A Survey on Data-driven COVID-19 and Future Pandemic Management

YUDONG TAO, University of Miami
CHUANG YANG, The University of Tokyo
TIANYI WANG and ERIK COLTEY, Florida International University
YANXIU JIN, YINGHAO LIU, RENHE JIANG, ZIPEI FAN, XUAN SONG, and RYOSUKE SHIBASAKI, The University of Tokyo
SHU-CHING CHEN, Florida International University
MEI-LING SHYU, University of Miami
STEVEN LUIS, Florida International University

The COVID-19 pandemic has resulted in more than 440 million confirmed cases globally and almost 6 million reported deaths as of March 2022. Consequently, the world experienced grave repercussions to citizens' lives, health, wellness, and the economy. In responding to such a disastrous global event, countermeasures are often implemented to slow down and limit the virus's rapid spread. Meanwhile, disaster recovery, mitigation, and preparation measures have been taken to manage the impacts and losses of the ongoing and future pandemics. Data-driven techniques have been successfully applied to many domains and critical applications in recent years. Due to the highly interdisciplinary nature of pandemic management, researchers have proposed and developed data-driven techniques across various domains. However, a systematic and comprehensive survey of data-driven techniques for pandemic management is still missing. In this article, we review existing data analysis and visualization techniques and their applications for COVID-19 and future pandemic management with respect to four phases (namely, Response, Recovery, Mitigation, and Preparation) in disaster management. Data sources utilized in these studies and specific data acquisition and integration techniques for COVID-19 are also summarized. Furthermore, open issues and future directions for data-driven pandemic management are discussed.

CCS Concepts: • **General and reference → Surveys and overviews**; • **Applied computing → Health informatics**; • **Computing methodologies → Artificial intelligence**; **Modeling and simulation;**

Additional Key Words and Phrases: Data analytics, data visualization, pandemic management, COVID-19

This work is partially supported by National Science Foundation (NSF), Grant Numbers CNS-1952089 and CNS-2125165, and Strategic International Collaborative Research Program (SICORP) of Japan Science and Technology Agency (JST), Grant Number JPMJSC2002.

Authors' addresses: Y. Tao and M.-L. Shyu, Department of Electrical and Computer Engineering, University of Miami, 1251 Memorial Drive, Coral Gables, FL, 33146; emails: {yxt128, shyu}@miami.edu; C. Yang, Y. Jin, Y. Liu, R. Jiang, Z. Fan, X. Song, and R. Shibasaki, Center for Spatial Information Science, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan; emails: {chuang.yang, jinyanxiu, hero.liu, jiangrh, fanzipei0725, songxuan, shiba}@csis.u-tokyo.ac.jp; T. Wang, E. Coltey, and S. Luis, Knight Foundation School of Computing & Information Sciences, Florida International University, 11200 SW 8th Street, Miami, FL, 33199; emails: {wtian002, ecolt003, luiss}@cs.fiu.edu; S.-C. Chen, Data Science and Analytics Innovation Center, University of Missouri-Kansas City, 5110 Rockhill Road, Kansas City, MO, 64110; email: s.chen@umkc.edu.

0360-0300/2022/12-ART141 $15.00
https://doi.org/10.1145/3542818

ACM Reference format:
Yudong Tao, Chuang Yang, Tianyi Wang, Erik Coltey, Yanxiu Jin, Yinghao Liu, Renhe Jiang, Zipei Fan, Xuan Song, Ryosuke Shibasaki, Shu-Ching Chen, Mei-Ling Shyu, and Steven Luis. 2022. A Survey on Data-driven COVID-19 and Future Pandemic Management. *ACM Comput. Surv.* 55, 7, Article 141 (December 2022), 36 pages.
https://doi.org/10.1145/3542818

1 INTRODUCTION

Pandemics, the epidemics of infectious diseases, are considered a type of disaster, i.e., anthropogenic hazards caused by human action or inaction, leading to unanticipated outcomes and losses. Compared to natural hazards such as hurricane/typhoon, flood, and earthquake, pandemics could affect many more regions and people. For example, the "Spanish" influenza pandemic in 1918 was estimated to cause the deaths of at least 50 million people all over the world [1]. **Severe Acute Respiratory Syndrome (SARS)** in 2002 led to less than 1,000 deaths but around USD 30–100 billion in economic losses [2]. Although human society has become more capable in managing pandemics than a hundred years ago, there are still many gaps in capability, some of which demand the investigation of novel technologies [3]. Unfortunately, we were hit by the **Coronavirus disease 2019 (COVID-19)** global pandemic in late 2019, before we were able to be more prepared.

The Pathogen of COVID-19 is a variant of coronavirus, i.e., SARS-CoV-2. It is highly contagious and very difficult for public officials to curb [4]. Thus, the **World Health Organization (WHO)** announced that COVID-19 had caused a global pandemic [5]. Since the outbreak, COVID-19 has resulted in grave repercussions globally to people's lives, health, and the economy. As of June 2021, there are more than 180 million confirmed cases and almost 4 million reported deaths caused by COVID-19 worldwide [6]. The elderly and people with underlying medical conditions such as heart difficulties, diabetes, and hypertension are the most vulnerable due to their increased likelihood of developing severe and deadly illnesses when being infected. Meanwhile, the world economy had suffered the most severe recession since the Great Recession in 2008, and the International Monetary Fund has estimated the 2020 global GDP growth contraction at −3.5%, i.e., trillions of dollars of reduction in GDP [7].

In response to a global pandemic with such an unanticipated large amount of cases, countermeasures are necessary to mitigate the effect of COVID-19, manage its contagions, and begin recovery. Social distancing policies such as quarantine and lockdowns have been deployed to reduce the speed of virus spread [8]. These policies focus on restricting human-to-human contacts, protecting vulnerable communities, and allowing the development of therapeutics and vaccines simultaneously [9]. While these countermeasures are effective in slowing down the spread, they could result in adverse impacts on people's livelihood, mental health, and the economy. Some of these unfavorable consequences may include the permanent closure of small businesses, an increased rate of unemployment, and loss of income to maintain basic living expenses [10]. The complex consequences of social distancing policies make it difficult to make optimal decisions for pandemic response, management, and recovery.

Meanwhile, 200 vaccines have been in various stages of development from teams around the world. Among all candidates, 50 vaccines have conducted trials on humans, and there are 18 of them under Phase III trials to evaluate their efficacy. Many vaccines have achieved satisfactory efficacy rates [11] and are now being used in practice. However, vaccination requires more than research and development. To achieve immunity for the global population, 10 to 11 billion high-quality and safe doses need to be manufactured to effectively interrupt the transmission of COVID-19. However, current manufacturing capacity is estimated to be 2 to 4 billion doses per year [12].

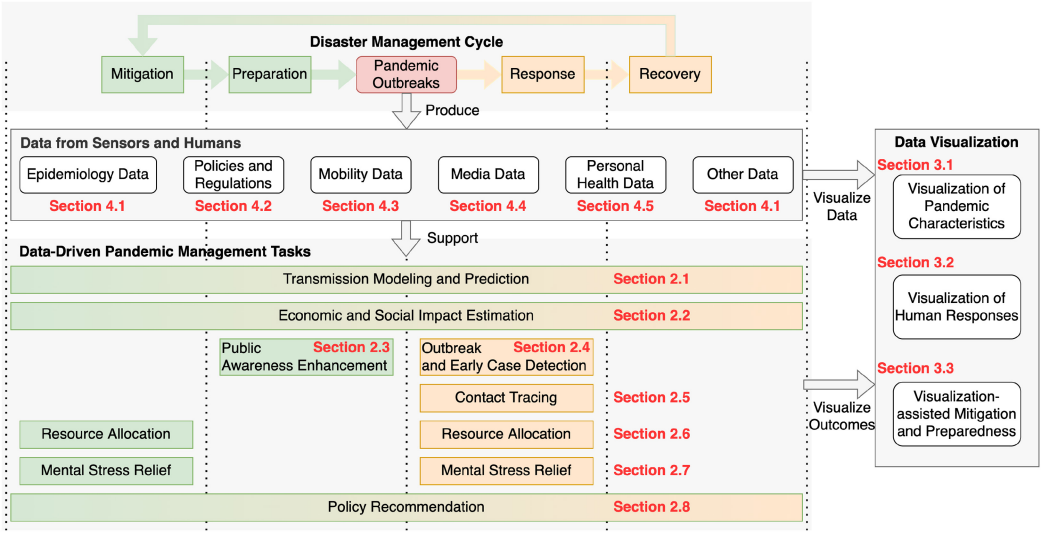

Fig. 1. Disaster management cycle and data-driven techniques for COVID-19 and future pandemics.

This limitation raises more issues, such as vaccination access and equity, dissemination, and so on [11]. To resolve these problems and provide practical solutions, further research is necessary.

The disaster management cycle [13] can be leveraged to separate pandemic management into four phases: *Response*, *Recovery*, *Mitigation*, and *Preparation*, as shown at the top of Figure 1. In each phase, a series of decisions are made to reduce pandemic consequences based on available information.

Right after the occurrence of infected cases, it is necessary to begin the *Response* phase by understanding the immediate risks and threats to people from the pandemic. As the *Response* phase progresses, the immediate and acute issues are resolved. Thereafter, in the *Recovery* phase, the focus moves to counteracting the negative impact of the pandemic and bringing society back to normal. Strategic plans to address the severe impacts of the pandemic should also be developed in this phase. In the *Mitigation* phase, it is necessary to take action to protect people and their property. Steps should also be taken to reduce vulnerability to future pandemics as well. The *Preparation* phase requires that the population be educated and trained based on a firm understanding about the disaster and its impacts. Global understanding is necessary to be prepared for future pandemics. A global pandemic such as COVID-19 could create multiple outbreaks in various areas and at different time periods after discovery (e.g., the influenza pandemic becomes seasonal after it is controlled for decades). Pandemic management requires continuous effort to respond to ongoing outbreaks, recover from previous outbreaks, and mitigate and prepare for the future.

Data science and data-driven techniques such as machine learning, **Reinforcement Learning (RL), Deep Neural Network (DNN)**, and **Natural Language Processing (NLP)** have been successfully applied to many domains in recent years [14]. These techniques provide powerful tools that process, analyze, synthesize, and perceive various types of data. As shown in Figure 1, data-driven techniques can be applied to support decision-making for critical pandemic management tasks throughout the disaster management cycle, supported by data generated during pandemics and disaster management processes. Meanwhile, data visualization techniques can help visualize both the data and outcomes of pandemic management tasks, as well as assist in *Mitigation* and *Preparation*. Useful information for battling COVID-19 [15], along with large volumes

Table 1. Summary of COVID-19 and Pandemic Management Tasks Using Data-driven Techniques

Section	Pandemic Management Tasks	Related Phases	Main Related Datasets
Section 2.1	Transmission Modeling and Prediction	All the phases	Epidemiology & Mobility
Section 2.2	Economic and Social Impact Estimation	All the phases	Media, Economic & Demographic
Section 2.3	Public Awareness Enhancement	Preparation	Epidemiology & Media
Section 2.4	Outbreak and Early Case Detection	Response	Media & Personal Health
Section 2.5	Contact Tracing	Response	Mobility
Section 2.6	Resource Allocation	Mitigation & Response	Epidemiology, Mobility & Demographic
Section 2.7	Mental Stress Relief	Mitigation & Response	Large-Scale Text Corpus
Section 2.8	Policy Recommendation	All the phases	Epidemiology, Policy & Mobility

of data collected by sensors and produced by users [8, 16, 17] are now available. The availability of data and data-driven techniques has triggered plenty of research efforts towards developing data-driven methods to mitigate COVID-19 transmission, facilitate pandemic management, and support decision-making processes.

This article summarizes and emphasizes the contributions of data science and data-driven techniques for modern pandemic management. We aim to help motivate further research to bridge the gaps in responding to and preparing for ongoing and future pandemics. Existing data-driven techniques for pandemic management have been surveyed, especially those using recent advances in big data, machine learning, deep learning, and artificial intelligence. In total, 152 and 22 papers acquired from Web of Science and arXiv/medRxiv databases, respectively, have been included and discussed in this survey (some of them can also be retrieved from Scopus, Elsevier, IEEE, and ACM databases as well). These papers were selected based on their quality and relevance to both pandemic management and data-driven techniques. Different from the existing surveys regarding COVID-19 management [15, 18–21], this article summarizes the data-driven techniques for COVID-19 management from a unique perspective, i.e., the disaster management cycle (See Figure 1). A broad range of applications in pandemic management is introduced and discussed in this article. Moreover, instead of solely discussing the data-driven techniques or the pandemic management, the data-driven techniques and the related data sources are presented with their connections to the management process.

The rest of this article is organized as follows: Section 2 discusses existing data analysis techniques for COVID-19 and pandemic management, organized by their applications. Data visualization tools and methods for COVID-19 and pandemic management are presented in Section 3. Thereafter, existing datasets for COVID-19 and pandemic management and continuous efforts to collect data and monitor the status of COVID-19 are summarized in Section 4. Section 5 introduces several open issues that emerge in COVID-19 management, as well as a few identified future research directions for data-driven pandemic management. Finally, Section 6 summarizes the current data-driven COVID-19 and future pandemic management.

2 DATA ANALYTICS FOR PANDEMIC MANAGEMENT

In this section, we discuss data analytics for the *Response*, *Recovery*, *Mitigation*, and *Preparation* of COVID-19 and future pandemics. We review several critical pandemic management applications and tasks, including transmission modeling and prediction, economic and social impact estimation, policy suggestion and recommendation, public awareness enhancement, and others. Since data-driven techniques for these applications have been continuously developed and are frequently published, we mainly review the most relevant high-impact papers. Table 1 summarizes existing

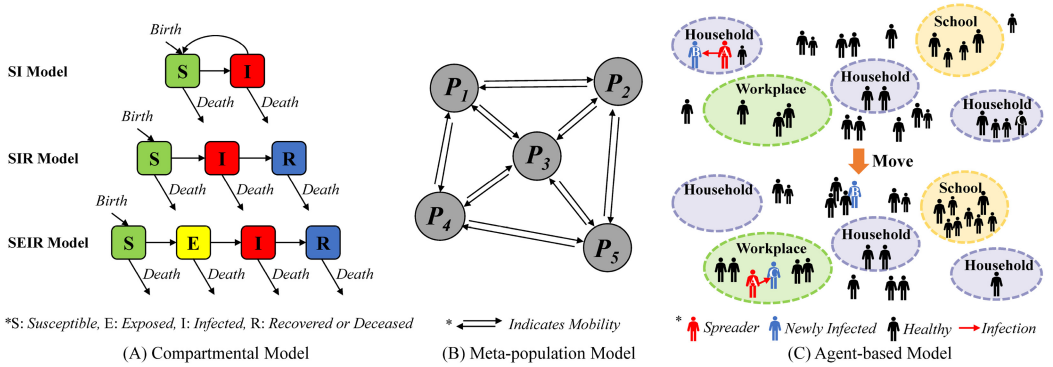

Fig. 2. Illustration of three types of infectious disease transmission models.

COVID-19 and future pandemic management tasks that use data analysis techniques, their roles in the disaster management cycle, and the related datasets used to support them.

2.1 Transmission Modeling and Prediction

Modeling and predicting the transmission of infectious diseases is critical for preventing, controlling, and managing pandemics. It can help people grasp the dynamic transmission characteristics of infectious diseases, along with simulate and predict the spread under different scenarios. Based on thorough investigation, we separate the data-driven modeling and prediction approaches into two approaches. One is the **mechanistic approach**, where predefined mathematical models or rules formulate the transmission dynamic. The data (Table 2 Main datasets) mainly accounts for model initialization and the estimation of critical transmission parameters. In light of the modeling scales, i.e., the overall, regional, and individual level, this approach can be further divided into three types: *Classical Compartmental Model* (Section 2.1.1), *Meta-population Model* (Section 2.1.2), and *Agent-based Model* (Section 2.1.3), respectively, as illustrated in Figure 2. The other is the **machine learning approach** (Section 2.1.4), which focuses on perceiving and modeling implied correlations and dependencies of data to simulate or predict the future development of infectious diseases. Note that in Tables 2 and 3, for brevity, we use different abbreviations to represent some data types: **Epidemiological data (*EPI.*), Mobility data (*MOB.*), Demographic data (*DG.*), Socioeconomic data (*SE.*), Policy data (*POL.*), Temperature data (*TEMP.*)**.

2.1.1 Classical Compartmental Model. The classical compartmental models consider the spread of infectious diseases from an overall perspective. It divides the population into several mutually disjoint compartments, representing different health statuses. During simulation, people flow between compartments, making the population size of each status vary over time. Such a process is formalized by several differential equations, which depict the transition order and rate between statuses. Figure 2(A) illustrates three typical compartmental models: SR, SIR, and SEIR, which are widely used during COVID-19 [22–25]. Taking the SEIR model as an example, it includes four compartments that depict the health status: **Susceptible (S), Exposed (E), Infected (I)**, and **Recovered or death (R)**. Three transmission parameters quantify the rate of transition in the unit of time: spreading rate (S->E), incubation rate (E->I), and recovery/death rate (I->R). During COVID-19, various modified compartmental models with different compartment design schemes also emerged, e.g., Reference [24] took into account the super-spreaders, and Reference [25] considered the quarantine.

