COVID-19 Modeling: A Review

LONGBING CAO*, University of Technology Sydney
QING LIU, University of Technology Sydney

The SARS-CoV-2 virus and COVID-19 disease have posed unprecedented and overwhelming demand, challenges and opportunities to domain, model and data driven modeling. This paper provides a comprehensive review of the challenges, tasks, methods, progress, gaps and opportunities in relation to modeling COVID-19 problems, data and objectives. It constructs a research landscape of COVID-19 modeling tasks and methods, and further categorizes, summarizes, compares and discusses the related methods and progress of modeling COVID-19 epidemic transmission processes and dynamics, case identification and tracing, infection diagnosis and medical treatments, non-pharmaceutical interventions and their effects, drug and vaccine development, psychological, economic and social influence and impact, and misinformation, etc. The modeling methods involve mathematical and statistical models, domain-driven modeling by epidemiological compartmental models, medical and biomedical analysis, AI and data science in particular shallow and deep machine learning, simulation modeling, social science methods, and hybrid modeling.

Additional Key Words and Phrases: COVID-19, SARS-CoV-2, coronavirus, pandemic, epidemic transmission, artificial intelligence (AI), data science, machine learning, deep learning, modeling, epidemiological modeling, forecasting, prediction, deep learning, biomedical analysis, statistical modeling, mathematical modeling, data-driven discovery, domain-driven modeling, simulation, influence analysis, impact modeling

1 INTRODUCTION

Here, we give a brief overview of the COVID-19 pandemic, the global effort on modeling COVID-19, and the scope, motivation and contributions of this comprehensive review.

1.1 COVID-19 Pandemic

The coronavirus disease 2019, designated as COVID-19, is a new epidemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus. SARS-CoV-2 and COVID-19 have overwhelmingly shocked and shaken the entire world. After the first outbreak in Wuhan China in Dec 2019, the disease spread rapidly across the world in only two months due to its strong human-to-human transmission ability. The World Health Organization (WHO) declared COVID-19 a pandemic on 11 March 2020. To date, COVID-19 has infected more than 194M people with 4M having lost their lives[1]. The continuous iterative mutative infections are even more seriously troubling 215 countries and territories with increasingly unexpected resurgences and virus mutations continuously challenging pandemic containment, vaccinations and treatments. With these widespread and continuous

---

*Corresponding to: Longbing Cao (Longbing.Cao@uts.edu.au).
[1]https://www.worldometers.info/coronavirus/#countries

---

infections, the increasingly contagious mutations and resurgences, and the slow rollout of global vaccination to achieve herd immunity, COVID-19 is and will continuously and fundamentally transform global public health, the economy and society with an increasingly unprecedented impact on every aspect of life. COVID-19 has not only exerted unprecedented pressure on global healthcare systems it has also fundamentally challenged the relevant scientific research on understanding, modeling, diagnosing and controlling the virus and disease. Many questions and challenges arise about COVID-19 in relation to the nature of the coronavirus; the virus' epidemic characteristics, transmission and influence processes; the disease's medical and genomic characteristics, dynamics and evolution, and the pros and cons of existing containment, diagnosis, treatment and precaution strategies.

Compared with the epidemics that have been defeated in recent decades, such as the severe acute respiratory syndrome (SARS) and the Middle East respiratory syndrome (MERS), SARS-CoV-2 and COVID-19 are more complicated and more transmissible from human to human [167]. Its strong uncertainty, transmission and mutation are some of the key factors that make COVID-19 a continuous, unprecedented and evolving pandemic. Another key challenge of COVID-19 is its long incubation period ranging from 1 to 14 days or even longer and high asymptomatic proportion. In the incubation period, infectious individuals are contagious but show no symptoms. Consequently, susceptible individuals may not be aware that they are infected, hindering their timely identification and contact tracing. Unfortunately, the number of asymptomatic and mild symptomatic infections is significant. Asymptomatic infectives are a concerning hidden source of the widespread infection and resurgence of COVID-19 [120], making it extremely difficult to be absolutely clear about the source, diagnosis and mitigation of COVID-19. This brief comparison poses challenging questions to be answered scientifically, including how does COVID-19 differ from other epidemic or endemic diseases, what is the nature of its asymptomatic phase and infection, what are more effective methods to contain and treat the virus and disease, and how does the virus mutate and react to vaccines.

