

[21] Kinsa Health, "U.S. Health Weather Map," Mar. 2020. [Online]. Available: https://healthweather.us/?mode=Atypical

[22] NCBI, "LitCovid," 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/research/coronavirus/

[23] Q. Chen, A. Allot, and Z. Lu, "Keep up with the latest coronavirus research," *Nature*, vol. 579, no. 7798, p. 193, 2020.

[24] nCoV2019Data, "ncov2019 epidemiological data," Mar. 2020. [Online]. Available: https://github.com/beoutbreakprepared/nCoV2019

[25] sudalairajkumar Data, "Novel corona-virus dataset," Mar. 2020. [Online]. Available: https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset

[26] NewYork-Times, "New york times dataset," Mar. 2020. [Online]. Available: https://github.com/nytimes/covid-19-data

[27] E. Chen, K. Lerman, and E. Ferrara, "COVID-19: The first public Coronavirus Twitter dataset," 2020, *arXiv:2003.07372*.

[28] D. S. Goodsell *et al.*, "RCSB protein data bank: Enabling biomedical research and drug discovery," *Protein Science*, vol. 29, no. 1, pp. 52–65, 2020.

[29] NPGEO, "Dataset of infections in germany," 2020. [Online]. Available: https://npgeo-corona-npgeo-de.hub.arcgis.com/datasets/dd4580c810204019a% 7b8eb3e0b329dd6_0/data

[30] C. B. *et al.*, "Coronasurveys: Monitoring COVID-19 incidence via open polls," 2020. [Online]. Available: http://coronasurveys.com/

[31] A. Koubaa, "Understanding the covid19 outbreak: A comparative data analytics and study," 2020, *arXiv:2003.14150*.

[32] N. Fenton, G. A. Hitman, M. Neil, M. Osman, and S. McLachlan, "Causal explanations, error rates, and human judgment biases missing from the covid-19 narrative and statistics."

[33] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[34] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[35] C. Jacobs, "Coronada," 2020. [Online]. Available: https://github.com/BayesForDays/coronada

[36] E. Aramaki, S. Maskawa, and M. Morita, "Twitter catches the flu: Detecting influenza epidemics using Twitter," in *Proc. Conf. Empirical Methods Natural Lang. Process*. Association for Computational Linguistics, 2011, pp. 1568–1576.

[37] A. Culotta, "Towards detecting influenza epidemics by analyzing Twitter messages," in *Proc. 1st Workshop Social Media Analytics*, 2010, pp. 115–122.

[38] V. Lampos, T. De Bie, and N. Cristianini, "Flu detector-tracking epidemics on Twitter," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, 2010, pp. 599–602.

[39] K. Sharma, S. Seo, C. Meng, S. Rambhatla, A. Dua, and Y. Liu, "Coronavirus on social media: Analyzing misinformation in Twitter conversations," 2020, *arXiv:2003.12309*.

[40] K. Zarei, R. Farahbakhsh, N. Crespi, and G. Tyson, "A first instagram dataset on covid-19," 2020, *arXiv:2004.12226*.

[41] D. Zhao *et al.*, "A comparative study on the clinical features of COVID-19 pneumonia to other pneumonias," *Clin. Infect. Dis.*, vol. 71, no. 15, pp. 756–761, 2020.

[42] C. Manning, "Understanding human language: Can NLP and deep learning help?" in *Proc. 39th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2016, p. 1.

[43] "arXiv, accessed on: 12-April-2020." [Online]. Available: https://arxiv.org/

[44] "medRxiv, accessed on: 12-April-2020." [Online]. Available: https://www.medrxiv.org/

[45] "bioRXiv, accessed on: 12-April-2020." [Online]. Available: https://www.biorxiv.org/

[46] "Wikipedia database downloaod," 2020. [Online]. Available: https://en.wikipedia.org/wiki/Wikipedia:Database_download

[47] L. Wang and A. Wong, "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images," 2020, *arXiv:2003.09871*.

[48] D. Mery, *Computer Vision for X-Ray Testing*, vol. 10. Cham, Switzerland: Springer, 2015, pp. 978–3.

[49] "COVID Chest-Xray Dataset," Mar. 2020. [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset

[50] A. MVD, "Covid-19 x rays," Jun. 2020. [Online]. Available: https://www.kaggle.com/andrewmvd/convid19-X-rays

[51] J. Zhao, Y. Zhang, X. He, and P. Xie, "COVID-CT-Dataset: A CT scan dataset about COVID-19," 2020, *arXiv:2003.13865*.

[52] S. Morozov *et al.*, "Mosmeddata: Chest ct scans with covid-19 related findings," 2020, *arXiv:2005.06465*.

[53] Nhsx, "Covid-19 x rays," Jun. 2020. [Online]. Available: https://nhsx.github.io/covid-chest-imaging-database/

[54] E. A. Foundation, "Covid19 impact survey," Jun. 2020. [Online]. Available: https://covid19impactsurvey.org