Table 2. Summary of Mechanistic Models for COVID-19 Transmission Modeling and Prediction

Category	Ref.	Compartments	Highlighted Task	Main Datasets				
				EPI.	*MOB.*	*DG.*	*SE.*	*POL.*
Classical Compartmental Model	[22]	SIR	Modeling	✓				✓
	[23]	SEIR	Simulation	✓				
	[24]	Extended SEIR	Modeling	✓				
	[25]	Extended SEIR	Modeling	✓				
Meta-population Model	[26]	SEIR	Simulation	✓	✓	✓		
	[27]	SEIR	Simulation	✓	✓	✓		
	[28]	Extended SEIR	Modeling	✓	✓	✓		
	[29]	Extended SEIR	Modeling	✓	✓	✓		✓
	[30]	Extended SEIR	Simulation	✓	✓	✓		✓
Agent-based Model	[31]	SEIR	Simulation	✓	✓			
	[32]	Extended SEIR	Simulation	✓	✓	✓		
	[33]	Extended SEIR	Simulation	✓	✓	✓		
	[34]	Extended SEIR	Simulation	✓		✓		
	[35]	User-defined	Simulation	✓		✓		
	[36]	Extended SIR	Simulation	✓		✓	✓	

To describe the observed infection condition with the epidemic model, the model parameters need to be estimated first. Afterwards, it is possible to conduct retrospective analysis of real-world propagation. For instance, by utilizing a least-square-based method with Poisson noise, Kuniya et al. estimated the transmission parameters of the SEIR model, revealing the early transmission dynamics of Japan [23]. Using the historical epidemical data as priors, Dehning et al. combined the SIR model with Bayesian inference to detect the change points of spreading rate over time [22]. The estimated results matched well with the announcement time of government interventions, and the corresponding spreading rate quantified the effect of interventions. According to the estimated parameters, future transmission can be simulated/predicted. For example, *holding the parameters constant*, References [23, 25] predicted the future magnitude of the epidemic for India and Japan, respectively. Different from fixed parameters directly, Reference [22] used samples from the parameters' posterior distribution to predict the future scenarios. Moreover, by *manually modifying the model parameters*, the expected effects of interventions under different intensities were also simulated. In particular, Chatterjee et al. evaluated the potential impact of different quarantine compliance on the healthcare system [25].

The classical compartmental model has shown great modeling and prediction performance on global scale. Additionally, benefiting from the model's simplicity, the computation is fast. However, it ignores the spatial distribution variation of individuals, treating all people homogeneous and uniformly mixed. Therefore, it could not capture the disease's spatial dissemination process.

2.1.2 Meta-population Model. The meta-population model, as shown in Figure 2(B), is a general term for the spatially extended compartment model. It treats the target population as spatially distinct patches (i.e., administrative regions in real-world scenarios), resorting to inter-patch mobility for explaining the spatial spread process. However, the inner-patch transmission model remains unchanged, still using the compartment model.

Table 2 shows the state-of-the-art works using this model, which adopt various compartment classes and multiscale mobility data to accommodate different research objectives. Here, we summarize the objectives into three categories. (a) *Estimating the complex and variable epidemic characteristics of COVID-19.* Specifically, Li et al. divided the infection compartment of SEIR into documented (i.e., confirmed or observed infections) and undocumented. The infection compartment is then combined with inter-city migration data to investigate the infectivity and impact of undocumented cases for different outbreak stages [28]. An ablation test towards mobility data was performed to demonstrate the superiority of spatial models versus single-location ones. Similar to

the goal of Reference [28], Gatto et al. considered ex-onset transmission and post-onset transmission at different symptom severity levels in the compartment design to examine the corresponding differences in transmissibility [29]. It is noteworthy that in light of the uncertainty of case reporting (e.g., untested cases), Reference [29] selected indicators like the number of hospitalized, death, and discharged, to estimate the transmission parameters. Most research typically used reported case numbers for parameter estimation [26, 28]. (b) *Assessing the impact of mobility restrictions.* For example, early in the COVID-19 pandemic, by combining the SEIR model with ground mobility data and airline data, Chinazzi et al. conducted what-if analysis to evaluate the impact of the Wuhan travel ban on the spread at both national and international scales [26]. Using commuting data released by official and mobile phone location data in the same period, Gatto et al. retrospectively reproduced the epidemic dynamics at the provincial level in Italy, revealing the effect of progressive mobility restrictions [29]. Instead of directly coupling inter-regional mobility networks like References [26, 28, 29], Chang et al. introduced POI into the network and considered POI as a medium for cross-regional dissemination [27]. Precisely, based on the aggregated mobile phone location data, they constructed a bipartite graph with time-varying edges. The **census block groups (CBGs)** and POIs are treated as nodes, and the dynamic edge indicates the number of visits from CBG to POI in that hour. On this basis, the impact of mobility restriction at the POI level could be assessed. Moreover, by integrating the demographic data, the demographic disparities in mobility and infections are investigated. (c) *Simulating the spread under different reopening plans.* As the epidemic develops, relaxing emergency containment measures and making restart strategies becomes crucial. An extended work of Reference [29] predicted the possible rebound after lifting the lockdown in the presence of increased transmission rates, estimated the corresponding isolation efforts needed to maintain the status quo [30]. By disrupting the network [27], Chang et al. further projected the effect of different POI reopening strategies, such as constraining the maximum occupancy and reopening specified categories only. To enhance the usefulness of the model, they developed an interactive dashboard to help policymakers evaluate the effect of mobility restriction policies in near real-time [37]. However, the computational performance becomes a bottleneck due to the large scale of the network. Thus, an incremental updating and parallelized version of their previous model is devised to reduce the simulation time. To summarize, the existing meta-population work integrates mobility data to model the critical parameters and the geography spread of the epidemic. Afterwards, what-if analysis and prediction of the future could be performed by disrupting the model parameters (e.g., mobility network). It is important to note that handling mobility is not all you need to analyze the past or project the future, but only a medium for controlling spatial dissemination. Changes in human behavior such as social distance [29], wearing masks [37], are also vital to consider.

2.1.3 Agent-based Model. The **agent-based model (ABM)** is an approach for emulating the interactions of agents to observe and understand the behavior of complex systems. Distinct from the other two mechanistic models, ABM models the spread of infectious diseases at the individual scale. These individuals are heterogeneous, with attributes such as age, gender, and occupation. These individuals could also be the members of specific mixing groups, such as household, workplace, and school. As time passes, the agent moves among groups and the contacts within a group lead to the diseases spread. Most existing works [32, 34, 35] use demographic data to synthesize these agents, called synthetic population. For example, Chang et al. generated over 24 million agents for modeling and simulating the spread dynamics of COVID-19 over household, workplace, and school in Australia [32]. Instead of synthesizing "fake" people and their interactions (i.e., contact), some works resort to actual GPS trajectories to trace the movement and contacts of agents [31, 33]. In these cases, the contacts and disease spreading are not only labeled with the geographic

Table 3. Summary of Machine Learning Models for COVID-19 Transmission Modeling and Prediction

Main Target	Ref.	Method	Metrics	Granularity		Main Model Features			
				Spatial	Temporal	*EPI.*	*MOB.*	*DG.*	*TEMP.*
Epidemic Parameters	[38]	DL	RMSE, R^2	Country	Daily	✓	✓	✓	
	[39]	DL	RMSE, MAE, MSE	State	Daily		✓		
	[40]	DL	KL divergence	Grid	Weekly	✓	✓	✓	
Epidemic Curve	[41]	Non-DL	RMSE, MAE	State	Daily	✓	✓		✓
	[42]	Non-DL	RMSE, MAE	County	Weekly	✓	✓	✓	✓
	[43]	DL	RMSE, Pearson's r	Country	Daily	✓			
	[44]	DL	RMSLE, Pearson's r	County	Daily	✓	✓		
	[45]	DL	RMSE, MAE, RAE	County	Daily, Weekly	✓	✓		

semantics, but could also be modeled at finer spatial granularity, such as grid [31] and POI [33] levels.

By controlling the behavior of agents, the effects of an ensemble of non-pharmaceutical interventions has been simulated, such as quarantine [32–36], contact tracing [33–35], testing [33, 35], school closures [32, 35], and telecommuting [31]. In particular, López et al. devised an ABS to analyze the impact of different digital tracing app installation rates [34]. They found that higher adult coverage could deliver indirect protection for the elderly. Unlike most ABM work that only treats a person as an agent, Silva et al. modeled individuals, government, households, business, and health care systems as separate agents. Economic relationships among them were constructed to explore the economic impact of different intervention strategies [36]. Compared to the other two mechanical methods, the ABM method can support a wider range of strategies in a more intuitive manner (manipulate directly at the individual level), while obtaining an output with detailed demographics. For example, mask-wearing [35, 36] and vaccination [35] could reflect on adjusting the transmissibility per contact directly. Contact tracing can be achieved by registering the historical contact [33–35], and isolation can be accomplished by cutting off all the connections [32–36]. Nevertheless, ABM is often accompanied by a significant computational burden, as there are a large number of heterogeneous individuals with various statuses, interaction rules, and contact history to be maintained simultaneously. To solve such a problem, a common practice is to introduce scaling factors that allow one person to represent multiple people, accelerating by reducing the number of agents [35]. However, this may also pose sparsity issues; for more information, please refer to Reference [35].

2.1.4 Machine Learning Model. In this section, we discuss the application of **Machine Learning (ML)** model for transmission modeling and prediction from two aspects, **task** and **data**,break respectively. **Task 1**: *Model parameters simulation/prediction.* According to the analysis on the mechanical method, when performing simulation and prediction tasks, the mechanical method mainly relies on the manual configuration of key model parameters to accommodate complex scenarios [23, 26, 27, 30], posing high demands for expert experiences. Moreover, the granularity of some parameters may make manual settings harder, e.g., fine-grained mobility network [27, 30]. In such a context, several research efforts have been conducted to utilize deep learning techniques (e.g., **Deep Neural Network-DNN** [38, 40], **Graph Neural Network-GNN** [39]) to learn the potential relationship between complex scenarios and model parameters at different spatial-temporal granularity (see Table 3, Epidemic Parameters). The learning targets include disease parameters [38], mobility patterns [39, 40]. For example, a spatial-temporal **conditional Generative Adversarial Network (cGAN)** was proposed to model the relationship between various real-world conditions (e.g., policy interventions, COVID-19 statistics) and fine-grained mobility [40]. Then, the mobility data under different intervention conditions can be generated for propagation simulations. **Task 2**: *End-to-End epidemic curves prediction.* Instead of estimating the model parameters

first and then bringing them into the mechanical model for prediction [38, 39], many works directly apply machine learning models to predict future epidemic curves (see Table 3, Epidemic Curves). They can be divided into two types according to whether they use **Deep learning (DL) or not (Non-DL)**. For the Non-DL model, classical statistical models enhanced by multisource real-world data have been designed for COVID-19 scenarios. For example, mobility marked Hawkes Processes model [41], Spatiotemporal Autoregressive model integrating inter-and intra-county human interactions [42]. Whereas DL models are primarily used to uncover hidden dependencies (e.g., temporal [43, 46], spatial-temporal [44, 45]) of multisource data. For example, variants of GNN [44, 45] have been proposed to capture the potential spatial-temporal dependencies between mobility trend data and future case count. **Data** is the cornerstone of machine learning models. Table 3 provides a checklist of different types of data used by the aforementioned models. It reveals the importance of mobility and epidemiological data for ML model features. In particular, the types of commonly used mobility data include OD data [39, 41, 42, 44]—capturing cross-regional transmission, and POI visits count—sensing potential contacts [38, 40, 44, 45]. In addition, there are also some works using temperature data [41, 42], socioeconomic and demographic data [38, 40, 42] for feature engineering. Temperature is related to the transmissibility of virus, while the socioeconomic and demographic data models region-specific features.

2.2 Economic and Social Impact Estimation

2.2.1 Economic Impact Estimation. The economic impact due to COVID-19 on various industries and its possible outcomes have been investigated by various studies. Data-driven techniques can be utilized to analyze and quantify these economic impacts and predict future trends. For example, Ou et al. utilized a multi-layer perceptron neural network to predict the impact of COVID-19 on the demand of gasoline in the United States [47]. The model uses multiple data sources as inputs, including epidemiology data, government orders, and demographic data. And the neural network produces a human mobility index as output, a metric that closely correlates with motor gasoline consumption and demand. They found out that a lock-down with sufficient length will reduce the gas demand initially but helps to boost the demand back to pre-pandemic level faster due to its effect on curbing the infection rate.

COVID-19 has also heavily affected the stock markets. Investors, regulators, and policymakers constantly monitored the performance of the financial markets during the pandemic, as there is a close correlation between COVID-19 and stock market performance [48]. Baek et al. tried to explain the changes in stock market volatility using a Markov Switching AR Model with various feature selection algorithms [49]. The model takes daily U.S. stock index values, macro-economic signals, and daily COVID-19 cases as input to specify volatility fluctuation for the United States stock markets based on the generated CRSP Value Weighted Market Index Returns. The study found out that the stock market is very sensitive and shows significant risks for all industries when the number of COVID-19 cases soars.

Studies investigating the impact of pandemics on consumption also demonstrate how COVID-19 could alter consumer behavior and impact the economy. Chen et al. attempted to estimate the impact of pandemics on goods and service spending based on daily transaction data in more than 200 cities in China [50]. A difference-in-differences regression model was adopted to evaluate the influence of COVID-19 on daily offline consumption. The simulation results suggest that consumption could greatly benefit from effective virus containment policies despite the consumption decreases in the early stage.

2.2.2 Social Impact Estimation. COVID-19 has also impacted various aspects of human society, where data analysis techniques can facilitate in both measuring and analyzing the effects of the

pandemic. One significant effect of COVID-19 is the increase in stress and mental health problems. Sentiment analysis can be applied to social media data as a probe to assess the psychological impact of COVID-19. Sentiment in social media posts can be generated using transformer-based DNN models. COVID-19 outbreaks are found to significantly and negatively affect estimated sentiment across social media [51]. Furthermore, increases in frequency and exposure length to social media during COVID-19 led to a higher prevalence of depression and anxiety in the general public [52]. Survey data collected from citizens from various regions in China suggest that such mental health problems are related to unnecessary stress caused by false information spread over social networks.

Another critical social impact of COVID-19 that can be perceived by data-driven techniques is the public's opinion towards the pandemic and associated policies and countermeasures. Social media data regarding policies, countermeasures, and their statistics can be analyzed to understand the public's opinions. Han et al. developed a topic classification model using Chinese word segmentation, **Latent Dirichlet Allocation (LDA)**, and random forest to analyze COVID-19-related Weibo data in a hierarchical manner [53]. Time-series analysis is performed on the number of posts and sentiment of the content within each topic, along with their spatial distribution. It is also possible to detect synchronous changes in public opinion with the progression of COVID-19 across different regions. Similar approaches to analyze social media data from various platforms including Twitter, Telegram, and so on, have been developed [54, 55]. The public opinion and its dynamics on different topics, such as quarantine, vaccination, can thus be obtained [54, 56].

2.3 Public Awareness Enhancement

Due to the high contingency of COVID-19, the public needs to learn how they can protect themselves to effectively mitigitate transmission. The government has provided guidelines for effective self-protection methods to the public, since the early stages of the pandemic. Pre-prints, instead of peer-reviewed journals, have been preferred to publish findings regarding COVID-19 to obtain earlier visibility [57]. However, limited understanding about the disease and the rapid information dissemination process make released information less reliable, especially for an unknown virus such as SARS-CoV-2. Meanwhile, social media platforms, such as Twitter, have become the main channel for people to acquire pandemic-related information and for healthcare professionals to disseminate their findings [58]. Misinformation can be extremely prevalent. For example, 25% of the top-viewed videos on YouTube contained misinformation about COVID-19 [59] with such phenomena occurring across all social media platforms [60]. As a result, misinformation has been created and distributed along with correct knowledge and the mix of true and false information could confuse the public and regular audiences. Moreover, rumors are found to be spread over social media much faster than the facts [61], which raises the "infodemic" challenge. Misinformation spread across social networks consequently has negative impacts on the management of COVID-19 [62]. Social media exposure is also found to have a strong association with misconceptions about COVID-19 [63].

While the "infodemic" challenge can be mitigated by perceiving knowledge from reliable sources and avoiding over-exposure to social media [64], most of the public have limited awareness of the problem and/or lack the capability to differentiate the reliability of the information. Fact-checking is another commonly used method to tackle the misinformation problem. However, it is difficult to monitor massive volumes of information produced over social media and pre-print publications regarding COVID-19 and pandemics, and the requirement of professional knowledge makes it difficult to differentiate facts from questionable and wrong information [65]. Therefore, data analysis techniques can be utilized to address the problem by detecting misinformation automatically [66]. Specifically, systems that identify non-factual information regarding a new topic and according to the most updated knowledge need to be developed. Alam et al. developed several

Table 4. Summary of Outbreak/Early Case Detection Models for COVID-19

Main Target	Ref.	Method	Metrics	Spatial Coverage	Temporal Granularity	Data Source
Outbreak Detection	[69]	Non-DL	Time-lag	State (US)	Weekly	Social Media
	[70]	Non-DL	Time-lag	Country (China)	Daily	Social Media
	[71]	Non-DL	DT, DLH, PA	City (Case 2)	5 Minutes (Case 2)	Water Sensor Networks
	[72]	DL	MAE	Region (Italy)	Daily	National Government
Early Case Detection	[73]	DL	CCE	-	-	Chest X-Ray
	[74]	DL	L2, CE, BCE	-	-	Chest X-Ray
	[75]	DL	Logarithmic Loss	-	-	Chest X-Ray, CT
	[76]	Non-DL	RHR-Diff, HROS	-	Hourly	Wearable Device
	[77]	Non-DL	AUC	-	Daily	Wearable Device

transformer-based models and fine-tuned them using an annotated dataset about COVID-19 information. Metadata such as original information sources, the social media platforms to publish the data, and available fact-checking data are integrated to improve model performance [67]. However, such an approach requires a large-scale annotated dataset, which can not be obtained in the early stage of the pandemic. To address this problem, a multi-modal feature fusion framework using an ensemble of weak learners was proposed [68]. In this framework, multi-modal features available in social media, including URL, number of retweets, hashtags, mentions, and so on, are leveraged to model the users' behaviors. A meta-model integrates decisions from a set of weak learners to generate the final prediction results.