Specifically, to slow the pandemic and bring infections under control, most governments have implemented many non-pharmaceutical interventions (NPI) such as social distancing, school and university closure, infective isolation or quarantine, banning public events and travel, and business lockdowns, etc. These interventions also incur significant socioeconomic costs and have a wide impact on businesses, triggering various debates on a trade-off between epidemic control and introduced negative impact, relaxed mitigation, or even herd immunity. However, it is quantitatively unclear what makes a better trade-off or timing and the appropriate extent of mitigation, what are the positive and negative impacts of NPI on pandemic control and socioeconomic wellness, and how has COVID-19 influenced other aspects of public life, work, mental and medical health, and the economy on both individual and population (e.g., a city, country to the globe) levels.

The SARS-CoV-2 virus and the COVID-19 disease present significant challenges for health care, the government, society and the economy, and also present both challenges and opportunities for the scientific and research communities. Coronavirus and COVID-19 have reshaped the focus of global scientific attention and efforts, including the exploration of the aforementioned challenges. The scientific research is comprehensive, spreading across almost every discipline from epidemiology to psychology and fostering new research areas and topics such as coronavirus epidemiology and genomics. Of the scientific efforts in 2020, *COVID-19 studies* emerged as the most important and active research area, where the growing volume of COVID-19 data is a valuable intangible asset for evidence-based virus

and disease understanding, fostering a critical research agenda and global interest in COVID-19 modeling. *COVID-19 modeling* aims to quantitatively understand and characterize the virus and disease characteristics, estimate and predict COVID-19 transmission, and identify cases and trends, intervention measures and their effects, and their impacts on social, economic, psychological and political aspects, etc. COVID-19 modeling plays an irreplaceable role in almost every aspect of the fight against the COVID-19 pandemic, in particular, in characterizing the intricate nature of COVID-19 and discovering insights for virus containment, disease treatment, drug and vaccine development, and mitigating its broad socioeconomic impact. These motivate this review of the global reaction to modeling COVID-19.

1.2 Global Effort on Modeling COVID-19

Hundreds of thousands of studies on coronavirus and COVID-19 and the related issues have been published in the literature in less than two years, as indicated by the WHO-collected global literature on the coronavirus disease COVID-19[2]. There are also an increasing number of publications, which review the relevant progress of specific aspects of coronavirus and COVID-19 studies. As of 22 February 2021, approximately 200k references have been published on issues relating to COVID-19 in various fields such as medical/biological science, computer science, economics, environment, policy, engineering, etc. to understand and study SARS-CoV-2 and COVID-19 and the associated problems. The computer science communities alone have contributed to approximately 7k publications, including work on modeling COVID-19 using so-called AI and data science techniques, in particular classic and deep analytical and machine learning methods. Data-driven discovery [36, 68] and *COVID-19 data science* play a major role in COVID-19 modeling, aiming to discover valuable knowledge and insights from various kinds of publicly available data including daily cases, texts, biomedical images, mobility, and so on. As further illustrated in this paper, every modeling technique has been applied to COVID-19 in some way. For example, classic epidemic models are tailored for COVID-19 to model its macroscopic transmission and predict the trends of the spread of the virus; generative models with Bayesian hierarchical structures are applied to capture the effects of NPI; deep natural language processing approaches are adopted to understand the growth, nature and spread of COVID-19 and people's reactions based on the textual data from social media such as Twitter. Modeling also helps to understand and characterize every aspect of coronavirus and COVID-19, from its epidemiological characteristics to the underlying genomic reactions and mutations and drug and vaccine development.

This review also shows that there are many challenges, gaps and opportunities in modeling COVID-19. First, publicly available COVID-19 data is limited with partial, inconsistent, even erroneous, biased, noisy and uncertain observations and statistics due to the limited, imbalanced and non-universal test ability, and non-unified reporting standards and statistical errors, especially at the beginning of the epidemic and in those undeveloped regions and countries. Second, the aforementioned long incubation period from infection to the onset of symptoms and the large number of asymptomatic to mild infections make correct and instant reporting difficult and leads to a significant number of undetected and unreported cases, degrading data quality and trust. Third, coronavirus and COVID-19 exhibit unique complexities, which differ from existing epidemics, including its transmission, infectivity in ethnic populations, external NPIs, people's NPI reactions and behavioral changes as a

---

[2]https://search.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/

result of COVID-19 mitigation policies, and the rapid and mysterious mutation and spread of coronavirus. Lastly, the modeled problems and areas are fragmented, and although the modeling techniques and results are highly comprehensive, they are divided and evolving.