[55] "Let's beat covid-19 - take the survey to help one million doctors." [Online]. Available: https://letsbeatcovid.net/, Accessed: Jul. 4, 2020.
[56] "Cov-clear | citizen science to help track and clear covid," 2020. [Online]. Available: https://cov-clear.com/, Accessed: Jul. 4, 2020.
[57] A. Zhavoronkov *et al.*, "Potential COVID-2019 3c-like protease inhibitors designed using generative deep learning approaches," *Insilico Medicine Hong Kong Ltd A*, vol. 307, p. E1, 2020.
[58] Z. Zhao, X. Li, F. Liu, G. Zhu, C. Ma, and L. Wang, "Prediction of the covid-19 spread in african countries and implications for prevention and controls: A case study in south africa, egypt, algeria, nigeria, senegal and kenya," *Sci. Total Environ.*, 2020, Art. no. 138959.
[59] Z. Usmani, "Pakistan coronavirus citywise data," Jun. 2020. [Online]. Available: https://www.kaggle.com/zusmani/pakistan-coronavirus-citywise-data
[60] R. data science coalition, "Uncover covid19 challenge," Mar 2020. [Online]. Available: https://www.kaggle.com/roche-data-science-coalition/uncover
[61] "Covid19 global forecasting challenge, the white house office of science and technology," Mar. 2020. [Online]. Available: https://www.kaggle.com/c/covid19-global-forecasting-week-2/overview
[62] M. Haghighatlari *et al.*, "Chemml: A machine learning and informatics program package for the analysis, mining, and modeling of chemical and materials data," *Wiley Interdisciplinary Reviews: Comput. Mol. Sci.*, 2019, Art. no. e1458.
[63] M. Cadotte, "Early evidence that COVID-19 government policies reduce urban air pollution," Mar. 2020. [Online]. Available: eartharxiv.org/nhgj3
[64] W. Inc., "World's air pollution: Real-time air quality index," 2020. [Online]. Available: https://waqi.info
[65] N. Oliver *et al.*, "Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle" *Sci. Adv.*, vol. 6, no. 23, 2020, Art. no. eabc0764, doi: 10.1126/sciadv.abc0764.
[66] "Data world," 2020. [Online]. Available: https://data.world/datasets/mobile
[67] S. Wurtzer, V. Marechal, J.-M. Mouchel, and L. Moulin, "Time course quantitative detection of sars-cov-2 in parisian wastewaters correlates with covid-19 confirmed cases," *medRxiv*, 2020.
[68] A. A. Khorana, "Artificial intelligence for cancer-associated thrombosis risk assessment–author's reply," *Lancet Haematology*, vol. 5, no. 9, pp. e391–e392, 2018.
[69] J. Dave, V. N. Dubey, D. Coppini, and J. Beavis, "Predicting diabetic neuropathy risk level using artificial neural network based on clinical characteristics of subjects with diabetes," *Diabetic Med.*, vol. 36, no. S1, p. 144, 2019.
[70] K. Wattanakit, G. Harshavardhan, S. Mungee, and M. Imtiaz, "Artificial intelligence based clinical risk assessment in predicting cardiac related chest pain in patients presenting to emergency room," *Circulation*, vol. 140, no. Suppl_1, pp. A14 100–A14 100, 2019.
[71] S. Latif *et al.*, "Mobile technologies for managing non-communicable diseases in developing countries," in *Mobile Applications and Solutions for Social Inclusion*. Pennsylvania, PA, USA: IGI Global, 2018, pp. 261–287.
[72] "*Startup Uses Fever Detection Technology To Stop Spread of Coronavirus*, accessed on: 4-April-2020," 2020. [Online]. Available: https://tinyurl.com/feverdetectiontechnology
[73] C. S. Currie *et al.*, "How simulation modelling can help reduce the impact of COVID-19," *J. Simul.*, pp. 1–15, 2020.
[74] W. O. Kermack, A. G. McKendrick, and G. T. Walker, "A contribution to the mathematical theory of epidemics," *Proc. Roy. Soc. London. Series A, Containing Papers Math. Phys. Character*, vol. 115, no. 772, pp. 700–721, 1927.
[75] A. J. Kucharski *et al.*, "Early dynamics of transmission and control of COVID-19: A mathematical modelling study," *Lancet Infect. Dis.*, vol. 20, no. 5, pp. 5537–558, 2020.
[76] K. J. Friston *et al.*, "Dynamic causal modelling of COVID-19," vol. 5, no. 89, 2020, Art. no. 89.
[77] J. Dehning *et al.*, "Inferring COVID-19 spreading rates and potential change points for case number forecasts," 2020, *arXiv:2004.01105*. [Online]. Available: http://dx.doi.org/10.1101/2020.04.02.20050922
[78] K. J. Friston *et al.*, "Second waves, social distancing, and the spread of Covid-19 across America," University College London, Tech. Rep., 2020. [Online]. Available: https://www.fil.ion.ucl.ac.uk/spm/covid-19
[79] Z. Tufekci, "*Don't Believe the COVID-19 Models: That's not What they're for*, The Atlantic, Data Published: Apr. 2, 2020," 2020. [Online]. Available: https://www.theatlantic.com/technology/archive/2020/04/coronavirus-mode% ls-arent-supposed-be-right/609271/
[80] M. Koerth, L. Bronner, and J. Mithani, "*Why It's So Freaking Hard To Make A Good COVID-19 Model*." [Online]. Available: https://fivethirtyeight.com/features/why-its-so-freaking-hard-to-make-a% -good-covid-19-model/, 2020.
[81] A. Noulas, C. Moffatt, D. Hristova, and B. Gonçalves, "Foursquare to the rescue: Predicting ambulance calls across geographies," in *Proc. Int. Conf. Digital Health*, 2018, pp. 100–109.
[82] N. Ferguson *et al.*, "Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand," 2020.
[83] H. Heesterbeek *et al.*, "Modeling infectious disease dynamics in the complex landscape of global health," *Science*, vol. 347, no. 6227, 2015, p. aaa4339, 2015.
[84] N. Ghaffarzadegan and H. Rahmandad, "Simulation-based estimation of the spread of covid-19 in Iran," *medRxiv*, 2020. [Online]. Available: https://www.medrxiv.org/content/early/2020/03/27/2020.03.22.20040956
[85] J. Hellewell *et al.*, "Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts," *Lancet Global Health*, vol. 8, no. 4, pp. e488–e496, 2020.
[86] L. Ferretti *et al.*, "Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing," *Sci.*, 2020, Art. no. 6491.
[87] A. S. S. Rao and J. A. Vazquez, "Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey in the populations when cities/towns are under quarantine," *Infection Control Hospital Epidemiology*, pp. 1–18, 2020.
[88] E. Yoneki and J. Crowcroft, "Epimap: Towards quantifying contact networks for understanding epidemiology in developing countries," *Ad Hoc Netw.*, vol. 13, pp. 83–93, 2014.
[89] "*AI Could Help With the Next Pandemicbut not With This one*, MIT Technology Review, accessed on: 1-April-2020." [Online]. Available: https://www.technologyreview.com/s/615351/ai-could-help-with-the-next-p% andemicbut-not-with-this-one/
[90] R. Sun, W. Wang, M. Xue, G. Tyson, S. Camtepe, and D. Ranasinghe, "Vetting security and privacy of global covid-19 contact tracing applications," 2020, *arXiv:2006.10933*.
[91] X. Li and D. Zhu, "Covid-xpert: An ai powered population screening of covid-19 cases using chest radiography images," 2020, *arXiv:2004.03042*.
[92] A. Wilder-Smith and D. Freedman, "Isolation, quarantine, social distancing and community containment: Pivotal role for old-style public health measures in the novel Coronavirus (2019-ncov) outbreak," *J. Travel Med.*, vol. 27, no. 2, 2020.
[93] F.-J. Schmitt, "A simplified model for expected development of the SARS-CoV-2 (corona) spread in germany and us after social distancing," 2020, *arXiv:2003.10891*.
[94] C. St Louis and G. Zorlu, "Can twitter predict disease outbreaks?" *Bmj*, vol. 344, 2012, Art. no. e2353.
[95] A. Signorini, A. M. Segre, and P. M. Polgreen, "The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic," *PloS One*, vol. 6, no. 5, 2011, Art. no. e19467.
[96] J. Zarocostas, "How to fight an infodemic," *Lancet*, vol. 395, no. 10225, p. 676, 2020.
[97] L. Bode and E. K. Vraga, "See something, say something: Correction of global health misinformation on social media," *Health Commun.*, vol. 33, no. 9, pp. 1131–1140, 2018.
[98] P. M. Waszak, W. Kasprzycka-Waszak, and A. Kubanek, "The spread of medical fake news in social media–the pilot quantitative study," *Health Policy Technol.*, vol. 7, no. 2, pp. 115–118, 2018.
[99] N. Parveen and J. Waterson, "Uk phone masts attacked amid 5g-coronavirus conspiracy theory," 2020. [Online]. Available: https://www.theguardian.com/uk-news/2020/apr/04/uk-phone-masts-attacked-amid-5g-coronavirus-conspiracy-theory
[100] "Misinformation related to the 2019 20 coronavirus pandemic," 2020. [Online]. Available: https://en.wikipedia.org/wiki/Misinformation_related_to_the_2019percentE2percent80percent9320_coronavirus_pandemic
[101] G. Pennycook, J. McPhetres, Y. Zhang, and D. Rand, "Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy nudge intervention," 2020.
[102] W. J. McKibbin and R. Fernando, "The global macroeconomic impacts of COVID-19: Seven scenarios," 2020.
[103] A. Atkeson, "What will be the economic impact of COVID-19 in the US? rough estimates of disease scenarios," National Bureau of Economic Research, Cambridge, MA, USA, 2020.