2.4 Outbreak and Early Case Detection

Outbreak and early case detection are essential to ensure that regulatory guidance can be adapted to manage pandemic situations at a given location and time. For this survey, we differentiate outbreak and early case detection as separate yet intertwined tasks. Outbreak detection deals with population-level understanding of the dynamics of a pandemic event such as COVID-19. Early case detection is concerned with the state of a specific individual through their medical data. Table 4 compares some existing methodologies presented in literature to aid in outbreak and early case detection of COVID-19.

2.4.1 Outbreak Detection. Event detection using social media data has been applied since the beginning of the pandemic for outbreak detection and surveillance [68–70]. One method to predict outbreaks using Twitter data is described in Reference [69], which searched for tweets containing common words describing symptoms of COVID-19. Their results demonstrated that Twitter discussions in the different states in the U.S. reached an informal outbreak stage from 7–19 days before drastic increases in actual case reports for COVID-19. Search engine trend data has also been used to evaluate the lag correlation coefficient between social media trends and actual outbreaks [70].

Research in outbreak detection algorithms includes models that represent the pandemic outbreak as a spread of pathogens, e.g., a computer virus, across the network. The authors in Reference [71] propose a method for the optimal placement of sensors to detect an outbreak in a network as quickly as possible. This work uses the property of sub-modularity among many common objective functions that are used to evaluate an outbreak. These objective functions include **Detection Time (DT), Detection Likelihood (DLH)**, and **Population Affected (PA)**, allowing for an efficient non-greedy method to get near-optimal node placement. Reference [72] provides an algorithm for anomaly detection using Italy's COVID-19 dataset. DL techniques such as 3D Convolution layers predict the date of the start of an outbreak in a specific region.

2.4.2 Early Case Detection. Early case detection is a necessary component of any effective way to prevent pandemic outbreaks. For example, **Computed Tomography (CT)** scans and **Chest X-Rays (CXR)** have proven to be useful in identifying COVID-19 at the early stages. To this end,

Table 5. Common Smartwatch Devices and Their Health-related Biometric Data Capabilities

Capability	Apple Watch Series 7	Fitbit Versa 3	Samsung Galaxy Watch Active 2
Heart Rate	✓	✓	✓
Heart ECG	✓	✓	✓
Step Count	✓	✓	✓
Sleep Tracking	✓	✓	✓
GPS	✓	✓	✓
Microphone	✓	✓	✓
Oxygen Saturation	✓	✓	✓
Activity Tracking	✓		✓
Skin Temperature		✓	
Electrodermal Activity		✓	
Blood Pressure			✓

many works have been proposed, aiming to automate the process of classifying such images to detect COVID-19 early. DNNs have been broadly used to analyze CXR and CT scans to detect patients infected by COVID-19 accurately [73, 74]. DNN models for early case detection are usually trained on several CXR and/or CT image datasets. Image categories include healthy, COVID-19 infected, and other unrelated medical issues such as **Pneumocystis Pneumonia (PCP)** and **Acute respiratory distress syndrome (ARDS)**. **Convolutional Neural Networks (CNNs)** have been the primary method used for image-based case classification, given the ability for CNNs to operate on image data. Transfer learning is effective in training existing CNN models such as Xception-Net, Inception-V3, and ResNeXt [73] and achieves good performance on several testing datasets when classifying using **Categorical Cross-Entropy (CCE)**. Furthermore, advanced DL models have been developed for COVID-19 early detection such as Convolutional Long Short-Term Memory [75], which achieve an even higher accuracy. Alternatively, anomaly detection models can also be applied. Abnormal images are identified and healthcare professionals can further determine whether a given patient has been infected by COVID-19 or not [74].

Moreover, the proliferation of wearables such as smartwatches and fitness trackers allows for biometric data to be collected passively, as shown in Table 5. These devices may help predict COVID-19 in those who have not yet taken a test and are therefore not officially counted as a COVID-19 case [76–78]. Smartwatch data was utilized in Reference [76], which used **resting heart rate (RHR)**, **heart rate over steps (HROS)**, sleep, and activity metrics with self-reported symptoms. These metrics led to better **Area Under the Curve (AUC)** scores than using symptoms alone to differentiate healthy, sick, and COVID-19 patients. Reference [77] leverages Logistic Regression to generate the probability of COVID-19 hospitalization based on data collected from wearable devices and self-reported symptoms. Similarly, Iwendi et al. proposed a boosted random forest algorithm to predict the possible health condition (death or recovery) of a patient based on travel history and demographic and symptom data [79]. Smartphone oxygen saturation readings were also found to be a useful proxy to detect silent hypoxia, which is an early marker of COVID-19-related pneumonia [78].

2.5 Contact Tracing

Contact tracing is another critical component in handling COVID-19 and pandemics alike. It helps to monitor pandemic spread and therefore enables early detection, efficient disease spread prevention, and less medical personnel and facility burden. In most countries, mobile devices, including smartphones and smartwatches, are adopted as the go-to platform for hosting the contact tracing application. By taking advantage of different technologies such as **Global Positioning System**

(GPS), Bluetooth, and electronic transaction data, the authorities could perform real-time contact tracing and analysis with a very high accuracy. Therefore, utilizing data-driven approaches to facilitate the harnessing and analysis of the information collected by the tracing systems has been intensively researched. Alsdurf et al. proposed a COVID-19 digital contact tracing system that uses machine learning to generate a risk factor [80]. The risk factor indicates the possibility of infection and the temporal and spatial information regarding contagiousness in the past days. More specifically, a Dynamic Bayesian Network [81] is trained using synthetic epidemiological data (contains age, sex, health conditions, location data, etc.), which predicts the contagiousness and infection status of each human subject. The proposed application also emphasizes preserving user privacy by adapting several procedures. All personal information, including exact encounter time, is removed and only coarser GPS locations are used.

In the smartphone-based contact tracing application developed by Maghdid et al. [82], a lockdown prediction model is adopted to estimate the specific lockdown area. The model uses geographic data and crowding level of the user as input and then the K-means++ algorithm [83] is used to calculate the centroid position. By utilizing the identified clusters, the model calculates the frequency of each user approaching one another. If the frequency is too high for a specific cluster (greater than 10), then the system will suggest the area to be locked down.

Besides mobile-based contact tracing applications, other platforms such as surveillance cameras placed in busy public locations could also provide data that helps to achieve human contact tracing. A study by Pi et al. [84] proposed a contact tracing system that utilized footage from surveillance cameras located at street intersections. The system uses CNN, or more specifically, the YOLO-v3 model, to analyze video footage and therefore identifies human subjects and their movement trajectories. The authors claim that the model could achieve an average precision of 69.41%. The proposed system was also used to simulate the spread of COVID-19 in a healthy population. Transfer learning is used to overcome data availability issues such as privacy.

2.6 Resource Allocation

In response to COVID-19, many countries have encountered resource shortages and starvation. This shortage is extremely severe for medical resources, such as ventilators, individual protection equipment, and so on. Thus, resource allocation, i.e., where and when to allocate the available resources fairly, ethically, and consistently becomes an important problem to mitigate the negative impacts of COVID-19 and pandemics. Resource allocation can be formulated as an optimization problem and the transmission, demographic, and mobility data can be utilized to provide effective numerical solutions. Such optimization methods have been used in many past pandemics, where the pandemics are formulated as the transmission among regions and populations in a network. The geometric programming technique can be applied to produce optimal resource allocation solutions to mitigate and manage the pandemic transmission effectively [85, 86].

However, these solutions usually assume sufficient amounts of resources are available, which is not suitable for the resource-starving scenario during COVID-19. To take the efficiency of resource usage into consideration, Lorenzo et al. proposed to allocate the available COVID-19 testings to various regions in Italy by formulating it as a quadratic optimization problem, where the COVID-19 detection capability is optimized [87]. Meanwhile, given the number of available test kits of COVID-19 and the amounts of people to be tested, an optimization algorithm has been applied to determine the best size of group testing [88]. These data analysis and modeling techniques improve testing efficiency and efficacy under limited resources.

Vaccines are another important resource to be allocated to manage pandemics globally. While vaccines have currently become available, the dissemination and allocation of vaccines to allow for fair and global access, as well as ensuring the effectiveness of group immunity across various

areas, remains a huge challenge [89]. Similar approaches to resource allocation can also be utilized to facilitate planning and decision-making for vaccine dissemination. Vaccine dissemination can be modeled as a resource distribution and allocation problem, where vaccines are distributed from storage spaces such as warehouses to the public [90]. Nowzari et al. integrate geometric programming and transmission modeling techniques to find the most effective allocation strategies to prevent pandemic transmissions given a fixed budget [91]. This proposed method can facilitate vaccine dissemination across a large region and accounts for the randomness of pandemic transmission and vaccine immunization. Furthermore, Roy et al. proposed a time-varying optimization method to allocate the vaccines while avoiding resource starvation (whenever possible) [92]. Specifically, a clustering algorithm is incorporated to determine potential storage locations for vaccines to improve the efficiency and effectiveness of the dissemination. Then, given the pre-determined storage locations, vaccine dissemination is formulated as a linear optimization problem to allocate vaccines across storage locations better. Meanwhile, real-time transmission and vaccination factors are considered to update the allocation solution dynamically.

2.7 Mental Stress Relief

Due to the pandemic and the consequent social distancing policies, it can be observed that people get depressed and stressed. Mental health therapies have been proven effective to the relief of such stress and many alternative mental health therapies during the pandemic have been developed and studied [93, 94]. In addition, regional and national call centers have been set up to help relieve mental stress for people. For example, the Washington Department of Health set up a hotline to provide consults and address people's concerns [95]. However, there are insufficient resources to serve large amounts of people to mitigate their stresses and worries. And only limited numbers of individuals can be served simultaneously via either mental health therapies or hotlines. Therefore, many citizens fail to obtain necessary guidance to respond to COVID-19 and pandemics.

The chatbot is a new technique that leverages NLP to automatically generate reasonable and human-understandable responses to language input from human beings. The data-driven model trained on massive conversation corpora enables chatbots to communicate like humans and have been successfully applied to provide clinical consults and achieve a comparable quality of services as professional medical doctors [96]. Therefore, the chatbot technique has been deployed to relieve concerns related to COVID-19, providing help and professional advice to people. For example, the pre-screening of COVID-19 can be performed by chatbots, where citizens can report their symptoms, and the chatbot will automatically generate follow-up questions and suggestions according to the text [97]. Compared to conventional call centers and hotline services, the chatbot has the potential to be widely applied and used for large populations.

2.8 Policy Recommendation

While social distancing policies are effective in controlling the pandemic transmission among the population, they can lead to adverse side effects on the economy and society. As mentioned above, these side effects can be modeled based on the observed data. And thus, it is possible that policies can be determined by balancing their pros and cons for future outbreaks of COVID-19 and other pandemics.

To this end, Kompella et al. proposed to leverage reinforcement learning methods, where the optimal policies can be learned to minimize the designed goal of controlling the pandemic without suffering unaffordable social impacts in terms of hospital capacity [98]. According to the continuous observations and monitoring of epidemiology data and contact tracing data across various areas, policies can be dynamically changed. Transmission models are leveraged to simulate the impacts of the pandemic, which will be minimized. It is demonstrated that the application of RL

to construct dynamic policies achieve the goal of avoiding exceeding hospital capacity while minimizing economic costs.

Furthermore, a **Mixed Observability Markov Decision Process (MOMDP)** problem is utilized to formulate the policy recommendation problem that is capable of optimizing the overall welfare of all the populations [99]. In this approach, both damages to people's health and the direct economic losses from COVID-19 are considered as a type of economic loss. Therefore, the proposed RL framework would produce estimated Pareto-optimal policies. Policymakers can use these results to reduce the economic losses and human lives on one objective without sacrificing the other.

Moreover, the **Deep Reinforcement Learning (DRL)** technique that integrates reinforcement learning with a deep neural network has also been developed to help recommend and optimize social distancing policies. Uddin et al. proposed a DRL-based approach to simulate the economic impacts of social distancing, people's living quality, and used resources, in addition to the pandemic transmission [100]. Various DRL models, including Deep Q-Networks and Deep Deterministic Policy Gradient, were implemented and produced better results compared to full lockdown or other simple manually crafted policies.

3 DATA VISUALIZATION FOR PANDEMIC MANAGEMENT

In addition to the state-of-the-art data analytic techniques introduced in the last section, data visualization is another important data-driven technique that can facilitate pandemic management, especially enhancing situation awareness. By employing appropriate data visualization techniques, the public, emergency managers, domain experts, and first responders would understand, perceive and distribute data (i.e., ground observation, modeling results) in a much easier manner. With COVID-19 rampant, many visualization works emerged and can be broadly classified into three categories, as shown in Table 6. The first category is **Disease Characteristics**, which focuses on visualizing disease-related characteristics (e.g., symptoms, virus structure). The second category is **Human Responses**, which covers the visualization of human response behavior in the face of a pandemic. The third category is **Mitigation and Preparedness**, which concentrates on the visualization contributions toward mitigating and preparing for the pandemic.

3.1 Visualization of Pandemic Characteristics

Understanding epidemic characteristics is essential for the prevention and control of infectious disease. Here, we summarize the relevant visualization work in terms of 3D-architecture, variations, symptoms, region-based features, and transmission-based features of a virus.

3.1.1 3D Architecture. By integrating 2D microscopy scan data and additional geometric rules of a biological entity, Nguyen et al. proposed a novel visualization technique that could efficiently and accurately construct the 3D mesoscale structure of a biological entity [101]. They adopted the proposed method to visualize the 3D ultrastructure of the COVID-19 virus. In addition, Kouvril et al. proposed a method capable of visually narrating molecular structures in a documentary-like style to facilitate public understanding and dissemination [102].

3.1.2 Variations. Genetic variants of COVID-19 have been appearing and spreading around the world. Tracing and analyzing the variants is crucial for adjusting response measures and guiding related research (e.g., effective drug and vaccine development). Driven by the variants data collected worldwide, Reference [103] developed interactive data visualization platforms to assist domain experts in exploring and analyzing the structural distribution of genetic variations of SARS-CoV-2. The phylogenetic tree view has been employed by References [104, 105] to track the circulating lineages of COVID-19 and elucidate the relationships among the variants and the evolution over

Table 6. Summary of Data Visualization Tools and Methods for COVID-19 and Future Pandemic Management

Category	Subdivision	Tools/Methods
Disease Characteristics	3D Architecture	[101, 102]
	Variations	[103–105]
	Symptoms	[106, 107]
	Region-based Features	[108, 109]
	Transmission-based Features	[31, 110–113]
Human Responses	Public Responses	[114–116]
	Government Responses	[8, 117]
	Academia Responses	[21, 118–121]
Mitigation and Preparedness	For General Public	[122]
	For Policymaker	[31, 123–125]

time. In addition, the stacked area chart [104, 105] has been applied to convey how the dominant variants shift with time.

3.1.3 Symptoms. Massive COVID-19 confirmed cases yielded data about clinical symptoms (e.g., fever, muscle or body aches, loss of smell, and taste). Data visualization techniques can assist people in better understanding and perceiving the significance of these data. For example, the CDC has created illustrations[1] to describe the typical symptoms of COVID-19, which can raise public awareness and guide self-checking. Bijoy et al. designed and developed an interactive visual analytics tool that integrates views such as symptom word clouds and clustered symptom maps. This tool aims to track and analyze the spatial distribution, temporal evolution, and spatio-temporal differences of COVID-19-related symptoms in around 462 million tweet data [107]. In the UK, the daily self-reported health data from over 1.6 million individuals were collected by Drew et al. in a mobile application [106]. They visualized the geographic distribution of collected samples' symptoms in real-time and found that such surveillance could be helpful for the discovery of early infection hotspots.

3.1.4 Region-based Features. On the basis of the continuous collection of infection case information, we can aggregate incidence rate, mortality rates, demographics, and other statistical data at different spatial scales. The visualization of these data is widely integrated into interactive dashboards [108, 109], enabling tracking, analyzing, and comparing of the pandemic progression at different spatial scales (e.g., county level, country level). In particular, for representing spatial variables such as incidence rate and mortality rates, the choropleth map or user-defined markers are the most widely used ones. For example, the incidence rates around the world are encoded with circles of different radii [108]. In addition, the log scale is introduced in some dashboards[2] to display data with different orders of magnitude simultaneously for comparative analysis. Other than monitoring in the geographic and temporal dimensions, some dashboards, such as that by the Georgia Department of Public Health,[3] integrate demographic information (e.g., gender, age, and race distribution) of infections to depict the impact of COVID-19 on different groups.