These brief observations indicate the critical need to model COVID-19 and the urgency of forming a comprehensive understanding of the progress being made in COVID-19 modeling, the research gaps and the open issues. This overview is crucial for not only furthering COVID-19 modeling research but also for informing insights on scientific and public strategies and actions to better battle this pandemic and future pandemics.

1.3 Our Findings

In this review, we seek answers and indications to the following major questions:

- What is the research landscape of COVID-19 modeling, i.e., what COVID-19 problems can be modeled and what modeling techniques can address these COVID-19 issues?
- How well do AI and data science, specifically machine learning and deep learning, deepen and broaden the understanding and management of the COVID-19 pandemic?
- How do varied techniques perform differently in modeling COVID-19?
- What are the gaps in modeling COVID-19?
- Where can AI and data science make new, more or better difference in containing COVID-19?

This comprehensive review obtains a relatively full spectrum of the virus challenges, data issues, techniques, gaps and opportunities in relation to modeling COVID-19. In addition to many specific observations obtained through this review, as discussed in the following sections, here we highlight the following high-to-low level observations and quantitative indications of results in modeling different COVID-19 problems and data.

- *COVID-19 problems and complexities*: as summarized in Section 3.1, the spectrum of problems covers typical aspects of epidemic dynamics and transmission, virus and disease diagnosis, infection identification, contact tracing, virus mutation and resurgence, medical diagnosis and treatment, pharmaceutical interventions, pathological and biomedical analysis, drug and vaccine development, non-pharmaceutical interventions, and socioeconomic influence and impact. We further summarize the COVID-19 characteristics and complexities in Section 2.1, including complex hidden epidemic attributes, high contagion, high mutation, high proportion of asymptomatic to mild symptomatic infections, varied and long incubation periods, ethnic sensitivity, and other high uncertainties, which shows significant differences between SARS-CoV-2 and other existing viruses and epidemics.
- *COVID-19 data and challenges*: the core data is related to the daily reported number of asymptomatic infections, the number of confirmed, recovered and deceased cases, patients' demographics, pathological, clinical and genomic results of the virus and disease tests, and patients' activities and hospitalized information, etc.; external data comprises NPI policies and events, the resident's responses and behaviors, public activities, texts from online and health services, weather, and environment, etc. Since the data spectrum is indeed comprehensive, almost all data complexities widely explored in general modeling have also been involved in modeling COVID, including data uncertainty, dynamics and nonstationarity, various data quality issues such as incompleteness, inconsistency, inequality and incomparability, lack of ground truth information, and limited size of daily reports. For further discussion, see Section 2.2.

- *Modeling challenges*: the COVID-19 complexities and data challenges bring about various modeling issues and challenges, as summarized in Section 2.3, including modeling low-to-poor quality data, modeling small and limited data, learning with weak-to-no prior knowledge and ground truth, modeling hierarchical and diverse forms of heterogeneities and interactions between multi-source core and external data, and disclosing hidden and unknown attributes and dynamics of the virus and disease.
- *Modeling techniques*: as summarized in Section 3.2 and detailed in Sections 4 to 9, the spectrum of techniques is wide enough to cover conventional mathematical and statistical modeling, simulation methods and epidemiological modeling, modern data-driven discovery and machine learning, and the recent advances in deep learning, in addition to social science methods including psychological, economic and behavior modeling methods. Among the 200k publications on COVID-19 with 22k on modeling COVID-19 [38], about 4%, 6%, and 60% of the entire literature involves the research and application on broad computer science, social science, and medical science problems and methods, respectively. Of the 22k publications on modeling, about 50% applies machine learning, deep learning, mathematical modeling, epidemic modeling to address medical problems. In addition, mathematical models, machine learning, deep learning, and epidemiological models are mostly explored in modeling COVID-19, contributing some 59%, 18%, 9% and 15% to the 22k publications, respectively. In the top-10 modeling methods, regression methods contribute to 34%, CCNs to 3%, Bayesian models to 3%, SIR models to 4%, and simulation to 9% of all modeling publications.
- *Modeling tasks*: on one hand, modeling tasks address almost all of the aforementioned COVID-19 problems and complexities, amounting to 15% on epidemic modeling, 9% on diagnosis and identification, 15% on influence and impact, 7% on simulation, and 1% on resurgence and mutation; on the other hand, the literature covers overwhelming analytical and learning tasks including roughly 2% on unsupervised learning and clustering, 6% on classification, 2% on multi-source and multi-modal data modeling and multi-task learning, and 6% on forecasting and prediction.
- *Epidemic attributes*: as summarized in Section 2.1, it is estimated that the reproduction number (probably larger than 3 in the original waves and over 2 in the resurgence after receiving vaccination) is much higher than SARS and MERS, the incubation may last for an average of 5 to even beyond 14 days, the asymptomatic infections may be much higher than 20% with even up to 80% undocumented infections in some countries, some of virus mutants may have increase the transmission rate by more than 50% over the original strain.
- *Machine learning performance*: as shown in Section 5.2, shallow machine learning methods report an accuracy of over 90% in predicting COVID-19 outbreaks, over 96% in disease diagnosis on clinical reports, over 98% in diagnosis on medical images, and close to 99% in diagnosis further involving latent features with specific settings and data; in addition, as shown in Sections 5.3 and 6.2, DNN variants achieve significant prediction performance on COVID-19 images and signals, e.g., an accuracy of over 92% on cough sound using LSTM, over 99% on chest X-ray images using CNN and imagenet variants, and less than 5% MAE on real unlabelled lung CT images by attention and gated U-net.
- Non-pharmaceutical intervention effect: as discussed in Section 7.1, NPIs such as business lockdowns, school closures, limiting gatherings, and social distancing are crucial to contain the virus outbreaks and reducing COVID-19 case numbers; e.g., reducing the reproduction number by 13%-42% individually or even 77% jointly by