[104] "WHO and Rakuten Viber Fight COVID-19 Misinformation With Inter-1318 active Chatbot," 2020, Accessed on: Apr. 4, 2020. [Online]. Available: https://tinyurl.com/WHORatukenChatBot

[105] CPR News, "Health Care Workers' Stress Compounded By Long Days And Concerns About People Not Taking COVID-19 Seriously," 2020, Accessed on: Apr. 1, 2020. [Online]. Available: https://www.cpr.org/2020/03/23/colorado-coronavirus-stress-healthcare-w% orkers-covid-19-spread/

[106] "AliveCor, accessed on: 1-April-2020." [Online]. Available: https://www.alivecor.com/

[107] AliveCor, "*New FDA Guidance Allows Use of KardiaMobile 6L to Measure QTc in COVID-19 Patients*," 2020, Accessed on: Apr. 27, 2020. [Online]. Available: https://cutt.ly/pofvIbm

[108] "CLEW Medical, accessed on: 1-April-2020." [Online]. Available: https://clewmed.com/

[109] FreeStyle, "*Diabetes Manegmet and COVID-19*, accessed on: 27-April-2020." [Online]. Available: https://www.freestylelibre.co.uk/libre/freestyle-libre-blog/Managing-di% abetes-and-covid-19.html

[110] Kate Johnson, "*COVID-19: Home Pulse Oximetry Could Be Game Changer, Says ER Doc*," 2020. Accessed on: Apr. 27, 2020. [Online]. Available: https://www.medscape.com/viewarticle/929309?src=soc_tw_200424_mscpedt_n% ews_mdscp_pulseoximeter&faf=1

[111] HHS News, "*Secretary Azar Announces Historic Expansion of Telehealth Access to Combat COVID-19*, accessed on: 27-April-2020." [Online]. Available: https://cutt.ly/Gofbqqt

[112] Federal Communications Commission, "*COVID-19 Telehealth Program*," 2020, Accessed on: Apr. 27, 2020. [Online]. Available: https://www.fcc.gov/covid-19-telehealth-program/

[113] P. Webster, "Virtual health care in the era of covid-19," *Lancet*, vol. 395, no. 10231, pp. 1180–1181, 2020.

[114] J. B. Mitchell, "Artificial intelligence in pharmaceutical research and development," 2018.

[115] K.-K. Mak and M. R. Pichika, "Artificial intelligence in drug development: present status and future prospects," *Drug Discovery Today*, vol. 24, no. 3, pp. 773–780, 2019.

[116] A. Zhavoronkov, "Artificial intelligence for drug discovery, biomarker development, and generation of novel chemistry," 2018.

[117] G. Tyson, A. Taweel, S. Miles, M. Luck, T. Van Staa, and B. Delaney, "An agent-based approach to real-time patient identification for clinical trials," in *Proc. Int. Conf. Electron. Healthcare*, 2011, pp. 138–145.

[118] D. A. Berry, "Bayesian clinical trials," *Nature Reviews Drug Discovery*, vol. 5, no. 1, pp. 27–36, 2006.

[119] D. M. Gysi *et al.*, "Network medicine framework for identifying drug repurposing opportunities for covid-19," 2020, *arXiv:2004.07229*.

[120] S. Robertson, H. Azizpour, K. Smith, and J. Hartman, "Digital image analysis in breast pathology from image processing techniques to artificial intelligence," *Translational Res.*, vol. 194, pp. 19–35, 2018.

[121] M. Usman, S. Latif, M. Asim, B.-D. Lee, and J. Qadir, "Retrospective motion correction in multishot MRI using generative adversarial network," *Scientific Rep.*, vol. 10, no. 1, pp. 1–11, 2020.