3.1.5 Transmission-based Features. To better contain the spread of infectious diseases, it is vital to understand the modes of transmission. Researchers have conducted visualization studies on

[1] https://www.cdc.gov/coronavirus/2019-ncov/downloads/COVID19-symptoms.pdf.
[2] https://aatishb.com/covidtrends/.
[3] https://dph.georgia.gov/covid-19-daily-status-report.

virus transmission at different scales: *individual* and *region* level. At the *individual* level, several works have been proposed to study the transmission chain from person to person [110, 111]. By working with public health departments in Germany, Antweiler et al. proposed a new visual analytic method to identify COVID-19 infection clusters in contact tracing networks [111]. Baumgart et al. proposed a novel visualization system to explore and analyze pathogen transmission pathways in hospitals [110]. The system integrates several practical views, such as the transmission pathway view inspired by storyline visualization for efficient and intuitive contact tracing. At the *region* level, based on the real-world trajectory of Japan, Yang et al. constructed an interactive transmission network exploration view to analyze the secondary effect of different mobility restrictions [31]. On account of a sizable agent-based pandemic spread dataset, Guo et al. devised a visual analytic method to discover interesting spatial interaction patterns among regions [112]. Given the propagation source airport, the import risk model, and the effective distance between airports, Brockmann et al. visualize the transmission tree between airports worldwide [113]. Here, the authors present the distribution of import risk and the most likely spreading routes.

3.2 Visualization of Human Responses

During the pandemic, the general public decreased their risk of infection and transmission by wearing masks and reducing mobility. The government issued a series of prevention and control policies to contain the pandemic. The academic community also actively researches pandemic response strategies. Different social participants have adjusted their responses to cope with this unprecedented pandemic. Timely perception and analysis of these responses can assist us in better knowing ourselves and strengthening the strain capacity of communities. This section summarizes the visualization work for monitoring and analyzing human responses in three subdivisions: the *public responses*, *government responses*, and *academia responses*.

3.2.1 Public Responses. A web portal to monitor the county-level mobility pattern changes in the United States has been developed [114]. The dashboard is driven by large-scale location service data and visualizes the county-level mobility change in the choropleth map. Similarly, the research team from the University of Maryland developed an impact analysis platform to inform mobility and social distance change affected by COVID-19 spread and government policies [115]. Social media platforms are the windows for people to express themselves during the epidemic. After extracting the sentiment feature, Naseem et al. drew a word cloud of positive, negative, and neutral COVID-19-related tweets to sense and analyze people's expressions in different states of mind [116]. In addition, Lee et al. investigated how COVID-19-related visualizations circulated on social media, finding the cognitive divergence of different groups in similar materials [126]. Although visual analysis of social media data has been an area of extensive research, not much work has been adopted during COVID-19. We expected to see more state-of-the-art analytics tools [127, 128] adopted for pandemic management.

3.2.2 Government Responses. The government has issued a series of control policies throughout the pandemic. These policies were further collected and structured by Reference [8]. The exploratory analysis of these policies could reveal many insights. For example, by mapping a pixel-based heatmap of the containment and health policies' intensity across countries, Hale et al. discovered convergence among them within the same two-week period, even though the pandemic was quite different from country to country at that period [8]. In addition, using Lux, a recommendation-based interactive data visualization exploration library, Lee et al. explored the relationship of policies' stringency with the local's life expectancy and levels of inequality [117]. They found that a more well-developed public health infrastructure prompted a stricter response

to the early pandemic. Nevertheless, three "outlier" countries were discovered with limited public health resources but with high stringency.

3.2.3 Academia Responses. A significant amount of research has been conducted to address the emerging COVID-19 challenges in academia. Visualization techniques have been widely adopted to portray the development status of these research efforts. For example, by combining the bibliometric analysis and a series of visualization views (i.e., treemap, heatmap, choropleth map, network graph), Haghani et al. surveyed the research hotspots, term network, and geographical distribution of COVID-19-related research [118]. Radanliev et al. conducted a dedicated bibliometric investigation of scientific literature involving COVID-19 mortality, immunity, and vaccine development [21]. The relationships under the literature were mined and visualized, such as the keywords co-occurrence and international collaboration. Besides, combined with a hierarchical topic model for literature organization, Bras et al. developed a theme-based visualization method allowing quick discovery of COVID-19 research topics, trends, and resources [119]. Clinical trials are another critical component for fighting against the pandemic. Thorlund et al. have developed a dashboard to monitor the clinical trials' progress [120] worldwide to avoid unnecessary duplication of effort and promote the perception of ongoing trials. The dashboard comprises a spatial view showing geographical distribution, a network view of relationships between clinical trials, and several charts showing basic statistical information. Likewise, an interactive web application was developed to monitor vaccine development progress [121].

3.3 Visualization-assisted Mitigation and Preparedness

As an efficient tool for perceiving, understanding, and transferring information, data visualization can affect public awareness of infectious diseases and thus influence mitigation and preparedness. Moreover, visualizations can also guide the government to make decisions. In this section, relevant data visualization techniques for the mitigation and preparedness of COVID-19 from the perspective of the public and policy-maker are summarized.

3.3.1 For General Public. The dashboards, charts, and diagrams, including the visualization works mentioned above, could be used to communicate information to the general public either directly or through secondary processing. Therefore, these visualizations can influence the behavior and opinions of the public on COVID-19. In particular, Zhang et al. collected and analyzed 668 COVID-19 visualizations created for the public, revealing the detailed landscape of COVID-19 Crisis Visualizations, which is insightful for future pandemics [122].

3.3.2 For Policymaker. A number of visualization efforts have emerged to help policy-makers make decisions under a more informative context. Afzal et al. created a novel visual analytical environment capable of simulating the spread of COVID under various conditions. Some of the tunable parameters include changeable locations of initial cases, enabled/disabled air transport, different speed of spread, and other deterministic measures [123]. Through collaboration with infectious disease experts, Yang et al. proposed an interactive visual analytics system for simulating the impact of different mobility restrictions on epidemic [31]. The system was built on real-world human trajectory data. It enables users to interactively generate restricted human mobility and pandemic transmission data when a certain set of policies is enacted. Furthermore, users could perform in-depth analysis to visually explore the deployed policy's secondary effect. Another powerful tool is GLEaMviz [125], a public software that enables the exploration of realistic scenarios of infectious disease transmission at a global scale. It provides a simple, intuitive, and visual way to enhance the disease modeling and simulation setting. GLEaMviz also evaluates simulation results using various maps, charts, and data analysis tools. An online dashboard was also created to

Table 7. A Summary of COVID-19 Data Sources and Datasets

Category	Type	Sources	Resolution and Granularity
Epidemiology Data	**Case Statistics**	John Hopkins University [108]	Daily & Regional
		Our World in Data [129, 130]	Daily & Regional
	Case Report	[131–133]	Individual
Policies and Regulations		OxCGRT [8]	Daily & Regional
		ACAPS [134]	Policy-based
Mobility Data	**Trip Surveys**	[29, 135, 136]	Daily & Regional
	Aggregated Footprint Locations	Google and Apple [137, 138]	Daily & Regional
		Descartes Labs and SafeGraph [139]	Hourly/Daily & Regional
	Transportation Flux	Derived from Footprint [140]	Depends on Footprint Data
	Trajectories	Taxi and Bike [141, 142]	Per-Trip
		Social Media [16, 143]	Real-Time & Regional/Precise Location
Media Data	**Social Media**	Twitter, Weibo, Instagram [144–148]	Real-Time Posts from Users
		Data with Annotations [147, 149–151]	Real-Time Posts from Users
	News Media	Multiple Channels [152]	Articles from Various Institutes
Personal Health Data	**Medical Imaging**	CXR and CT for Diagnosis [153–157]	Images with Diagnosis
		Ultrasound Images for Diagnosis [158]	Images with Diagnosis
		CXR Images for Complicated Tasks [159, 160]	Images for Various Purpose
	Electronic Health Records	OpenSafely [161]	Individual Data without Direct Access
Other		OxCOVID19 [162]	Daily & Regional

provide real-time, geo-located risk information [124]. The aforementioned dashboard allows setting a gathering event size interactively for estimating the risk of having at least one COVID-19 case present in a gathering at the county level in the US. This is instructive for the government in adjusting size limits for activities.

4 DATA SOURCES FOR COVID-19 AND PANDEMIC MANAGEMENT

Data from different modalities collected from various sources are one of the necessary fundamental components of data-driven methods for COVID-19 and pandemic management. Therefore, governments, research communities, the private sector, and other entities have conducted collaborative efforts to collect data from various sources, develop tools and software to facilitate continuous data acquisition and aggregation, and build datasets for different applications. In this section, a list of important and well-known datasets and data acquisition tools for COVID-19 pandemic management is introduced, from daily confirmed cases to mobility data during COVID-19 to policy responses by regions and countries. The existing data are mainly within six data categories and a summary of all the existing datasets is presented in Table 7.

4.1 Epidemiology Data

4.1.1 Case Statistics. The most commonly used epidemiology data are the number of cases in various groups and categories. Case number groups and categories include the number of tested cases, confirmed cases, hospitalized cases, deaths, recoveries, vaccinated, and others. These data are usually disclosed by local authorities, media, and government agencies every day for a certain region (e.g., cities, counties, and countries). However, raw data from various regions and countries have heterogeneous formats and are not directly ready for data-driven methods. The first and well-known attempt to integrate such information globally was accomplished by John Hopkins University, where data from various countries are aggregated semi-automatically [108]. Thus, convenient data access and analysis on global epidemiology data are enabled. A similar approach is developed by "Our World in Data," where hospitalization and vaccination data are added to the dataset [129, 130].

4.1.2 Case Report. There are also datasets that exist for meta-populations at the level of cities, counties, or individual-level epidemiological data of COVID-19 cases in China [131–133]. These

datasets contain more than 13,000 individual laboratory-confirmed cases in China (outside Hubei Province). A rich set of features are present in the datasets, including when each patient showed symptoms and what symptoms and when the patient was confirmed to get COVID-19. Data also exist for hospital admission, patient travel history, and patient demographic information (e.g., age, sex). These datasets are updated regularly to include more cases.

4.2 Policies and Regulations

Since the emergence of COVID-19, policies and regulations have been announced as countermeasures to control the transmission of the pandemic. As an essential factor that significantly impacts the transmission patterns, it is valuable to develop a dataset to characterizes the policies in different regions to facilitate applications. The **Oxford Coronavirus Government Response Tracker (OxCGRT)** [8] tracks government response categorized into 17 aspects, such as school closure policies, quarantine policies, economic policies, testing regimes, and so on. The changes in policies and regulations are tracked on a daily basis to maintain timing and stringency over the pandemic period. Such datasets can help explain the differences in the impacts of COVID-19 across countries and regions, with respect to the government's responses and decision-making. As a result, better decision-making can be potentially made for future pandemics and the losses caused by future pandemics can be reduced.

Another approach to aggregate government countermeasures of COVID-19 is built by the **Assessment Capacities Project (ACAPS)**.[4] Specifically, the COVID-19 Government Measures Dataset [134] is developed, where multiple countermeasures of COVID-19 pandemics leveraged by different governments are put together. Compared to the OxCGRT dataset, the ACAPS dataset includes additional information such as the limitation in medical exports. Meanwhile, while both testing policies are recorded in both datasets, the ACAPS dataset provides some detailed information, such as whether the health screenings for travelers are enforced. Different from the OxCGRT dataset, which provides structural data and each aspect of the policies is normalized into several categories, the ACAPS dataset describes each policy in text.

4.3 Mobility Data

4.3.1 Trip Surveys. Conducting trip surveys is the most conventional way to obtain mobility data, where the journeys of the surveyed individuals on a given day or given time period are recorded. During COVID-19, several surveys are collected to understand the human mobility patterns and the regional strategies for pandemic management [29, 135, 136]. However, the survey data usually suffer significant delays between the journey and the data collection. Also, the coverage of the survey is limited and has data biases due to the limited sample size.

4.3.2 Aggregated Footprint Locations. The mobility patterns of human beings are closely related to pandemic transmission, since in-person contacts are the primary way of transmission for most infectious diseases, including COVID-19. Due to the privacy issue, public datasets are usually aggregated by countries and regions. Mobility data can be leveraged to analyze the relation between mobility patterns and epidemiological patterns and help public health officials understand community response to mobility constraint policies, such as lockdown. One of the earliest-released datasets to report trending mobility during COVID-19 is the COVID-19 community mobility reports by Google [137]. In this report, mobility indicators are reported at the region and country levels, based on the number of visitors in locations of six categories, including residential areas, workplaces, retail, and so on. In total, data from more than 150 countries have been included. Since

[4] https://www.acaps.org.

data are aggregated, no identifiable personal information remains in the dataset. Furthermore, indicators measure the number of visitors compared to a common baseline, instead of reporting absolute numbers, to further ensure data to be anonymous. Meanwhile, Apple generates a similar mobility trends report dataset [138]. Mobility data provided by Apple have a finer resolution at the city level compared to county-level data from Google. However, Apple only has general mobility information and does not categorize it into different types of locations.

More detailed information about digital footprint is made available by Descartes Labs and SafeGraph [139]. Mobility statistics, such as maximum distances traveled, and so on, are made available for the United States at the county level on a daily basis by Descartes Labs. SafeGraph leverages their digital footprint data to generate aggregated data, including time spent staying at home, the number of visitors at **Point-of-Interest (POI)** locations, and so on. County-level and census-truck-level data are aggregated every day or hour for the United States and Canada. A dashboard application is developed to visualize data from Descartes Labs and SafeGraph [139].

4.3.3 Transportation Flux. Compared to regional footprint statistics, transportation flux data can better describe the dynamics of movement, recording the volume of traffic on the roads or between regions. OD matrix is a common way to represent the transportation flux, and based on footprint data, OD matrix about inter-region mobility can also be measured and estimated [140].

4.3.4 Trajectories. The most fine-grained mobility data are trajectories. Due to recent advances in remote sensing and GPS technologies, trajectory data can be potentially obtained by service providers. Such data are usually not available to the public due to data privacy and security concerns. However, some of the trajectory datasets from public transportation, such as taxis and bikes, can be accessed by researchers and facilitate fine-grained analyses on human mobility patterns [141, 142]. Such data provide the finest information about the human mobility of a certain population, which can not be fully obtained via other mobility data types. This fine-grained mobility can be utilized to investigate the relationship between mobility, COVID-19 transmission, and social activities in detail.

Furthermore, the trajectories of an individual can also be accessed using social media records if geo-locations of their posts are available. For example, Twitter allows the user to include location information within their tweets. Following this idea, Qazi et al. released a dataset, including huge amounts of COVID-related tweets that contain users' locations [16]. Similarly, Wang et al. acquired location information from the Foursquare application [143]. Meanwhile, it is also possible to retrieve customized data via the Twitter Developer Platform.

4.4 Media Data

4.4.1 Social Media. Social media platforms allow users to freely post their thoughts about any topic and record their daily lives. In addition to the trajectories available in social media posts, social media data are valuable information sources to understand public opinion as well. While most social media data can be publicly accessed, the data might not be ready for data analysis and machine learning. Thus, researchers have developed large-scale AI-ready datasets to facilitate this problem. Some examples include Twitter streaming dataset [144], multilingual Twitter conversation dataset regarding COVID-19 [145], Weibo dataset [148], and multimedia Instagram COVID-19 dataset [146]. These datasets characterize various aspects of the social dynamics of COVID-19 and usually contain tens of or hundreds of millions of social media posts. These posts include both textual contents and metadata, and they are retrieved by using different keywords. But the data collection processes of various datasets are designated for various purposes. For example, Reference [145] is developed to analyze the dynamics of user conversations on social media, which helps identify rumors, misinformation, and negative sentiments distributed and spread on social media

platforms. Furthermore, some of the datasets have been continuously maintained and updated to include new social media data [144, 145]. As of June 2021, this dataset has included more than 1.1 billion tweets posted by users across the countries.

While social media data can be easily accessed, one challenge of building datasets for data analysis and machine learning is to obtain sufficient amounts of pandemic-specific annotations and labels for data modeling. Given the large amounts of misinformation spread over social media, a few annotated datasets for detecting COVID-19 misinformation in social media have been built [149, 150]. Human annotators have classified social media records in the datasets into different categories to indicate whether and how much misinformation is included in the posts. Another task addressed by annotated social media data is to detect the worries and negative sentiments regarding COVID-19 in the posts. Understanding public sentiment and stances is critical to the management of the global crisis, such as COVID-19. Reference [151] released a Twitter dataset, where the stance of each tweet is captured by the ClaimsKG algorithm [163]. While the pre-trained model can be used to achieve the same goal, the model performance can be further improved with additional annotated data. Kleinberg et al. developed a dataset of annotated tweet-sized texts, called Real World Worry Dataset [147]. Text to describe the sentiment and emotion about COVID-19 were written by 2,500 participants from the United Kingdom along with their rating at a nine-point scale to score their sentiment regarding the situation of COVID-19. Such datasets can be utilized to train and fine-tune the NLP model to better recognize the sentiments of contents in the tweets and other social media data.

4.4.2 News Media. News Media is another main channel to disseminate information regarding COVID-19. Therefore, these articles are also valuable to understand the public awareness of COVID-19 and pandemics. However, different presses and institutes publish in their own electronic and paper-based channels. To integrate news from different channels, Jingyuan developed a COVID-19 news and article dataset [152]. This dataset includes articles from a total of 52 major news channels across the countries and a few well-known national and international organizations, such as WHO and the CDC.

4.5 Personal Health Data

4.5.1 Medical Imaging. Sincethe COVID-19 outbreak, automated diagnosis and early detection of COVID-19 based on medical imaging data have become an important research topic. Most of the datasets are being updated continuously. CXR and CT are the two major imaging techniques used for COVID-19 detection and diagnosis. As one of the pioneer works of building public medical imaging datasets, Cohen et al. combined data from four sources [164], namely, Corona Cases,[5] the Italian Society of Medical and Interventional Radiology,[6] Radiopaedia,[7] and EuroRad.[8] This dataset contains more than 500 CXR and CT data from patients with COVID-19 and other infectious diseases such as SARS, Middle East Respiratory Syndrome, and Acute Respiratory Distress Syndrome. Metadata of each image in the dataset have been attached, including the age of the patient, the data collection date and location, and more. Many machine learning and DNN models for COVID-19 detection and diagnosis are (partially) trained and validated on this dataset [155, 164, 165]. Similarly, Zhao et al. developed a COVID-CT-Dataset containing 349 COVID-19 positive cases from 216 patients and 695 negative samples from multiple [17]. Many such small-scale datasets with hundreds to thousands of CXR and/or CT images have been developed, composed of samples from positive

[5]https://coronacases.org/.
[6]https://www.sirm.org/category/senza-categoria/covid-19/.
[7]https://radiopaedia.org/.
[8]https://www.eurorad.org/.

COVID-19 cases, negative pneumonia cases, and negative cases of normal patients [165–167]. These medical imaging datasets have been well summarized in Reference [153].

However, to achieve more reliable diagnosis and detection, it is necessary to build larger-scale datasets so the biases in trained machine learning models can be mitigated. Cohen et al. leverage a combination of large-scale negative CXR datasets [154], such as MIMIC-CXR [168], PADCHEST [169], and so on, to pre-train the deep learning model for enhanced diagnosis performance and reliability. However, due to the limited amounts of positive samples in both training and testing datasets, many problems of the model remain. Wang et al. developed a COVIDx dataset containing 13,975 CXR data, and almost every image is collected from different patients. [155]. Based on such a dataset, a deep learning model, called COVID-Net, has been trained to detect COVID-19 infected cases from the imagery, and the model can achieve 93.3% test accuracy. Meanwhile, many large-scale datasets used in publications for deep-learning-based COVID-19 diagnosis are private [156, 157]. Some of these private datasets are composed of data collected from multiple locations and can better validate the effectiveness of data-driven models.

Except for the most commonly used CXR and CT data, Born et al. developed an effective COVID-19 detection algorithm based on ultrasound images, and a small-scale dataset called **Point-of-care Ultrasound (POCUS)** was introduced [158]. POCUS integrates data from multiple sources and contains 1,103 video data, where 654 positive COVID-19 samples are included. Ultrasound is a non-invasive medical imaging technique that does less harm to patients than CSR and CT. The ultrasound technique is also more portable, and collecting ultrasound data is cheaper than collecting CSR and CT medical images, which makes it more accessible to professionals in developing countries.

In addition to diagnosis and detection of COVID-19, data-driven methods can be applied for many complicated tasks such as prognosis for severity [159], segmentation of infected areas in the scans [160], and so on. However, due to the high cost and requirements of professional knowledge, these datasets are all very small, including tens or hundreds of samples. While the current data-driven models rely on large amounts of data to achieve good performance, it is challenging to develop larger-scale annotated datasets with affordable costs.

4.5.2 Electronic Health Records. While medical imaging is an effective data source to provide informative data for COVID-19 diagnosis and detection, comprehensive decision-making and analysis should depend on complete medical history and thorough information about a patient, which can be accessed via personal health records. However, personal health records contain sensitive information that usually cannot be shared with the public. To tackle this challenge, a pseudonymized data analysis platform that allows the researchers to conduct data mining and develop data-driven methods is proposed [161]. Data analysis scripts can be submitted to the platform, and the corresponding aggregated results can be obtained. Specifically, there are more than 17 million adult electronic health records available on the platform. Therefore, the users will not have direct access to the data to maintain data privacy. At the same time, a broad range of data-driven methods can be performed on massive amounts of electronic health records.

4.6 Other Data

Other than the directly related data as described above, many supplementary data can be helpful for specific applications of COVID-19 and pandemic management. These data can be obtained via various domain-specific sources. In this survey, only a few commonly used datasets are introduced. First, population and demographic information are closely related to analyzing data patterns and developing robust models for human behaviors and pandemic transmission across various regions. Such data are usually collected and released by government agencies, such as the

United States Census Bureau,[9] Eurostat,[10] and so on. To model the social and economic impacts of pandemics, survey data and derived social and economic indexes can be used to measure the effects of pandemics [170, 171]. Another common data related to COVID-19 and pandemic management is meteorological data, i.e., weather information, which can be obtained via a numerical weather prediction dataset from NOAA,[11] the UK Met Office,[12] and more. All these data are available across various sites on the Internet, and it is non-trivial to aggregate all of them together. To this end, the OxCOVID19 dataset is proposed to integrate demographic, socioeconomic, weather data, as well as other information, such as mobility and government responses, for more than 50 countries [162]. There are many other domain-specific data related to COVID-19 and pandemic management, which are out of the scope of this survey and not included in this section.

5 OPEN ISSUES AND FUTURE RESEARCH DIRECTIONS

Since the emergence of the global COVID-19 pandemic, large amounts of innovative research and development work using data-driven techniques for COVID-19 management have been released and published. However, there remain many open issues and new directions that can be resolved and explored to leverage data-driven methods for pandemic management. Further research and development efforts should be done to prepare ourselves for future pandemics. In this section, a few important open issues and future research directions are discussed to provide insights about what needs to be done and which data-driven techniques can potentially help.

5.1 Trustworthy and Reliable Machine Learning

When data analysis results are used for decision-making in COVID-19 and pandemic management, it can be life-changing. Data-driven models that produce life-changing results should be reliable and trustworthy. How the data can be used in decision-making is critical to its outcomes, necessitating transparent and trustworthy machine learning models. The "black box" model, such as DNN, needs to be opened to provide certain transparency of how the results are generated. For example, it was recently discovered that most of the existing models built for COVID-19 detection using machine learning algorithms for diagnosis cannot be practically used in the clinical setting due to their biases and uncertainties [172]. The first step in deploying advanced machine learning models in safety-critical applications is to ensure that trustworthy results can be produced. Many existing publications lack appropriate model validations and evaluations and use internal datasets for evaluating model performance. Consequently, undetected biases in COVID-19 data could lead to misleading results [173].

A trustworthy and reliable machine learning requires fair, explainable, auditable, and safe processes and outcomes [174]. Specifically, in the context of pandemic management, fairness avoids discrimination and biases within data and prevents additional risks for certain groups of people. Explainability focuses on techniques to provide a human-understandable rationale for the model's outcomes to allow disaster managers to utilize them better. Auditability allows for monitoring and supervision of the operation of machine learning models for pandemic management by the third parties. Finally, safety focuses on preventing the models from malicious attacks.

To achieve each aspect of these desired characteristics for a machine learning model, various techniques have been proposed. For example, to ensure the model fairness, a regulation term was designed and added to a logistic regression classifier for achieving unbiased classification [175].

[9] https://www.census.gov/about/policies/open-gov/open-data.html.
[10] https://ec.europa.eu/eurostat/web/population-and-housing-census/census-data/database.
[11] https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/numerical-weather-prediction.
[12] https://www.metoffice.gov.uk/services/data.

Two main solutions are proposed to explain results from machine learning models, i.e., ex-ante and ex-post. Ex-ante approaches refer to the models with explicit decision paths from inputs to the results, such as the decision tree model, while ex-post approaches examine the local or global behaviors of the model to generate explanations [176]. Auditability can be obtained by examining the model's sensitivity of results on input features, where "decision provenance" can be leveraged [177]. Model safety can be ensured by improving the model's confidentiality (discussed in the next subsection) and integrity. The integrity of a model can be improved by incorporating error detection methods [178] and fault-tolerant model training strategies [179]. While current solutions can improve model trustworthiness, there are many challenges in each aspect to be further addressed. Meanwhile, many problems, such as how the model can achieve multiple aspects simultaneously, how the trustworthy model achieves comparable performance as the "black box" model, and so on, remain unknown. A more detailed survey for general trustworthy machine learning techniques and their limitations is available in Reference [174].

5.2 Data Privacy

All data-driven techniques depend on the quantity and quality of available data. High-quality data are one of the most important components to develop effective methods and models for managing COVID-19 and future pandemics. However, the increase of data availability and feasibility could raise concerns about data privacy, especially for health and mobility data [180]. When machine learning is applied, data privacy becomes more difficult to protect from attacks, since machine learning algorithms could automatically memorize sensitive personal information within training datasets, either intentionally or unintentionally. Malicious users can extract and recover such information from the machine learning models [181]. Furthermore, public awareness of such a problem is limited due to the lack of professional knowledge about machine learning. Therefore, the public might volunteer their sensitive personal information without knowing it.

Many privacy-preserving machine learning models have been proposed to tackle the data privacy issue in machine learning. Existing techniques mainly use three approaches, i.e., encryption, obfuscation, and aggregation, and a detailed survey of state-of-the-art privacy-preserving machine learning solutions is provided in Reference [181]. In brief, encryption protects data privacy by encrypting either training data or the trained machine learning model and preventing malicious access to the data. Obfuscation protects privacy by intentionally adding noise to the data and/or models. Aggregation intends to train the machine learning models in a distributed environment, where the final models do not directly access the raw data. However, novel and advanced privacy attacks are being discovered, and continuous research in privacy-preserving techniques should be conducted to protect the public from personal information leakage.

5.3 In-crisis Community Identification

Based on Table 1, we can observe that data-driven methods for disaster recovery are underresearched. Recovery from COVID-19 and pandemics involves many complex decision-making processes, which can be supported by data-driven techniques. Specifically, we find it important to identify those communities that need the most help from society and understand the problems faced by them [182]. During a global pandemic like COVID-19, various communities could encounter different unanticipated problems. Meanwhile, countermeasures to mitigate the pandemic transmission could further trigger other problems such as adverse effects to the economy. Therefore, to provide immediate and dynamic assists to these communities, data-driven methods that automate the in-crisis community identification can be helpful. Many community detection algorithms have been proposed within a social network [183] and based on mobility data [184]. Most of these methods follow a similar framework: (a) Treating each trajectory as a node; (b) Calculating

the distance between nodes with a predefined distance function; and (c) Using a graph clustering method to get the communities. For the members of in-crisis communities during a pandemic, significant and abrupt changes in their behaviors can be expected [185, 186], which can be utilized to identify which communities or subgroups of a community are in crisis. Given the identified communities, the behavioral changes during the pandemics can be monitored based on the collected data from various sources, such as social media, mobility, and so on. Whenever abnormal and unusual changes in the behavior of certain communities are observed, such communities might have encountered a crisis. These behavioral changes could be observed from various aspects in data. For example, suppose many members in a community turn to have negative sentiment, which can be measured by the contents of tweets posted by the community members. In that case, such a community might be suspected to be in crisis. Furthermore, when candidates of in-crisis communities are identified, social media data and other data available for the community can be analyzed to further understand their demands and problems.

5.4 Compound Disaster Management

As COVID-19 remains a threat to the public, natural disasters could disrupt the normal procedures of pandemic management. It is important to develop strategies to minimize the losses of life and properties from natural disasters in the context of COVID-19 and any future pandemics [187]. However, managing natural disasters usually involves measures, such as the displacement of large populations, which directly conflicts with the countermeasures of pandemics [188]. The risks of pandemic transmission will greatly increase if people are clustered and housed within a limited space, where social distancing is no longer feasible. It has been reported that mass evacuation and sheltering processes during COVID-19 could cause significant increases in infected cases [189]. Therefore, a combination of countermeasures can be deployed to manage the compound disasters instead of a single solution. For example, in the event of a hurricane, shelter-in-place and evacuation can be incorporated for citizens to keep safe when the shelter capacity needs to be limited to mitigate the risks of COVID-19, where risks from both natural disasters and pandemics need to be balanced.

While the complexity of the problem increases, it is harder for human disaster managers to make optimal decisions. As a result, the emergency managers could face increased stress and pressure from the unknown consequences of their decisions to manage compound disaster situations [190]. Therefore, data-driven methods can be leveraged to provide insights and suggestions for disaster planning. By integrating pandemic transmission models with human mobility and disaster-related data, the situations of natural disasters and pandemics can be simulated and estimated. These results can thereafter be used to predict the combined risks of both natural disasters and pandemics together. However, such data-driven tools to support decision-making are not yet ready. Research on managing compound disasters should be further investigated.

6 CONCLUSION

While COVID-19 has taken a toll on the world health, economy, and many other aspects, it raised broad attention to pandemic management in the research communities. Given the recent advances in data science and machine learning, data-driven techniques have been developed and have played a more important role in managing global pandemics. Unlike conventional pandemic management methods that rely heavily on human expertise and are not scalable, data-driven techniques can process data from all over the world and provide assistance to manage global pandemics. These techniques can scale to large amounts of countries and regions and massive numbers of people.

This article surveys the state-of-the-art data-driven techniques that facilitate the management of global pandemics, especially the ongoing COVID-19 pandemic. It starts with the current status

of COVID-19 and its global impacts and then moves to various phases of disaster and pandemic management where data-driven methods can contribute. Then, recent data-driven methods and algorithms that apply to some critical applications for pandemic management are presented. These applications and corresponding data-driven methods are related to the four phases in the disaster management cycle. We discuss the data used for each application, which facilitates the management of COVID-19 and future pandemics. For applications such as transmission modeling, transmission prediction, outbreak detection, and early case detection, many data-driven methods have been proposed. The methodologies, advantages, and disadvantages of these existing methods have been compared. Meanwhile, three-category data visualization techniques for pandemic management and situation awareness enhancement and existing tools used for COVID-19 are presented. Thereafter, datasets among six categories that contribute to the COVID-19 and pandemic management have been introduced, including epidemiology data, policies and regulations, mobility data, media data, personal health data, and other data. This article discusses several open issues and future research directions of the data-driven COVID-19 and future pandemic management. Several existing and potential solutions to these topics are provided, but many of them have yet to be fully addressed, as data-driven techniques are not yet ready but could potentially help in the future.

To summarize, data-driven methods and algorithms open numerous opportunities across various applications for COVID-19 and future pandemic management. However, many challenges in data science and machine learning need to be solved to appropriately aid in solving these issues. Data-driven methods that could objectively and dynamically respond to the pandemic at scale will certainly be one of the key components in the future of pandemic management.

REFERENCES

[1] Niall P. A. S. Johnson and Juergen Mueller. 2002. Updating the accounts: Global mortality of the 1918–1920 "Spanish" influenza pandemic. *Bull. Hist. Med.* 76, 1 (2002), 105–115.

[2] Marcus Richard Keogh-Brown and Richard David Smith. 2008. The economic impact of SARS: How does the reality match the predictions? *Health Polic.* 88, 1 (2008), 110–120.

[3] Barbara Jester, Timothy Uyeki, and Daniel Jernigan. 2018. Readiness for responding to a severe pandemic 100 years after 1918. *Amer. J. Epidem.* 187, 12 (2018), 2596–2602.

[4] Joacim Rocklöv and Henrik Sjödin. 2020. High population densities catalyse the spread of COVID-19. *J. Trav. Med.* 27, 3 (2020).

[5] Domenico Cucinotta and Maurizio Vanelli. 2020. WHO declares COVID-19 a pandemic. *Acta Bio Med.: Atenei Parmen.* 91, 1 (2020), 157–160.

[6] WHO Emergency Response Team. 2021. COVID-19 Weekly Epidemiological Update. Retrieved from https://www.who.int/docs/default-source/coronaviruse/situation-reports/20210518_weekly_epi_update_40.pdf?sfvrsn=ba1c16cd_10.

[7] International Monetary Fund. 2021. *World Economic Outlook: Managing Divergent Recoveries*. International Monetary Fund, Washington, DC.

[8] Thomas Hale, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, Saptarshi Majumdar, and Helen Tatlow. 2021. A global panel database of pandemic policies (Oxford COVID-19 government response tracker). *Nat. Hum. Behav.* 5, 4 (2021), 529–538.

[9] Giuseppe Pascarella, Alessandro Strumia, Chiara Piliego, Federica Bruno, Romualdo Del Buono, Fabio Costa, Simone Scarlata, and Felice Eugenio Agrò. 2020. COVID-19 diagnosis and management: A comprehensive review. *J. Intern. Med.* 288, 2 (2020), 192–206.

[10] Wolfram Kawohl and Carlos Nordt. 2020. COVID-19, unemployment, and suicide. *Lancet Psychiat.* 7, 5 (2020), 389–390.

[11] Jerome H. Kim, Florian Marks, and John D. Clemens. 2021. Looking beyond COVID-19 vaccine phase 3 trials. *Nat. Med.* 27, 2 (2021), 205–211.