[122] M. Usman, B.-D. Lee, S. S. Byon, S. H. Kim, and B. IlLee, "Volumetric lung nodule segmentation using adaptive roi with multi-view residual learning," vol. 10, no. 1, pp. 1–15, 2020.

[123] Z. Z. Qin *et al.*, "Using artificial intelligence to read chest radiographs for tuberculosis detection: A multi-site evaluation of the diagnostic accuracy of three deep learning systems," *Scientific Reports*, vol. 9, no. 1, pp. 1–10, 2019.

[124] R. Singh *et al.*, "Deep learning in chest radiography: Detection of findings and presence of change," *PloS One*, vol. 13, no. 10, 2018.

[125] F. Shi *et al.*, "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19," 2020, *EEE Rev. Biomed. Eng.*.

[126] T. Ai *et al.*, "Correlation of chest CT and RT-PCR testing in Coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases," *Radiology*, vol. 296, 2020, Art. no. 200642.

[127] Y. Li and L. Xia, "Coronavirus disease 2019 (covid-19): Role of chest ct in diagnosis and management," *Amer. J. Roentgenology*, vol. 214, no. 6, pp. 1280–1286, 2020.

[128] S. Wang *et al.*, "A deep learning algorithm using CT images to screen for corona virus disease (COVID-19)," *medRxiv*, 2020.

[129] X. Xu *et al.*, "Deep learning system to screen Coronavirus disease 2019 pneumonia," *Appl. Intell.*, p. 1, 2020.

[130] J. Chen *et al.*, "Deep learning-based model for detecting 2019 novel Coronavirus pneumonia on high-resolution computed tomography: a prospective study," *medRxiv*, 2020.

[131] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Berlin, Germany: Springer, 2018, pp. 3–11.

[132] O. Gozes *et al.*, "Rapid AI development cycle for the Coronavirus (COVID-19) pandemic: Initial results for automated detection & patient monitoring using deep learning CT image analysis," 2020, *arXiv:2003.05037*.

[133] L. Li *et al.*, "Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT," *Radiology*, vol. 214, 2020, Art. no. 200905.

[134] D. of Health., "The ionising radiation (medical exposure) regulations 2017," 2017.

[135] E. El-Din Hemdan, M. A. Shouman, and M. E. Karar, "COVIDX-Net: A framework of deep learning classifiers to diagnose COVID-19 in X-Ray images," 2020, *arXiv:2003.11055*.

[136] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[137] A. G. Howard *et al.*, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[138] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of Coronavirus disease (COVID-19) using x-ray images and deep convolutional neural networks," 2020, *arXiv:2003.10849*.

[139] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[140] M. Farooq and A. Hafeez, "COVID-ResNet: A deep learning framework for screening of COVID19 from radiographs," 2020, *arXiv:2003.14395*.

[141] P. K. Sethy and S. K. Behera, "Detection of coronavirus disease (covid-19) based on deep features," *Preprints*, 2020.

[142] L. Inc., "Lunit CXR for COVID-19; Chest X-ray solutions," 2020. [Online]. Available: https://insight.lunit.io/covid19/

[143] V. Inc., "VUNO Med-LungQuant & Chest X-ray solutions," 2020. [Online]. Available: https://covid19.vunomed.com/

[144] S. Kazemi Rashed, J. Frid, and S. Aits, "English dictionaries, gold and silver standard corpora for biomedical natural language processing related to SARS-CoV-2 and COVID-19," 2020, *arXiv:2003.09865*.

[145] C. E. Lopez, M. Vasu, and C. Gallemore, "Understanding the perception of COVID-19 policies by mining a multilanguage Twitter dataset," 2020, *arXiv:2003.10359*.

[146] J. E. C. Saire and R. C. Navarro, "What is the people posting about symptoms related to Coronavirus in Bogota, Colombia?" 2020, *arXiv:2003.11159*.

[147] L. Schild, C. Ling, J. Blackburn, G. Stringhini, Y. Zhang, and S. Zannettou, "'go eat a bat, chang!': An early look on the emergence of Sinophobic behavior on web communities in the face of COVID-19," 2020.

[148] M. Cinelli *et al.*, "The COVID-19 social media infodemic," 2020, *arXiv:2003.05004*.

[149] L. Singh *et al.*, "A first look at COVID-19 information and misinformation sharing on Twitter," 2020, *arXiv:2003.13907*.

[150] X. Li *et al.*, "Network bioinformatics analysis provides insight into drug repurposing for COVID-2019," 2020.

[151] M. M. Hossain, "Current status of global research on novel Coronavirus disease (COVID-19) : A bibliometric analysis and knowledge mapping," *Available at SSRN 3547824*, 2020.

[152] B. P. Roquette, H. Nagano, E. C. Marujo, and A. C. Maiorano, "Prediction of admission in pediatric emergency department with deep neural networks and triage textual data," *Neural Networks*, 2020.

[153] B. W. Schuller, D. M. Schuller, K. Qian, J. Liu, H. Zheng, and X. Li, "Covid-19 and computer audition: An overview on what speech & sound analysis could contribute in the SARS-CoV-2 Corona crisis," 2020, *arXiv:2003.11117*.

[154] I. Song, "Diagnosis of pneumonia from sounds collected using low cost cell phones," in *Proc. Int. Joint Conf. Neural Netw.*, 2015, pp. 1–8.

[155] R. Rana *et al.*, "Automated screening for distress: A perspective for the future," *Eur. J. Cancer Care*, vol. 28, no. 4, 2019, Art. no. e13033.

[156] S. Amiriparian *et al.*, "Cast a database: Rapid targeted large-scale big data acquisition via small-world modelling of social media platforms," in *Proc. 7th Int. Conf. Affect. Comput. Intell. Interact.*, 2017, pp. 340–345.

[157] P. Moradshahi, H. Chatrzarrin, and R. Goubran, "Improving the performance of cough sound discriminator in reverberant environments using microphone array," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. Proc.*, 2012, pp. 20–23.

[158] T. Olubanjo and M. Ghovanloo, "Tracheal activity recognition based on acoustic signals," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2014, pp. 1436–1439.
[159] A. Imran *et al.*, "AI4COVID-19: AI Enabled Preliminary Diagnosis for COVID-19 from Cough Samples via an App," *Inf. Med. Unlocked*, vol. 20, 2020, Art. no. 100378.
[160] S. Latif, J. Qadir, S. Farooq, and M. A. Imran, "How 5g wireless (and concomitant technologies) will revolutionize healthcare?" *Future Internet*, vol. 9, no. 4, pp. 93–117, 2017.
[161] Y. Ye, S. Hou, Y. Fan, Y. Qian, Y. Zhang, S. Sun, Q. Peng, and K. Laparo, "α-satellite: An AI-driven system and benchmark datasets for hierarchical community-level risk assessment to help combat COVID-19," 2020, *arXiv:2003.12232*.
[162] "Google uses location data to show which places are complying with stay-at-home orders and which aren't, *The Verge*," 2020, Accessed on: Apr. 14, 2020. [Online]. Available: https://www.theverge.com/2020/4/3/21206318/google-location-data-mobilit% y-reports-covid-19-privacy
[163] H. S. Maghdid, K. Z. Ghafoor, A. S. Sadiq, K. Curran, and K. Rabie, "A novel AI-enabled framework to diagnose Coronavirus COVID 19 using smartphone embedded sensors: Design study," in *Proc. IEEE 21st Int. Conf. Inf. Reuse Integr. Data Sci. (IRI)*, 2020, pp. 180–187.
[164] "Poland: App helps police monitor home quarantine," 2020. Accessed on: Apr. 1, 2020. [Online]. Available: https://privacyinternational.org/examples/3473/poland-app-helps-police-% monitor-home-quarantine/
[165] M. N. Kamel Boulos and E. M. Geraghty, "Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics," *Int. J. Health Geographics*, vol. 19, no. 8, 2020.
[166] J. Ye, "China's QR health code," published on 19-Feb. 19, 2020, Accessed on: Apr. 6,2020. [Online]. Available: https://www.abacusnews.com/tech/chinas-qr-health-code-system-brings-rel% ief-some-and-new-problems/article/3051020
[167] GovTech, "*TraceTogether - Behind the Scenes Look at its Development Process*, published on 25-March-2020; accessed on: 6-April-2020."
[168] T. Cohen, "Israelis Using Voluntary Coronavirus Monitoring App," 01-Apr. 1, 2020, Accessed on: Apr. 6, 2020. [Online]. Available: https://tinyurl.com/reutersCoronaIsraelApps
[169] H. Cho, D. Ippolito, and Y. W. Yu, "Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs," 2020, *arXiv:2003.11511*.
[170] R. A. Calvo, S. Deterding, and R. M. Ryan, "Health surveillance during COVID-19 pandemic," *BMJ*, vol. 369, 2020. [Online]. Available: https://www.bmj.com/content/369/bmj.m1373
[171] "Dp-3t documentation," 2020. [Online]. Available: https://github.com/DP-3T/documents
[172] J. Bell, D. Butler, C. Hicks, and J. Crowcroft, "Tracesecure: Towards privacy preserving contact tracing," 2020.
[173] A. Berke, M. Bakker, P. Vepakomma, R. Raskar, K. Larson, and A. Pentland, "Assessing disease exposure risk with location histories and protecting privacy: A cryptographic approach in response to a global pandemic," 2020, *arXiv:2003.14412*.
[174] J. John, T. Kathryn, K. Pushmeet, H. Demis, and A. Team, "Computational predictions of protein structures associated with COVID-19," Mar. 2020. [Online]. Available: https://cutt.ly/yofIYsB
[175] A. W. Senior *et al.*, "Protein structure prediction using multiple deep neural networks in the 13th critical assessment of protein structure prediction (CASP13)," *Proteins: Struct., Function, Bioinf.*, vol. 87, no. 12, pp. 1141–1148, 2019.
[176] N. Bung, S. R. Krishnan, G. Bulusu, and A. Roy, "De novo design of new chemical entities (NCEs) for SARS-CoV-2 using artificial intelligence," 2020. [Online]. Available: https://chemrxiv.org/articles/De_Novo_Design_of_New_Chemical_Entities_N% CEs_for_SARS-CoV-2_Using_Artificial_Intelligence/11998347/2
[177] H. C. Metsky, C. A. Freije, T.-S. F. Kosoko-Thoroddsen, P. C. Sabeti, and C. Myhrvold, "CRISPR-based surveillance for COVID-19 using genomically-comprehensive machine learning design," *bioRxiv*, 2020. [Online]. Available: https://www.biorxiv.org/content/early/2020/03/02/2020.02.26.967026.1
[178] F. Hu, J. Jiang, and P. Yin, "Prediction of potential commercially inhibitors against SARS-CoV-2 by multi-task deep model," 2020, *arXiv:2003.00728*.
[179] H. Zhang *et al.*, "Deep learning based drug screening for novel coronavirus 2019-nCov," 2020. [Online]. Available: https://www.preprints.org/manuscript/202002.0061/v1