[12] Coalition for Epidemic Preparedness Innovations. 2020. CEPI survey assesses potential COVID-19 vaccine manufacturing capacity. Retrieved from https://cepi.net/news_cepi/cepi-survey-assesses-potential-covid-19-vaccine-manufacturing-capacity/.

[13] Tao Li, Ning Xie, Chunqiu Zeng, Wubai Zhou, Li Zheng, Yexi Jiang, Yimin Yang, Hsin-Yu Ha, Wei Xue, Yue Huang, Shu-Ching Chen, Jainendra Navlakha, and S. S. Iyengar. 2017. Data-driven techniques in disaster information management. *Comput. Surv.* 50, 1 (2017).

[14] Samira Pouyanfar, Saad Sadiq, Yilin Yan, Haiman Tian, Yudong Tao, Maria Presa Reyes, Mei-Ling Shyu, Shu-Ching Chen, and S. S. Iyengar. 2018. A survey on deep learning: Algorithms, techniques, and applications. *Comput. Surv.* 51, 5 (2018).

[15] Junaid Shuja, Eisa Alanazi, Waleed Alasmary, and Abdulaziz Alashaikh. 2020. COVID-19 open source data sets: A comprehensive survey. *Appl. Intell.* 51 (2020), 1296–1325.

[16] Umair Qazi, Muhammad Imran, and Ferda Ofli. 2020. GeoCov19: A dataset of hundreds of millions of multilingual COVID-19 tweets with location information. *SIGSPATIAL Spec.* 12, 1 (2020), 6–15.

[17] Jinyu Zhao, Yichen Zhang, Xuehai He, and Pengtao Xie. 2020. COVID-CT-Dataset: A CT scan dataset about COVID-19. *arXiv*, Article 2003.13865 (2020).

[18] Teodoro Alamo, Daniel G. Reina, and Pablo Millán. 2020. Data-driven methods to monitor, model, forecast and control COVID-19 pandemic: Leveraging data science, epidemiology and control theory. *arXiv*, Article 2006.01731 (2020).

[19] Gitanjali R. Shinde, Asmita B. Kalamkar, Parikshit N. Mahalle, Nilanjan Dey, Jyotismita Chaki, and Aboul Ella Hassanien. 2020. Forecasting models for coronavirus disease (COVID-19): A survey of the state-of-the-art. *SN Comput. Sci.* 1, 4 (2020), 1–15.

[20] Aishwarya Kumar, Puneet Kumar Gupta, and Ankita Srivastava. 2020. A review of modern technologies for tackling COVID-19 pandemic. *Diab. Metab. Synd.: Clin. Res. Rev.* 14, 4 (2020), 569–573.

[21] Petar Radanliev, David De Roure, and Rob Walton. 2020. Data mining and analysis of scientific research data records on COVID-19 mortality, immunity, and vaccine development-in the first wave of the COVID-19 pandemic. *Diab. Metab. Synd.: Clin. Res. Rev.* 14, 5 (2020), 1121–1132.

[22] Jonas Dehning, Johannes Zierenberg, F. Paul Spitzner, Michael Wibral, Joao Pinheiro Neto, Michael Wilczek, and Viola Priesemann. 2020. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science* 369, 6500 (2020).

[23] Toshikazu Kuniya. 2020. Prediction of the epidemic peak of coronavirus disease in Japan, 2020. *J. Clin. Med.* 9, 3 (2020).

[24] Faïçal Ndaïrou, Iván Area, Juan J. Nieto, and Delfim F. M. Torres. 2020. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos, Solit. Fract.* 135 (2020).

[25] Kaustuv Chatterjee, Kaushik Chatterjee, Arun Kumar, and Subramanian Shankar. 2020. Healthcare impact of COVID-19 epidemic in India: A stochastic mathematical model. *Med. J. Armed Forces India* 76, 2 (2020), 147–155.

[26] Matteo Chinazzi, Jessica T. Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kunpeng Mu, Luca Rossi, Kaiyuan Sun, Cécile Viboud, Xinyue Xiong, Hongjie Yu, M. Elizabeth Halloran, Ira M. Longini Jr., and Alessandro Vespignani. 2020. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* 368, 6489 (2020), 395–400.

[27] Serina Chang, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. 2021. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* 589, 7840 (2021), 82–87.

[28] Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman. 2020. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* 368, 6490 (2020), 489–493.

[29] Marino Gatto, Enrico Bertuzzo, Lorenzo Mari, Stefano Miccoli, Luca Carraro, Renato Casagrandi, and Andrea Rinaldo. 2020. Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proc. Nat. Acad. Sci.* 117, 19 (2020), 10484–10491.

[30] Enrico Bertuzzo, Lorenzo Mari, Damiano Pasetto, Stefano Miccoli, Renato Casagrandi, Marino Gatto, and Andrea Rinaldo. 2020. The geography of COVID-19 spread in Italy and implications for the relaxation of confinement measures. *Nat. Commun.* 11, 1 (2020), 1–11.

[31] Chuang Yang, Zhiwen Zhang, Zipei Fan, Renhe Jiang, Quanjun Chen, Xuan Song, and Ryosuke Shibasaki. 2022. EpiMob: Interactive Visual Analytics of Citywide Human Mobility Restrictions for Epidemic Control. *IEEE Trans. Visualiz. Comput. Graph.* (2022). DOI: http://dx.doi.org/10.1109/TVCG.2022.3165385

[32] Sheryl L. Chang, Nathan Harding, Cameron Zachreson, Oliver M. Cliff, and Mikhail Prokopenko. 2020. Modelling transmission and control of the COVID-19 pandemic in Australia. *Nat. Commun.* 11, 1 (2020), 1–13.

[33] Alberto Aleta, David Martin-Corral, Ana Pastore y Piontti, Marco Ajelli, Maria Litvinova, Matteo Chinazzi, Natalie E. Dean, M. Elizabeth Halloran, Ira M. Longini Jr, Stefano Merler, et al. 2020. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat. Hum. Behav.* 4, 9 (2020), 964–971.

[34] Jesús A. Moreno López, Beatriz Arregui García, Piotr Bentkowski, Livio Bioglio, Francesco Pinotti, Pierre-Yves Boëlle, Alain Barrat, Vittoria Colizza, and Chiara Poletto. 2021. Anatomy of digital contact tracing: Role of age, transmission setting, adoption, and case detection. *Sci. Adv.* 7, 15 (2021), eabd8750.

[35] Cliff C. Kerr, Robyn M. Stuart, Dina Mistry, Romesh G. Abeysuriya, Katherine Rosenfeld, Gregory R. Hart, Rafael C. Núñez, Jamie A. Cohen, Prashanth Selvaraj, Brittany Hagedorn, et al. 2021. Covasim: An agent-based model of COVID-19 dynamics and interventions. *PLOS Computat. Biol.* 17, 7 (2021), e1009149.

[36] Petrônio C. L. Silva, Paulo V. C. Batista, Hélder S. Lima, Marcos A. Alves, Frederico G. Guimarães, and Rodrigo C. P. Silva. 2020. COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions. *Chaos, Solit. Fract.* 139 (2020).

[37] Serina Chang, Mandy L. Wilson, Bryan Lewis, Zakaria Mehrab, Komal K. Dudakiya, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, Madhav Marathe, and Jure Leskovec. 2021. Supporting COVID-19 policy response with large-scale mobility-based modeling. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2632–2642.

[38] Salah Ghamizi, Renaud Rwemalika, Maxime Cordy, Lisa Veiber, Tegawendé F. Bissyandé, Mike Papadakis, Jacques Klein, and Yves Le Traon. 2020. Data-driven simulation and optimization for COVID-19 exit strategies. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 3434–3442.

[39] Renhe Jiang, Zhaonan Wang, Zekun Cai, Chuang Yang, Zipei Fan, Tianqi Xia, Go Matsubara, Hiroto Mizuseki, Xuan Song, and Ryosuke Shibasaki. 2021. Countrywide origin-destination matrix prediction and its application for COVID-19. In *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 319–334.

[40] Han Bao, Xun Zhou, Yingxue Zhang, Yanhua Li, and Yiqun Xie. 2020. COVID-GAN: Estimating human mobility responses to COVID-19 pandemic through spatio-temporal conditional generative adversarial networks. In *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*. 273–282.

[41] Amray Schwabe, Joel Persson, and Stefan Feuerriegel. 2021. Predicting COVID-19 spread from large-scale mobility data. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 3531–3539.

[42] Behzad Vahedi, Morteza Karimzadeh, and Hamidreza Zoraghein. 2021. Spatiotemporal prediction of COVID-19 cases using inter- and intra-county proxies of human interactions. *Nat. Commun.* 12, 1 (2021), 1–15.

[43] S. Dhamodharavadhani, R. Rathipriya, and Jyotir Moy Chatterjee. 2020. COVID-19 mortality rate prediction for India using statistical neural network models. *Front. Pub. Health* 8 (2020), 441.

[44] Amol Kapoor, Xue Ben, Luyang Liu, Bryan Perozzi, Matt Barnes, Martin Blais, and Shawn O'Banion. 2020. Examining COVID-19 forecasting using spatio-temporal graph neural networks. *arXiv preprint arXiv:2007.03113* (2020).

[45] Yue Cui, Chen Zhu, Guanyu Ye, Ziwei Wang, and Kai Zheng. 2021. Into the unobservables: A multi-range encoder-decoder framework for COVID-19 prediction. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 292–301.

[46] R. Sujath, Jyotir Moy Chatterjee, and Aboul Ella Hassanien. 2020. A machine learning forecasting model for COVID-19 pandemic in India. *Stochast. Environ. Res. Risk Assess.* 34, 7 (2020), 959–972.

[47] Shiqi Ou, Xin He, Weiqi Ji, Wei Chen, Lang Sui, Yu Gan, Zifeng Lu, Zhenhong Lin, Sili Deng, Steven Przesmitzki, and Jessey Bouchard. 2020. Machine learning model to project the impact of COVID-19 on US motor gasoline demand. *Nat. Energ.* 5, 9 (2020), 666–673.

[48] Badar Nadeem Ashraf. 2020. Stock markets' reaction to COVID-19: Cases or fatalities? *Res. Int. Bus. Fin.* 54 (2020).

[49] Seungho Baek, Sunil K. Mohanty, and Mina Glambosky. 2020. COVID-19 and stock market volatility: An industry level analysis. *Fin. Res. Lett.* 37 (2020).

[50] Haiqiang Chen, Wenlan Qian, and Qiang Wen. 2021. The impact of the COVID-19 pandemic on consumption: Learning from high-frequency transaction data. In *AEA Papers and Proceedings*, Vol. 111. American Economic Association, Nashville, TN, 307–11.

[51] Yongjian Zhu, Liqing Cao, Jingui Xie, Yugang Yu, Anfan Chen, and Fengming Huang. 2021. Using social media data to assess the impact of COVID-19 on mental health in China. *Psychol. Med.* (2021), 1–8. DOI: http://dx.doi.org/10.1017/S0033291721001598

[52] Junling Gao, Pinpin Zheng, Yingnan Jia, Hao Chen, Yimeng Mao, Suhong Chen, Yi Wang, Hua Fu, and Junming Dai. 2020. Mental health problems and social media exposure during COVID-19 outbreak. *PLoS One* 15, 4 (2020).

[53] Xuehua Han, Juanle Wang, Min Zhang, and Xiaojie Wang. 2020. Using social media to mine and analyze public opinion related to COVID-19 in China. *Int. J. Environ. Res. Pub. Health* 17, 8 (2020).

[54] Amir Karami and Mackenzie Anderson. 2020. Social media and COVID-19: Characterizing anti-quarantine comments on Twitter. *Proc. Assoc. Inf. Sci. Technol.* 57, 1 (2020).

[55] Neha Puri, Eric A. Coomes, Hourmazd Haghbayan, and Keith Gunaratne. 2020. Social media and vaccine hesitancy: New updates for the era of COVID-19 and globalized infectious diseases. *Hum. Vacc. Immunother.* 16, 11 (2020), 2586–2593.

[56] Lynnette Hui Xian Ng and Jia Yuan Loke. 2020. Analyzing public opinion and misinformation in a COVID-19 telegram group chat. *IEEE Internet Comput.* 25, 2 (2020), 84–91.

[57] Albert K. M. Chan, Chris P. Nickson, Jenny W. Rudolph, Anna Lee, and Gavin M. Joynt. 2020. Social media for rapid knowledge dissemination: Early experience from the COVID-19 pandemic. *Anaesthesia* 75, 12 (2020), 1579–1582.

[58] Latika Gupta, Durga Prasanna Misra, Vishwesh Agarwal, Suma Balan, and Vikas Agarwal. 2021. Management of rheumatic diseases in the time of COVID-19 pandemic: Perspectives of rheumatology practitioners from India. *Ann. Rheumat. Dis.* 80, 1 (2021).

[59] Heidi Oi-Yee Li, Adrian Bailey, David Huynh, and James Chan. 2020. YouTube as a source of information on COVID-19: A pandemic of misinformation? *BMJ Glob. Health* 5, 5 (2020).

[60] Matteo Cinelli, Walter Quattrociocchi, Alessandro Galeazzi, Carlo Michele Valensise, Emanuele Brugnoli, Ana Lucia Schmidt, Paola Zola, Fabiana Zollo, and Antonio Scala. 2020. The COVID-19 social media infodemic. *Scient. Rep.* 10, 1 (2020).

[61] Md Saiful Islam, Tonmoy Sarkar, Sazzad Hossain Khan, Abu-Hena Mostofa Kamal, S. M. Murshid Hasan, Alamgir Kabir, Dalia Yeasmin, Mohammad Ariful Islam, Kamal Ibne Amin Chowdhury, Kazi Selim Anwar, Abrar Ahmad Chughtai, and Holly Seale. 2020. COVID-19–related infodemic and its impact on public health: A global social media analysis. *Amer. J. Trop. Med. Hyg.* 103, 4 (2020), 1621–1629.

[62] Abhay B. Kadam and Sachin R. Atre. 2020. Negative impact of social media panic during the COVID-19 outbreak in India. *J. Trav. Med.* 27, 3, Article taaa057 (2020).

[63] Aengus Bridgman, Eric Merkley, Peter John Loewen, Taylor Owen, Derek Ruths, Lisa Teichmann, and Oleg Zhilin. 2020. The causes and consequences of COVID-19 misperceptions: Understanding the role of news and social media. *Harv. Kenn. School Misinf. Rev.* 1, 3 (2020), 18.

[64] Emily A. Holmes, Rory C. O'Connor, V. Hugh Perry, Irene Tracey, Simon Wessely, Louise Arseneault, Clive Ballard, Helen Christensen, Roxane Cohen Silver, Ian Everall, Tamsin Ford, Ann John, Thomas Kabir, Kate King, Ira Madan, et al. 2020. Multidisciplinary research priorities for the COVID-19 pandemic: A call for action for mental health science. *Lancet Psychiat.* 7, 6 (2020), 547–560.

[65] Jay J. Van Bavel, Katherine Baicker, Paulo S. Boggio, Valerio Capraro, Aleksandra Cichocka, Mina Cikara, Molly J. Crockett, Alia J. Crum, Karen M. Douglas, James N. Druckman, John Drury, Oeindrila Dube, Naomi Ellemers, Eli J. Finkel, James H. Fowler, et al. 2020. Using social and behavioural science to support COVID-19 pandemic response. *Nat. Hum. Behav.* 4, 5 (2020), 460–471.

[66] Bin Guo, Yasan Ding, Lina Yao, Yunji Liang, and Zhiwen Yu. 2020. The future of false information detection on social media: New perspectives and trends. *Comput. Surv.* 53, 4 (2020).

[67] Firoj Alam, Shaden Shaar, Fahim Dalvi, Hassan Sajjad, Alex Nikolov, Hamdy Mubarak, Giovanni Da San Martino, Ahmed Abdelali, Nadir Durrani, Kareem Darwish, and Preslav Nakov. 2020. Fighting the COVID-19 infodemic: Modeling the perspective of journalists, fact-checkers, social media platforms, policy makers, and the society. *arXiv*, Article 2005.00033 (2020).

[68] Mabrook S. Al-Rakhami and Atif M. Al-Amri. 2020. Lies kill, facts save: Detecting COVID-19 misinformation in twitter. *IEEE Access* 8 (2020), 155961–155970.

[69] Erfaneh Gharavi, Neda Nazemi, and Faraz Dadgostari. 2020. Early outbreak detection for proactive crisis management using Twitter data: COVID-19 a case study in the US. *arXiv*, Article 2005.00475 (2020).

[70] Cuilian Li, Li Jia Chen, Xueyu Chen, Mingzhi Zhang, Chi Pui Pang, and Haoyu Chen. 2020. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. *Eurosurveillance* 25, 10 (2020).

[71] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. 2007. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, 420–429.

[72] Yildiz Karadayi, Mehmet N. Aydin, and Arif Selçuk Öğrencí. 2020. Unsupervised anomaly detection in multivariate spatio-temporal data using deep learning: Early detection of COVID-19 outbreak in Italy. *IEEE Access* 8 (2020), 164155–164177.

[73] Rachna Jain, Meenu Gupta, Soham Taneja, and D. Jude Hemanth. 2021. Deep learning based detection and analysis of COVID-19 on chest X-ray images. *Appl. Intell.* 51, 3 (2021), 1690–1700.