[180] B. Tang, F. He, D. Liu, M. Fang, Z. Wu, and D. Xu, "AI-aided design of novel targeted covalent inhibitors against SARS-CoV-2," *bioRxiv*, 2020.
[181] V. Boucher, "Open and Collaborative De Novo Discovery of Antiviral Agents for COVID-19 With Deep Reinforcement Learning and OpenAI Gym," MONTREAL.AI, 2020, Accessed on: Apr. 7, 2020. [Online]. Available: https://montrealartificialintelligence.com/covid19/
[182] F. Shan *et al.*, "Lung infection quantification of COVID-19 in ct images with deep learning," 2020, *arXiv:2003.04655*.
[183] J. Zhang, Y. Xie, Y. Li, C. Shen, and Y. Xia, "Covid-19 screening on chest x-ray images using deep learning based anomaly detection," 2020, *arXiv:2003.12338*.
[184] S. Wang *et al.*, "A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis," *medRxiv*, 2020.
[185] M. Barstugan, U. Ozkaya, and S. Ozturk, "Coronavirus (covid-19) classification using ct images by machine learning methods," 2020, *arXiv:2003.09424*.
[186] O. Gozes, M. Frid-Adar, N. Sagie, H. Zhang, W. Ji, and H. Greenspan, "Coronavirus detection and analysis on chest ct with deep learning," 2020, *arXiv:2004.02640*.
[187] K. E. Asnaoui, Y. Chawki, and A. Idri, "Automated methods for detection and classification pneumonia based on x-ray images using deep learning," 2020, *arXiv:2003.14363*.
[188] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest x-ray images using DeTraC deep convolutional neural network," 2020, *arXiv:2003.13815*.
[189] S. Rajaraman and S. Antani, "Weakly labeled data augmentation for deep learning: A study on covid-19 detection in chest x-rays," *Diagnostics*, vol. 10, no. 6, p. 358, 2020.
[190] Y. Oh, S. Park, and J. C. Ye, "Deep learning covid-19 features on cxr using limited training data sets," *IEEE Trans. Medical Imag.*, vol. 39, no. 8, pp. 2688–2700, Aug. 2020.
[191] A. M. Alqudah, S. Qazan, H. Alquran, I. A. Qasmieh, and A. Alqudah, "COVID-2019 detection using x-ray images and artificial intelligence hybrid systems."
[192] B. Ghoshal and A. Tucker, "Estimating uncertainty and interpretability in deep learning for Coronavirus (COVID-19) detection," 2020, *arXiv:2003.10769*.
[193] F. M. Salman, S. S. Abu-Naser, E. Alajrami, B. S. Abu-Nasser, and B. A. Ashqar, "Covid-19 detection using artificial intelligence," 2020.
[194] Y. Zhu and S. Newsam, "Densenet for dense flow," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 790–794.
[195] M. Karim *et al.*, "Deepcovidexplainer: Explainable covid-19 predictions based on chest x-ray images," 2020, *arXiv:2004.04582*.
[196] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Phys. Eng. Sci. Med.*, vol. 43, pp. 635–640, 2020.
[197] K. Jahanbin and V. Rahmanian, "Using Twitter and web news mining to predict COVID-19 outbreak," 2020.
[198] Y. Zhao and H. Xu, "Chinese public attention to COVID-19 epidemic: Based on social media," *medRxiv*, 2020.
[199] L. Li, Q. Zhang, X. Wang, J. Zhang, T. Wang, T.-L. Gao, W. Duan, K. K.-f. Tsoi, and F.-Y. Wang, "Characterizing the propagation of situational information in social media during COVID-19 epidemic: A case study on weibo," *IEEE Trans. Comput. Social Syst.*, vol. 7, no. 2, pp. 556–562, Apr. 2020.
[200] D. Prabhakar Kaila *et al.*, "Informational flow on Twitter–Corona virus outbreak–topic modelling approach," *Int. J. Adv. Res. Eng. Technol.*, vol. 11, no. 3, 2020.
[201] F. Stephany, N. Stoehr, P. Darius, L. Neuhäuser, O. Teutloff, and F. Braesemann, "The CoRisk-index: A data-mining approach to identify industry-specific risk assessments related to COVID-19 in real-time," 2020, *arXiv:2003.12432*.
[202] M. Hofmarcher *et al.*, "Large-scale ligand-based virtual screening for SARS-CoV-2 inhibitors using deep neural networks," 2020, *arXiv:2004.00979*.
[203] A. Mayr *et al.*, "Large-scale comparison of machine learning methods for drug target prediction on ChEMBL," *Chem. Sci.*, vol. 9, no. 24, pp. 5441–5451, 2018.
[204] K. Preuer, G. Klambauer, F. Rippmann, S. Hochreiter, and T. Unterthiner, "Interpretable deep learning in drug discovery," in *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning.* Berlin, Germany: Springer, 2019, pp. 331–345.
[205] T. Sterling and J. J. Irwin, "ZINC 15–ligand discovery for everyone," *J. Chem. Inf. Model.*, vol. 55, no. 11, pp. 2324–2337, 2015.

[206] B. R. Beck, B. Shin, Y. Choi, S. Park, and K. Kang, "Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model," *Comput. Structural Biotechnology J.*, vol. 18, pp. 784–790, 2020.
[207] J. Kim *et al.*, "Advanced bioinformatics rapidly identifies existing therapeutics for patients with coronavirus disease-2019 (COVID-19)," *J. Translational Medicine*, vol. 18, 2020, Art. no. 257.
[208] P. Richardson *et al.*, "Baricitinib as potential treatment for 2019-ncov acute respiratory disease," *Lancet*, vol. 395, no. 10223, pp. e30–e31, 2020.
[209] J. Stebbing *et al.*, "Covid-19: Combining antiviral and anti-inflammatory treatments," *Lancet Infectious Diseases*, vol. 20, pp. 400–402, 2020.
[210] V. Chenthamarakshan *et al.*, "Target-specific and selective drug design for covid-19 using deep generative models," 2020, *arXiv:2004.01215*.
[211] "Global coalition to accelerate COVID-19 clinical research in resource-limited settings," *The Lancet*, 2020. [Online]. Available: https://doi.org/10.1016/S0140-6736(20)30798-4
[212] S. Mahmoud *et al.*, "Multi-agent system for recruiting patients for clinical trials," Kings College London, London, U.K., 2014.
[213] J. A. Anguera, J. T. Jordan, D. Castaneda, A. Gazzaley, and P. A. Areán, "Conducting a fully mobile and randomised clinical trial for depression: Access, engagement and expense," *BMJ Innov.*, vol. 2, no. 1, pp. 14–21, 2016.
[214] "Scopus, accessed on: 12-April-2020," [Online]. Available: https://www.scopus.com/home.uri
[215] CDC, "CDC research Repositories, accessed on: 12-April-2020." [Online]. Available: https://www.cdc.gov/library/researchguides/2019novelcoronavirus/databas% esjournals.html
[216] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, no. Jan, pp. 993–1022, 2003.
[217] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of COVID-19 infection: Systematic review and critical appraisal," *BMJ*, vol. 369, 2020.
[218] N. Fenton and M. Neil, "The use of Bayes and causal modelling in decision making, uncertainty and risk," *Council Eur. Professional Inform. Societies Upgrade*, vol. 12, no. 5, pp. 10–21, 2011.
[219] D. Gunning, "Explainable artificial intelligence (XAI)," *Defense Adv. Res. Projects Agency, nd Web*, vol. 2, 2017.
[220] S. Latif, A. Qayyum, M. Usama, J. Qadir, A. Zwitter, and M. Shahzad, "Caveat emptor: The risks of using big data for human development," *IEEE Technol. Soc. Mag.*, vol. 38, no. 3, pp. 82–90, 2019.
[221] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, "Secure and robust machine learning for healthcare: A survey," *IEEE Rev. Biomed. Eng.*, 2020.
[222] Y. N. Harari, "The world after coronavirus, financial times," 2020. Accessed on: Apr. 1, 2020. [Online]. Available: https://www.ft.com/content/19d90308-6858-11ea-a3c9-1fe6fedcca75
[223] J. L. Raisaro *et al.*, "Medco: Enabling secure and privacy-preserving exploration of distributed clinical and genomic data," *IEEE/ACM Trans. Comput. Biology Bioinf.*, vol. 16, no. 4, pp. 1328–1341, Jul./Aug. 2018.
[224] D. Froelicher, J. R. Troncoso-Pastoriza, J. S. Sousa, and J.-P. Hubaux, "Drynx: Decentralized, secure, verifiable system for statistical queries and machine learning on distributed datasets," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 3035–3050, Mar. 2020.
[225] L. Floridi *et al.*, "AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations," *Minds Mach.*, vol. 28, no. 4, pp. 689–707, 2018.
[226] "Ethical Implications of the Use of AI to Manage the COVID-19 Outbreak," Apr. 2020. [Online]. Available: https://ieai.mcts.tum.de/wp-content/uploads/2020/04/April-2020-IEAI-Res% earch-Brief-Covid-19-FINAL.pdf
[227] "*Statement Regarding the Ethical Implementation of Artificial Intelligence Systems (AIS) for Addressing the COVID-19 Pandemic*, The Executive Committee of The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems," Apr. 2020. [Online]. Available: https://standards.ieee.org/content/dam/ieee-standards/standards/web/doc% uments/other/gieais-covid.pdf
[228] "*Call for Action: Toward Building the Data Infrastructure and Ecosystem We Need to Tackle Pandemics and Other Dynamic Society and Environmental Threats*. The Gov Lab, New York University," Mar. 2020. [Online]. Available: http://www.thegovlab.org/static/files/publications/ACallForActionCOVID1% 9.pdf
[229] "Pandemic data challenges," *Nature Mach. Intell.*, vol. 2, no. 4, pp. 193–193, Apr. 2020. [Online]. Available: https://doi.org/10.1038/s42256-020-0172-7
[230] R. Anderson, "Contact tracing in the real world," 2020. [Online]. Available: https://www.lightbluetouchpaper.org/2020/04/12/contact-tracing-in-the-r% eal-world/
[231] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," *Nature Mach. Intell.*, vol. 1, no. 9, pp. 389–399, 2019.
[232] T. Hagendorff, "The ethics of AI ethics: An evaluation of guidelines," *Minds Mach.*, vol. 30, pp. 99–120, 2020.
[233] N. Bostrom and E. Yudkowsky, "The ethics of artificial intelligence," *Cambridge Handbook Artif. Intell.*, vol. 1, pp. 316–334, 2014.
[234] C. Tucker, A. Agrawal, J. Gans, and A. Goldfarb, "Privacy, algorithms, and artificial intelligence," *in The Economics of Artificial Intelligence: An Agenda*. Chicago, IL, USA: Univ. Chicago Press, 2018, pp. 423–437,.
[235] R. Shokri and V. Shmatikov, "Privacy-preserving deep learning," in *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur.*, 2015, pp. 1310–1321.
[236] E. Segal *et al.*, "Building an international consortium for tracking coronavirus health status," *Nat Med.*, vol. 26, no. 8, pp. 1161–1165, 2020.
[237] Q.-Y. Peng *et al.*, "Findings of lung ultrasonography of novel corona virus pneumonia during the 2019–2020 epidemic," *Intensive Care Medicine*, vol. 46, no. 5, pp. 849–850, 2020.
[238] E. Poggiali *et al.*, "Can lung us help critical care clinicians in the early diagnosis of novel coronavirus (COVID-19) pneumonia?" *Radiology*, vol. 295, no. 3, pp. E6–E6, 2020.
[239] N. Poyiadji, G. Shahin, D. Noujaim, M. Stone, S. Patel, and B. Griffith, "COVID-19–associated acute hemorrhagic necrotizing encephalopathy: CT and MRI features," *Radiology*, vol. 296, 2020, Art. no. 201187.
[240] F. Ahmed, N. Ahmed, C. Pissarides, and J. Stiglitz, "Why inequality could spread COVID-19," *Lancet Public Health*, Apr. 2020. [Online]. Available: https://doi.org/10.1016/s2468-2667(20)30085-2
[241] J. Qadir *et al.*, "IEEE access special section editorial: Health informatics for the developing world," *IEEE Access*, vol. 5, pp. 27 818–27 823, 2017.
[242] S. Latif, R. Rana, J. Qadir, A. Ali, M. A. Imran, and M. S. Younis, "Mobile health in the developing world: Review of literature and lessons from a case study," *IEEE Access*, vol. 5, pp. 11 540–11 556, 2017.
[243] J. Quinn, V. Frias-Martinez, and L. Subramanian, "Computational sustainability and artificial intelligence in the developing world," *AI Mag.*, vol. 35, no. 3, p. 36, 2014.