[74] Jianpeng Zhang, Yutong Xie, Guansong Pang, Zhibin Liao, Johan Verjans, Wenxing Li, Zongji Sun, Jian He, Yi Li, Chunhua Shen, and Yong Xia. 2020. Viral pneumonia screening on chest x-rays using confidence-aware anomaly detection. *IEEE Trans. Med. Imaging* 40, 3 (2020), 879–890.

[75] Ahmed Sedik, Abdullah M. Iliyasu, Abd El-Rahiem, Mohammed E. Abdel Samea, Asmaa Abdel-Raheem, Mohamed Hammad, Jialiang Peng, Abd El-Samie, E. Fathi, and Ahmed A. Abd El-Latif. 2020. Deploying machine and deep learning models for efficient data-augmented detection of COVID-19 infections. *Viruses* 12, 7 (2020).

[76] Tejaswini Mishra, Meng Wang, Ahmed A. Metwally, Gireesh K. Bogu, Andrew W. Brooks, Amir Bahmani, Arash Alavi, Alessandra Celli, Emily Higgs, Orit Dagan-Rosenfeld, Bethany Fay, Susan Kirkpatrick, Ryan Kellogg, Michelle Gibson, Tao Wang, et al. 2020. Pre-symptomatic detection of COVID-19 from smartwatch data. *Nat. Biomed. Eng.* 4, 12 (2020), 1208–1220.

[77] Aravind Natarajan, Hao-Wei Su, and Conor Heneghan. 2020. Assessment of physiological signs associated with COVID-19 measured using wearable devices. *NPJ Digit. Med.* 3, 1 (2020), 1–8.
[78] Jason Teo. 2020. Early detection of silent hypoxia in COVID-19 pneumonia using smartphone pulse oximetry. *J. Med. Syst.* 44, 8 (2020), 1–2.
[79] Celestine Iwendi, Ali Kashif Bashir, Atharva Peshkar, R. Sujatha, Jyotir Moy Chatterjee, Swetha Pasupuleti, Rishita Mishra, Sofia Pillai, and Ohyun Jo. 2020. COVID-19 patient health prediction using boosted random forest algorithm. *Front. Pub. Health* 8 (2020), 357.
[80] Hannah Alsdurf, Yoshua Bengio, Tristan Deleu, Prateek Gupta, Daphne Ippolito, Richard Janda, Max Jarvie, Tyler Kolody, Sekoul Krastev, Tegan Maharaj, Robert Obryk, Dan Pilat, Valerie Pisano, Benjamin Prud'homme, Meng Qu, et al. 2020. Covid white paper. *arXiv*, Article 2005.08502 (2020).
[81] Sunyong Kim, Seiya Imoto, and Satoru Miyano. 2004. Dynamic Bayesian network and nonparametric regression for nonlinear modeling of gene networks from time series gene expression data. *Biosystems* 75, 1–3 (2004), 57–65.
[82] Halgurd S. Maghdid and Kayhan Zrar Ghafoor. 2020. A smartphone enabled approach to manage COVID-19 lockdown and economic crisis. *SN Comput. Sci.* 1, 5 (2020), 1–9.
[83] David Arthur and Sergei Vassilvitskii. 2006. *k-means++: The Advantages of Careful Seeding*. Technical Report 2006-13. Stanford InfoLab. Retrieved from http://ilpubs.stanford.edu:8090/778/.
[84] Yalong Pi, Nipun D. Nath, Shruthi Sampathkumar, and Amir H. Behzadan. 2021. Deep learning for visual analytics of the spread of COVID-19 infection in crowded urban environments. *Nat. Haz. Rev.* 22, 3, Article 04021019 (2021).
[85] Victor M. Preciado, Michael Zargham, Chinwendu Enyioha, Ali Jadbabaie, and George Pappas. 2013. Optimal vaccine allocation to control epidemic outbreaks in arbitrary networks. In *Proceedings of the 52nd IEEE Conference on Decision and Control*. IEEE, Piscataway, NJ, 7486–7491.
[86] Chinwendu Enyioha, Ali Jadbabaie, Victor Preciado, and George Pappas. 2015. Distributed resource allocation for control of spreading processes. In *Proceedings of the European Control Conference*. European Control Association, 2216–2221.
[87] Lorenzo Lampariello and Simone Sagratella. 2021. Effectively managing diagnostic tests to monitor the COVID-19 outbreak in Italy. *Oper. Res. Health Care* 28 (2021).
[88] Jens Niklas Eberhardt, Nikolas Peter Breuckmann, and Christiane Sigrid Eberhardt. 2020. Multi-stage group testing improves efficiency of large-scale COVID-19 screening. *J. Clin. Virol.* 128 (2020).
[89] Olivier J. Wouters, Kenneth C. Shadlen, Maximilian Salcher-Konrad, Andrew J. Pollard, Heidi J. Larson, Yot Teerawattananon, and Mark Jit. 2021. Challenges in ensuring global access to COVID-19 vaccines: Production, affordability, allocation, and deployment. *Lancet* 397, 10278 (2021), 13–19.
[90] Sunderesh S. Heragu, L. Du, Ronald J. Mantel, and Peter C. Schuur. 2005. Mathematical model for warehouse design and product allocation. *Int. J. Product. Res.* 43, 2 (2005), 327–338.
[91] Cameron Nowzari, Victor M. Preciado, and George J. Pappas. 2015. Optimal resource allocation for control of networked epidemic models. *IEEE Trans. Contr. Netw. Syst.* 4, 2 (2015), 159–169.
[92] Satyaki Roy, Ronojoy Dutta, and Preetam Ghosh. 2021. Optimal time-varying vaccine allocation amid pandemics with uncertain immunity ratios. *IEEE Access* 9 (2021), 15110–15121.
[93] Petar Radanliev and David De Roure. 2021. Alternative mental health therapies in prolonged lockdowns: Narratives from COVID-19. *Health Technol.* 11, 5 (2021), 1101–1107.
[94] Tim R. Wind, Marleen Rijkeboer, Gerhard Andersson, and Heleen Riper. 2020. The COVID-19 pandemic: The "black swan" for mental health care and a turning point for e-health. *Internet Intervent.* 20 (2020).
[95] Mike Reicher. 2020. Coronavirus call centers stumble in Washington state: Glitches, lack of staff, contradicting messages. Retrieved from https://www.seattletimes.com/seattle-news/times-watchdog/washington-coronavirus-hotlines-were-unprepared-for-onslaught-of-callers/.
[96] Jean-Emmanuel Bibault, Benjamin Chaix, Arthur Guillemassé, Sophie Cousin, Alexandre Escande, Morgane Perrin, Arthur Pienkowski, Guillaume Delamon, Pierre Nectoux, and Benoît Brouard. 2019. A chatbot versus physicians to provide information for patients with breast cancer: Blind, randomized controlled noninferiority trial. *J. Med. Internet Res.* 21, 11 (2019).
[97] Alistair Martin, Jama Nateqi, Stefanie Gruarin, Nicolas Munsch, Isselmou Abdarahmane, Marc Zobel, and Bernhard Knapp. 2020. An artificial intelligence-based first-line defence against COVID-19: Digitally screening citizens for risks via a chatbot. *Sci. Rep.* 10 (2020).
[98] Varun Kompella, Roberto Capobianco, Stacy Jong, Jonathan Browne, Spencer Fox, Lauren Meyers, Peter Wurman, and Peter Stone. 2020. Reinforcement learning for optimization of COVID-19 mitigation policies. *arXiv*, Article 2010.10560 (2020).
[99] Runzhe Wan, Xinyu Zhang, and Rui Song. 2020. Multi-objective reinforcement learning for infectious disease control with application to COVID-19 spread. *arXiv*, Article 2009.04607 (2020).

[100] M. Irfan Uddin, Syed Atif Ali Shah, Mahmoud Ahmad Al-Khasawneh, Ala Abdulsalam Alarood, and Eesa Alsolami. 2020. Optimal policy learning for COVID-19 prevention using reinforcement learning. *J. Inf. Sci.* (2020), 13. DOI: http://dx.doi.org/10.1177/0165551520959798

[101] Ngan Nguyen, Ondrej Strnad, Tobias Klein, Deng Luo, Ruwayda Alharbi, Peter Wonka, M. Maritan, Peter Mindek, L. Autin, D. Goodsell, and I. Viola. 2021. Modeling in the time of COVID-19: Statistical and rule-based mesoscale models. *IEEE Trans. Visualiz. Comput. Graph.* 27 (2021), 722–732.

[102] David Kouřil, Ondřej Strnad, Peter Mindek, Sarkis Halladjian, Tobias Isenberg, M. Eduard Gröller, and Ivan Viola. 2022. Molecumentary: Adaptable narrated documentaries using molecular visualization. *IEEE Trans. Visualiz. Comput. Graph.* (2022). DOI: 10.1109/TVCG.2021.3130670

[103] Stephanie Portelli, M. Olshansky, Carlos H. M. Rodrigues, Elston N. D'Souza, Yoochan Myung, Michael Silk, Azadeh Alavi, D. Pires, and D. Ascher. 2020. Exploring the structural distribution of genetic variation in SARS-CoV-2 with the COVID-3D online resource. *Nat. Genet.* 52, 10 (2020), 999–1001.

[104] A. Lauring and E. Hodcroft. 2021. Genetic variants of SARS-CoV-2-what do they mean? *JAMA* 325, 6 (2021), 529–531.

[105] Jitesh Chowdhury, Simon Scarr, and Jane Wardell. 2020. How the novel coronavirus has evolved. Retrieved from https://graphics.reuters.com/HEALTH-CORONAVIRUS/EVOLUTION/yxmpjqkdzvr/.

[106] David A. Drew, Long H. Nguyen, Claire J. Steves, Cristina Menni, Maxim Freydin, Thomas Varsavsky, Carole H. Sudre, M. Jorge Cardoso, Sebastien Ourselin, Jonathan Wolf, Tim D. Spector, Andrew T. Chan, and COPE Consortium. 2020. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science* 368, 6497 (2020), 1362–1367.

[107] Biddut Sarker Bijoy, Syeda Jannatus Saba, Souvika Sarkar, Md Saiful Islam, Sheikh Rabiul Islam, Mohammad Ruhul Amin, and Shubhra Kanti Karmaker Santu. 2021. COVID19α: Interactive spatio-temporal visualization of COVID-19 symptoms through tweet analysis. In *Proceedings of the 26th International Conference on Intelligent User Interfaces-Companion*. 28–30.

[108] Ensheng Dong, Hongru Du, and Lauren Gardner. 2020. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* 20, 5 (2020), 533–534.

[109] World Health Organization. 2021. WHO Coronavirus (COVID-19) Dashboard. Retrieved from https://covid19.who.int/.

[110] Tom Baumgartl, M. Petzold, Marcel Wunderlich, M. Höhn, D. Archambault, M. Lieser, A. Dalpke, S. Scheithauer, M. Marschollek, V. Eichel, N. Mutters, and T. V. Landesberger. 2021. In search of patient zero: Visual analytics of pathogen transmission pathways in hospitals. *IEEE Trans. Visualiz. Comput. Graph.* 27 (2021), 711–721.

[111] Dario Antweiler, David Sessler, Sebastian Ginzel, and Jörn Kohlhammer. 2021. Towards the detection and visual analysis of COVID-19 infection clusters. In *Proceedings of the EuroVis Workshop on Visual Analytics (EuroVA)*.

[112] D. Guo. 2007. Visual analytics of spatial interaction patterns for pandemic decision support. *Int. J. Geograph. Inf. Sci.* 21 (2007), 859–877.

[113] Dirk Brockmann. 2020. Most probable routes and effective distance. Retrieved from http://rocs.hu-berlin.de/viz/sgb/.

[114] Song Gao, J. Rao, Yuhao Kang, Yunlei Liang, and Jake Kruse. 2020. Mapping county-level mobility pattern changes in the United States in response to COVID-19. *SIGSPATIAL Spec.* 12 (2020), 16–26.

[115] Lei Zhang, Sepehr Ghader, Michael L. Pack, Chenfeng Xiong, Aref Darzi, Mofeng Yang, Qianqian Sun, AliAkbar Kabiri, and Songhua Hu. 2020. An interactive COVID-19 mobility impact and social distancing analysis platform. *medRxiv*, Article 2020.04.29.20085472 (2020).

[116] Usman Naseem, Imran Razzak, Matloob Khushi, Peter W. Eklund, and Jinman Kim. 2021. COVIDSenti: A large-scale benchmark twitter data set for COVID-19 sentiment analysis. *IEEE Trans. Computat. Soc. Syst.* 8, 4 (2021), 1003–1015.

[117] Doris Jung-Lin Lee, Dixin Tang, Kunal Agarwal, Thyne Boonmark, Caitlyn Chen, Jake Kang, U. Mukhopadhyay, Jerry Song, Micah Yong, Marti A. Hearst, and Aditya G. Parameswaran. 2021. Lux: Always-on visualization recommendations for exploratory data science. *arXiv*, Article 2105.00121 (2021).

[118] M. Haghani, M. Bliemer, F. Goerlandt, and J. Li. 2020. The scientific literature on Coronaviruses, COVID-19 and its associated safety-related research dimensions: A scientometric analysis and scoping review. *Safet. Sci.* 129, Article 104806 (2020).

[119] P. Bras, Azimeh Gharavi, D. Robb, Ana F. Vidal, Stefano Padilla, and M. Chantler. 2020. Visualising COVID-19 research. *arXiv*, Article 2005.06380 (2020).

[120] K. Thorlund, L. Dron, Jay J. H. Park, Grace Hsu, J. Forrest, and E. Mills. 2020. A real-time dashboard of clinical trials for COVID-19. *Lancet Digit. Health* 2, 6 (2020), e286–e287.

[121] M. Shrotri, T. Swinnen, B. Kampmann, and E. Parker. 2021. An interactive website tracking COVID-19 vaccine development. *Lancet Glob. Health* 9, 5 (2021), e590–e592.

[122] Yixuan Zhang, Yifan Sun, Lace Padilla, Sumit Barua, Enrico Bertini, and Andrea G. Parker. 2021. Mapping the landscape of COVID-19 crisis visualizations. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY.

[123] Shehzad Afzal, Sohaib Ghani, Hank C. Jenkins-Smith, David S. Ebert, Markus Hadwiger, and Ibrahim Hoteit. 2020. A visual analytics based decision making environment for COVID-19 modeling and visualization. In *Proceedings of the IEEE Visualization Conference*. IEEE, Piscataway, NJ, 86–90.

[124] Aroon Chande, Seolha Lee, Mallory Harris, Quan Nguyen, Stephen J. Beckett, Troy Hilley, Clio Andris, and Joshua S. Weitz. 2020. Real-time, interactive website for US-county-level COVID-19 event risk assessment. *Nat. Hum. Behav.* 4, 12 (2020), 1313–1319.

[125] W. V. D. Broeck, C. Gioannini, B. Gonçalves, M. Quaggiotto, V. Colizza, and Alessandro Vespignani. 2011. The GLEaMviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. *BMC Infect. Dis.* 11 (2011).

[126] Crystal Lee, Tanya Yang, Gabrielle D. Inchoco, Graham M. Jones, and Arvind Satyanarayan. 2021. Viral visualizations: How coronavirus skeptics use orthodox data practices to promote unorthodox science online. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.

[127] Y. Wu, Shixia Liu, Kai Yan, Mengchen Liu, and Fangzhao Wu. 2014. OpinionFlow: Visual analysis of opinion diffusion on social media. *IEEE Trans. Visualiz. Comput. Graph.* 20 (2014), 1763–1772.

[128] J. Zhao, Nan Cao, Zhen Wen, Yale Song, Y. Lin, and C. Collins. 2014. #FluxFlow: Visual analysis of anomalous information spreading on social media. *IEEE Trans. Visualiz. Comput. Graph.* 20 (2014), 1773–1782.

[129] Hannah Ritchie, Esteban Ortiz-Ospina, Diana Beltekian, Edouard Mathieu, Joe Hasell, Bobbie Macdonald, Charlie Giattino, Cameron Appel, Lucas Rodés-Guirao, and Max Roser. 2020. Coronavirus Pandemic (COVID-19). Retrieved from https://ourworldindata.org/coronavirus.

[130] Joe Hasell, Edouard Mathieu, Diana Beltekian, Bobbie Macdonald, Charlie Giattino, Esteban Ortiz-Ospina, Max Roser, and Hannah Ritchie. 2020. A cross-country database of COVID-19 testing. *Scient. Data* 7, 1 (2020), 1–7.

[131] Xiao Fan Liu, Xiao-Ke Xu, and Ye Wu. 2021. Mobility, exposure, and epidemiological timelines of COVID-19 infections in China outside Hubei province. *Scient. Data* 8, Article 54 (2021).

[132] Bo Xu, Moritz U. G. Kraemer, and Open COVID-19 Data Curation Group. 2020. Open access epidemiological data from the COVID-19 outbreak. *Lancet Infect. Dis* 20, 5 (2020), 534–534.

[133] Bo Xu, Bernardo Gutierrez, Sumiko Mekaru, Kara Sewalk, Lauren Goodwin, Alyssa Loskill, Emily L. Cohn, Yulin Hswen, Sarah C. Hill, Maria M. Cobo, Alexander E. Zarebski, Sabrina Li, Chieh-Hsi Wu, Erin Hulland, Julia D. Morgan, et al. 2020. Epidemiological data from the COVID-19 outbreak, real-time case information. *Scient. Data* 7, Article 106 (2020).