Siddique Latif received the bachelor's degree in electronic engineering from International Islamic University, Islamabad, Pakistan, in 2014, and the M.S. degree in electrical engineering from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2018, all in electrical engineering. He is currently working toward the Ph.D. degree with the University of Southern Queensland (USQ), Toowoomba, QLD, Australia and the Distributed Sensing Systems Research Group, Data61—CSIRO. Mr. Latif's bachelor's degree was sponsored by the National ICT Scholarship Program.

Muhammad Usman received the B.S. degree from International Islamic University, Islamabad, Pakistan, in 2014, and the M.S. degree from the COMSATS University Islamabad, Islamabad, Pakistan, in 2018, both in electrical engineering. He is currently working toward the Ph.D. degree with Seoul National University, Seoul, South Korea and AI Research Engineer, HealthHub Company, South Korea. He also worked as a Research Associate with IHSAN Lab, Information Technology University, Lahore, Pakistan, under the supervision of Prof. Junaid Qadir. His current research interests span medical image processing, medical image analysis, computer vision, and deep learning. Mr. Usman's B.S. degree was sponsored under the National ICT R&D Scholarship Program.

Sanaullah Manzoor (Member, IEEE) received the M.S. degree in computer system engineering from Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Swabi, Pakistan. He is currently working toward the Ph.D. degree with the Department of Computer Science, Information Technology University, Lahore, Pakistan. His main research areas include applied machine learning, 5G cellular networks, user mobility modeling, proactive resource management, and big data analytics.

Waleed Iqbal received the B.Sc. degree in electrical engineering from the University of the Punjab, Lahore, Pakistan, in 2013, and the M.S. degree in computer science from Information Technology University, Lahore, in 2018. He is currently working toward the Ph.D. degree with Social Data Science Lab (SDS), School of Electronic Engineering and Computer Science, Queen Mary University of London (QMUL), London, U.K. His current research interests span Internet measurements, social network analysis, scientometrics, and big data analytics.

Junaid Qadir (Senior Member, IEEE) is currently the Director of the IHSAN Lab and the Chairperson of the Electrical Engineering Department at the Information Technology University (ITU) of Punjab, Lahore, Pakistan. His primary research interests are in the areas of computer systems and networking, applied machine learning, using ICT for development (ICT4D), and engineering education. He has authored or coauthored more than 100 peer-reviewed articles at various high-quality research venues including more than 50 impact-factor journal publications at top international research journals, including the *IEEE Communication Magazine*, IEEE Journal on Selected Areas in Communication (JSAC), IEEE Communications Surveys and Tutorials (CST), and IEEE Transactions on Mobile Computing (TMC). He has been appointed as ACM Distinguished Speaker for a three-year term starting from 2020. He is a Senior Member of ACM.

Gareth Tyson is a Senior Lecturer with the Queen Mary University of London and a Fellow with the Alan Turing Institute, London, U.K. He is the Deputy Director of the Institute of Applied Data Science (IADS), and co-leads the Social Data Science Lab (SDS). His research has been awarded the Best Student Paper Award at the Web Conference 2020; the Best Paper Award at eCrime'19; and the Honourable Mention Award at the Web Conference 2018 (best paper in track). He has served on numerous organising and programme committees, such as ACM IMC, ACM CoNEXT, and ICWSM.

Ignacio Castro is currently a Lecturer in Data Analytics with the Queen Mary University of London, London, U.K. and a member of the Social Data Science Lab (SDS). His research interests focus on the interactions of networks and economics. His research has been awarded the Best Student Paper Award at the Web Conference 2020 and serves on numerous organising and programme committees, such as ACM SIGCOMM, ACM IMC, and ACM CoNEXT.

Adeel Razi (Member, IEEE) is currently an Associate Professor with the Turner Institute for Brain and Mental Health, Monash University, Melbourne, VIC, Australia, where he is the Director of Computational Neuroscience Laboratory and the Deputy Lead of the Brain Mapping and Modeling Research Program. He is currently Australian Research Council DECRA Fellow (2017 to 2020) and has also been awarded NHMRC Investigator (Emerging Leader) Fellowship (2021 to 2025). He is an Honorary Senior Research Fellow with the Wellcome Centre for Human Neuroimaging, University College London, where he also worked from 2012 to 2018. His research interest is cross-disciplinary – combining engineering, physics, and machine-learning approaches – motivated by questions grounded in neuroscience leading toward the understanding of how the brain works.

Maged N. Kamel Boulos (Senior Member, IEEE) received the master's degree in medical informatics from King's College, University of London, London, U.K., in 2000, and the Ph.D. degree in measurement and information in medicine from the City University of London, London, U.K., in 2002. He has more than 30 years of clinical and informatics experience, working as Medical Doctor (Dermatologist), Researcher, then Lecturer, then Associate Professor, then Professor of Health Informatics (Chair of Digital Health) with City, University of London, and the Universities of Bath, Plymouth, and UHI, Scotland, U.K., before moving to Guangzhou, China, where he is currently a Professor with Sun Yat-sen University. He has authored or coauthored more than 160 publications, with a GS h-index of 44. He is the founder and Editor-in-Chief of Springer Nature's *International Journal of Health Geographics*.

Adrian Weller is currently a Programme Director for AI at The Alan Turing Institute, the UK national institute for data science and AI, where he is also a Turing Fellow leading work on safe and ethical AI. He is a Principal Research Fellow in Machine Learning with the University of Cambridge, and with the Leverhulme Centre for the Future of Intelligence where he is Programme Director for Trust and Society. His interests span AI, its commercial applications and helping to ensure beneficial outcomes for society. He serves on several boards, including the Centre for Data Ethics and Innovation. He is Co-Director of the European Laboratory for Learning and Intelligent Systems (ELLIS) programme on Human-centric Machine Learning, and a member of the UNESCO Ad Hoc Expert Group on the Ethics of AI. Previously, he held senior roles in finance.

Jon Crowcroft (Fellow, IEEE) received the graduate degree in physics from Trinity College, University of Cambridge, Cambridge, U.K., in 1979, and the M.Sc. degree in computing and the Ph.D. degree from UCL, London, U.K., in 1981 and 1993, respectively. He has been the Marconi Professor of Communications Systems with the Computer Laboratory since October 2001. He has worked in the area of internet support for multimedia communications for more than 30 years. Three current research interest interest have been scalable multicast routing, practical approaches to traffic management, and the design of deployable end-to-end protocols. Current active research areas are opportunistic communications, social networks, privacy preserving analytics, and techniques and algorithms to scale infrastructure-free mobile systems. He leans toward a "build and learn" paradigm for research. From 2016 to 2018, he was the Programm Chair at the Turing, the U.K.'s national Data Science and AI Institute, and is now researcher-at-large there. He is a fellow of the Royal Society, ACM, British Computer Society, IET, and the Royal Academy of Engineering.