[134] ACAPS. 2020. Report on COVID19 Government Measures Updates. Retrieved from https://www.acaps.org/special-report/covid-19-government-measures-update.

[135] Fabio Della Rossa, Davide Salzano, Anna Di Meglio, Francesco De Lellis, Marco Coraggio, Carmela Calabrese, Agostino Guarino, Ricardo Cardona-Rivera, Pietro De Lellis, Davide Liuzza, Francesco Lo Iudice, Giovanni Russo, and Mario di Bernardo. 2020. A network model of Italy shows that intermittent regional strategies can alleviate the COVID-19 epidemic. *Nat. Commun.* 11, 1 (2020), 1–9.

[136] Srinivasan Venkatramanan, Adam Sadilek, Arindam Fadikar, Christopher L. Barrett, Matthew Biggerstaff, Jiangzhuo Chen, Xerxes Dotiwalla, Paul Eastham, Bryant Gipson, Dave Higdon, Onur Kucuktunc, Allison Lieber, Bryan L. Lewis, Zane Reynolds, Anil K. Vullikanti, et al. 2021. Forecasting influenza activity using machine-learned mobility map. *Nat. Commun.* 12, 1 (2021), 1–12.

[137] Google. 2021. COVID-19 Community Mobility Reports. Retrieved from https://www.google.com/covid19/mobility/.

[138] Apple. 2021. Mobility Trends Reports. Retrieved from https://covid19.apple.com/mobility.

[139] Song Gao, Jinmeng Rao, Yuhao Kang, Yunlei Liang, and Jake Kruse. 2020. Mapping county-level mobility pattern changes in the United States in response to COVID-19. *SIGSpatial Spec.* 12, 1 (2020), 16–26.

[140] Yuhao Kang, Song Gao, Yunlei Liang, Mingxiao Li, Jinmeng Rao, and Jake Kruse. 2020. Multiscale dynamic human mobility flow dataset in the US during the COVID-19 epidemic. *Scient. Data* 7, 1 (2020), 1–13.

[141] Guangyue Nian, Bozhezi Peng, Daniel Jian Sun, Wenjun Ma, Bo Peng, and Tianyuan Huang. 2020. Impact of COVID-19 on urban mobility during post-epidemic period in megacities: From the perspectives of taxi travel and social vitality. *Sustainability* 12, 19 (2020).

[142] Wen-Long Shang, Jinyu Chen, Huibo Bi, Yi Sui, Yanyan Chen, and Haitao Yu. 2021. Impacts of COVID-19 pandemic on user behaviors and environmental benefits of bike sharing: A big-data analysis. *Appl. Energ.* 285 (2021).

[143] Haotian Wang, Abhirup Ghosh, Jiaxin Ding, Rik Sarkar, and Jie Gao. 2021. Heterogeneous interventions reduce the spread of COVID-19 in simulations on real mobility data. *Scient. Rep.* 11, 1 (2021), 1–12.

[144] Juan M. Banda, Ramya Tekumalla, Guanyu Wang, Jingyuan Yu, Tuo Liu, Yuning Ding, Katya Artemova, Elena Tutubalina, and Gerardo Chowell. 2020. A large-scale COVID-19 Twitter chatter dataset for open scientific research–An international collaboration. *arXiv*, Article 2004.03688 (2020).

[145] Emily Chen, Kristina Lerman, and Emilio Ferrara. 2020. Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus Twitter data set. *JMIR Pub. Health Surveill.* 6, 2 (2020).

[146] Koosha Zarei, Reza Farahbakhsh, Noel Crespi, and Gareth Tyson. 2020. A first Instagram dataset on COVID-19. *arXiv*, Article 2004.12226 (2020).

[147] Bennett Kleinberg, Isabelle van der Vegt, and Maximilian Mozes. 2020. Measuring emotions in the COVID-19 real world worry dataset. *arXiv*, Article 2004.04225 (2020).

[148] Yong Hu, Heyan Huang, Anfan Chen, and Xian-Ling Mao. 2020. Weibo-COV: A large-scale COVID-19 social media dataset from Weibo. In *Proceedings of the 1st Workshop on NLP for COVID-19 at EMNLP*. Association for Computational Linguistics.

[149] Parth Patwa, Shivam Sharma, Srinivas Pykl, Vineeth Guptha, Gitanjali Kumari, Md Shad Akhtar, Asif Ekbal, Amitava Das, and Tanmoy Chakraborty. 2020. Fighting an infodemic: COVID-19 fake news dataset. *arXiv*, Article 2011.03327 (2020).

[150] Tamanna Hossain, Robert L. Logan IV, Arjuna Ugarte, Yoshitomo Matsubara, Sean Young, and Sameer Singh. 2020. COVIDLies: Detecting COVID-19 misinformation on social media. In *Proceedings of the 1st Workshop on NLP for COVID-19 at EMNLP*. Association for Computational Linguistics.

[151] Dimitar Dimitrov, Erdal Baran, Pavlos Fafalios, Ran Yu, Xiaofei Zhu, Matthäus Zloch, and Stefan Dietze. 2020. Tweetscov19–A knowledge base of semantically annotated tweets about the COVID-19 pandemic. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. Association for Computing Machinery, New York, NY, 2991–2998.

[152] Jingyuan Yu. 2020. Open access institutional and news media tweet dataset for COVID-19 social science research. *arXiv*, Article 2004.01791 (2020).

[153] Feng Shi, Jun Wang, Jun Shi, Ziyan Wu, Qian Wang, Zhenyu Tang, Kelei He, Yinghuan Shi, and Dinggang Shen. 2020. Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. *IEEE Rev. Biomed. Eng.* 14 (2020), 4–15.

[154] Joseph Paul Cohen, Lan Dao, Karsten Roth, Paul Morrison, Yoshua Bengio, Almas F. Abbasi, Beiyi Shen, Hoshmand Kochi Mahsa, Marzyeh Ghassemi, Haifang Li, and Tim Q. Duong. 2020. Predicting COVID-19 pneumonia severity on chest x-ray with deep learning. *Cureus* 12, 7 (2020).

[155] Linda Wang, Zhong Qiu Lin, and Alexander Wong. 2020. COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest x-ray images. *Scient. Rep.* 10, 1 (2020), 1–12.

[156] Harrison X. Bai, Robin Wang, Zeng Xiong, Ben Hsieh, Ken Chang, Kasey Halsey, Thi My Linh Tran, Ji Whae Choi, Dong-Cui Wang, Lin-Bo Shi, Ji Mei, Xiao-Long Jiang, Ian Pan, Qiu-Hua Zeng, Ping-Feng Hu, et al. 2020. Artificial intelligence augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other origin at chest CT. *Radiology* 296, 3 (2020), E156–E165.

[157] Eduardo Luz, Pedro Silva, Rodrigo Silva, Ludmila Silva, João Guimarães, Gustavo Miozzo, Gladston Moreira, and David Menotti. 2022. Towards an effective and efficient deep learning model for COVID-19 patterns detection in X-ray images. *Res. Biomed. Eng.* 38 (2022), 149–162.

[158] Jannis Born, Gabriel Brändle, Manuel Cossio, Marion Disdier, Julie Goulet, Jérémie Roulin, and Nina Wiedemann. 2020. POCOVID-Net: Automatic detection of COVID-19 from a new lung ultrasound imaging dataset (POCUS). *arXiv*, Article 2004.12084 (2020).

[159] Matthew D. Li, Nishanth Thumbavanam Arun, Mishka Gidwani, Ken Chang, Francis Deng, Brent P. Little, Dexter P. Mendoza, Min Lang, Susanna I. Lee, Aileen O'Shea, Anushri Parakh, Praveer Singh, and Jayashree Kalpathy-Cramer. 2020. Automated assessment and tracking of COVID-19 pulmonary disease severity on chest radiographs using convolutional siamese neural networks. *Radiol.: Artif. Intell.* 2, 4, Article e200079 (2020).

[160] Jun Ma, Yixin Wang, Xingle An, Cheng Ge, Ziqi Yu, Jianan Chen, Qiongjie Zhu, Guoqiang Dong, Jian He, Zhiqiang He, Tianjia Cao, Yuntao Zhu, Ziwei Nie, and Xiaoping Yang. 2020. Towards data-efficient learning: A benchmark for COVID-19 CT lung and infection segmentation. *arXiv*, Article 2004.12537 (2020).

[161] Elizabeth J. Williamson, Alex J. Walker, Krishnan Bhaskaran, Seb Bacon, Chris Bates, Caroline E. Morton, Helen J. Curtis, Amir Mehrkar, David Evans, Peter Inglesby, Jonathan Cockburn, Helen I. McDonald, Brian MacKenna, Laurie Tomlinson, Ian J. Douglas, et al. 2020. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 584, 7821 (2020), 430–436.

[162] Adam Mahdi, Piotr Błaszczyk, Paweł Dłotko, Dario Salvi, Tak-Shing Chan, John Harvey, Davide Gurnari, Yue Wu, Ahmad Farhat, Niklas Hellmer, Alexander Zarebski, Bernie Hogan, and Lionel Tarassenko. 2021. OxCOVID19 Database, a multimodal data repository for better understanding the global impact of COVID-19. *Scient. Rep.* 11, 1 (2021), 1–11.

[163] Andon Tchechmedjiev, Pavlos Fafalios, Katarina Boland, Malo Gasquet, Matthäus Zloch, Benjamin Zapilko, Stefan Dietze, and Konstantin Todorov. 2019. ClaimsKG: A knowledge graph of fact-checked claims. In *Proceedings of the 18th International Semantic Web Conference (Lecture Notes in Computer Science)*, Vol. 11779. Springer, New York, NY, 309–324.

[164] Joseph Paul Cohen, Paul Morrison, Lan Dao, Karsten Roth, Tim Q. Duong, and Marzyeh Ghassemi. 2020. COVID-19 image data collection: Prospective predictions are the future. *arXiv*, Article 2006.11988 (2020).

[165] Ioannis D. Apostolopoulos and Tzani A. Mpesiana. 2020. COVID-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Phys. Eng. Sci. Med.* 43, 2 (2020), 635–640.

[166] Shuai Wang, Bo Kang, Jinlu Ma, Xianjun Zeng, Mingming Xiao, Jia Guo, Mengjiao Cai, Jingyi Yang, Yaodong Li, Xiangfei Meng, and Bo Xu. 2021. A deep learning algorithm using CT images to screen for corona virus disease (COVID-19). *Eur. Radiol.* (2021), 1–9.

[167] Fei Shan, Yaozong Gao, Jun Wang, Weiya Shi, Nannan Shi, Miaofei Han, Zhong Xue, Dinggang Shen, and Yuxin Shi. 2020. Lung infection quantification of COVID-19 in CT images with deep learning. *arXiv*, Article 2003.04655 (2020).

[168] Alistair E. W. Johnson, Tom J. Pollard, Seth J. Berkowitz, Nathaniel R. Greenbaum, Matthew P. Lungren, Chih-ying Deng, Roger G. Mark, and Steven Horng. 2019. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Scient. Data* 6, 1 (2019), 1–8.

[169] Aurelia Bustos, Antonio Pertusa, Jose-Maria Salinas, and Maria de la Iglesia-Vayá. 2020. PadChest: A large chest x-ray image dataset with multi-label annotated reports. *Med. Image Anal.* 66 (2020), 101797.

[170] Anna Josephson, Talip Kilic, and Jeffrey D. Michler. 2021. Socioeconomic impacts of COVID-19 in low-income countries. *Nat. Hum. Behav.* 5, 5 (2021), 557–565.

[171] Amory Martin, Maryia Markhvida, Stéphane Hallegatte, and Brian Walsh. 2020. Socio-economic impacts of COVID-19 on household consumption and poverty. *Econ. Disast. Clim. Change* 4, 3 (2020), 453–479.

[172] Michael Roberts, Derek Driggs, Matthew Thorpe, Julian Gilbey, Michael Yeung, Stephan Ursprung, Angelica I. Aviles-Rivero, Christian Etmann, Cathal McCague, Lucian Beer, Jonathan R. Weir-McCall, Zhongzhao Teng, Effrossyni Gkrania-Klotsas, Alessandro Ruggiero, Anna Korhonen, et al. 2021. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nat. Mach. Intell.* 3, 3 (2021), 199–217.

[173] Beatriz Garcia Santa Cruz, Matías Nicolás Bossa, Jan Sölter, and Andreas Dominik Husch. 2021. Public COVID-19 X-ray datasets and their impact on model bias–A systematic review of a significant problem. *medRxiv*, Article 2021.02.15.21251775 (2021).

[174] Ehsan Toreini, Mhairi Aitken, Kovila P. L. Coopamootoo, Karen Elliott, Carlos Gonzalez Zelaya, and Aad van Moorsel. 2020. The relationship between trust in AI and trustworthy machine learning technologies. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, New York, NY, 272–283.

[175] Toshihiro Kamishima, Shotaro Akaho, Hideki Asoh, and Jun Sakuma. 2012. Fairness-aware classifier with prejudice remover regularizer. In *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, New York, NY, 35–50.

[176] Scott M. Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In *Proceedings of the Conference on Advances in Neural Information Processing Systems*. Curran Associates Inc., Red Hook, NY, 4765–4774.

[177] Jatinder Singh, Jennifer Cobbe, and Chris Norval. 2019. Decision provenance: Harnessing data flow for accountable systems. *IEEE Access* 7 (2019), 6562–6574.

[178] Battista Biggio, Giorgio Fumera, and Fabio Roli. 2013. Security evaluation of pattern classifiers under attack. *IEEE Trans. Knowl. Data Eng.* 26, 4 (2013), 984–996.

[179] Qiang Liu, Pan Li, Wentao Zhao, Wei Cai, Shui Yu, and Victor C. M. Leung. 2018. A survey on security threats and defensive techniques of machine learning: A data driven view. *IEEE Access* 6 (2018), 12103–12117.

[180] Marcello Ienca and Effy Vayena. 2020. On the responsible use of digital data to tackle the COVID-19 pandemic. *Nat. Med.* 26, 4 (2020), 463–464.

[181] Bo Liu, Ming Ding, Sina Shaham, Wenny Rahayu, Farhad Farokhi, and Zihuai Lin. 2021. When machine learning meets privacy: A survey and outlook. *Comput. Surv.* 54, 2 (2021).

[182] Yudong Tao, Renhe Jiang, Erik Coltey, Chuang Yang, Xuan Song, Ryosuke Shibasaki, Mei-Ling Shyu, and Shu-Ching Chen. 2021. Data-driven in-crisis community identification for disaster response and management. In *Proceedings of the 7th IEEE International Conference on Collaboration and Internet Computing*. IEEE, 96–104.

[183] Ophélie Fraisier, Guillaume Cabanac, Yoann Pitarch, Romaric Besançon, and Mohand Boughanem. 2017. Uncovering like-minded political communities on Twitter. In *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*. Association for Computing Machinery, New York, NY, 261–264.

[184] Zhongyuan Jiang, Xianyu Chen, Bowen Dong, Junsan Zhang, Jibing Gong, Hui Yan, Zehua Zhang, Jianfeng Ma, and S. Yu Philip. 2019. Trajectory-based community detection. *IEEE Trans. Circ. Syst. II: Express Briefs* 67, 6 (2019), 1139–1143.

[185] Maarit Mäkinen and Mary Wangu Kuira. 2008. Social media and postelection crisis in kenya. *Int. J. Press/Politics* 13, 3 (2008), 328–335.

[186] Rebecca Goolsby. 2010. Social media as crisis platform: The future of community maps/crisis maps. *ACM Trans. Intell. Syst. Technol.* 1, 1 (2010), 1–11.

[187] Carly A. Phillips, Astrid Caldas, Rachel Cleetus, Kristina A. Dahl, Juan Declet-Barreto, Rachel Licker, L. Delta Merner, J. Pablo Ortiz-Partida, Alexandra L. Phelan, Erika Spanger-Siegfried, Shuchi Talati, Christopher H. Trisos, and Colin J. Carlson. 2020. Compound climate risks in the COVID-19 pandemic. *Nat. Clim. Change* 10, 7 (2020), 586–588.

[188] Guy J. Abel, Michael Brottrager, Jesus Crespo Cuaresma, and Raya Muttarak. 2019. Climate, conflict and forced migration. *Glob. Environ. Change* 54 (2019), 239–249.

[189] James M. Shultz, James P. Kossin, Attila Hertelendy, Fredrick Burkle, Craig Fugate, Ronald Sherman, Johnna Bakalar, Kim Berg, Alessandra Maggioni, Zelde Espinel, Duane E. Sands, Regina C. LaRocque, Renee N. Salas, and Sandro Galea. 2020. Mitigating the twin threats of climate-driven Atlantic hurricanes and COVID-19 transmission. *Disast. Med. Pub. Health Prepared.* 14, 4 (2020), 494–503.

[190] Moreno Di Marco, Michelle L. Baker, Peter Daszak, Paul De Barro, Evan A. Eskew, Cecile M. Godde, Tom D. Harwood, Mario Herrero, Andrew J. Hoskins, Erica Johnson, William B. Karesh, Catherine Machalaba, Javier Navarro Garcia, Dean Paini, Rebecca Pirzl, et al. 2020. Opinion: Sustainable development must account for pandemic risk. *Proc. Nat. Acad. Sci.* 117, 8 (2020), 3888–3892.

Received 14 July 2021; revised 19 March 2022; accepted 23 May 2022