these control measures; and resulting in over 40% transmission reduction by restricting human mobility and interactions.

- *Emotional, social and economic impact*: the COVID-19 pandemic has generated an overwhelming negative impact on the public mental health (e.g., significantly increasing anxiety, stress, depression and suicide), economic growth and workforce (e.g., over 20% estimated annual GDP loss in 2020), public health systems, global supply chain, sociopolitical systems, and information disorder, as discussed in Section 7.1.

- *Modeling gaps*: as commented in Section 10.1, the review also finds various issues and limitations of existing research, e.g., an insufficient, biased and partial understanding of COVID-19 complexities and data challenges; a simple and direct application of modeling techniques on often simple data; lack of robust, generalizable and tailored designs and insights into the virus and disease nature and complexities.

- *Future opportunities*: the discussion in Section 10.2 indicates significant new opportunities, e.g., studying rarely to poorly addressed problems such as epidemiologically modeling mutated virus attributes, complex interactions between core and external factors, and the influence of external factors on epidemic dynamics and NPI effect; new directions and methods such as hybridizing multiple sources of data or methods to characterize the complex COVID systems; and novel AI, data science and machine learning research on large-scale simulation of the intricate evolutionary mechanisms in COVID, discovering robust and actionable evidence to dynamically personalize the control of potential resurgence and balance the economic and mental recovery and the virus containment.

Note, the above-quoted numerical results are illustrative, which do not represent the state-of-the-art performance. Interested readers may refer to [38] and specific references for more comprehensive information about how the global scientists have responded to model COVID-19 and [37] to understand what quantitative results COVID-19 modeling has identified in both the above questioned areas and other areas.

1.4 Contributions and Limitations

Several surveys have been conducted on COVID-19 modeling, which review the progress from specific perspectives, e.g., COVID-19 characteristics [70], epidemiology [173], general applications of AI and machine learning [157, 160] such as for epidemic and transmission forecasting and prediction [30, 47, 192], virus detection, spread prevention, and medical assistance [207], policy effectiveness and contact tracing [143], infection detection and disease diagnosis [28, 120], virology and pathogenesis [132], drug and vaccine development [111], and mental health [258]. The methods which have been reviewed include epidemiological modeling [173], general AI and machine learning methods [47, 106, 111, 151, 157, 160, 194], data science [122], computational intelligence [224], computer vision and image processing [209, 227], statistical models [151], and deep learning [103, 265]. These reviews paint a partial picture of what happened in their selected areas based on several references and specific techniques. However, there are currently no comprehensive surveys or critical analyses of the intricate challenges posed by the virus, the disease, the data and the modeling.

This review is the first attempt to provide a comprehensive picture of the problems by modeling coronavirus and COVID-19 data. We start by categorizing the characteristics and challenges of the COVID-19 disease, the data and the modeling in Section 2. A transdisciplinary landscape is formed to categorize and match both COVID-19 modeling tasks and objectives and categorize the corresponding methods and general frameworks in

Section 3. The review then focuses on structuring, analyzing and comparing the work on mathematical, data-driven (shallow and deep machine learning), domain-driven (epidemic, medical and biomedical analyses) modeling in Sections 4, 5 and 6, respectively. Section 7 further discusses the modeling on the influence and impact of COVID-19, Section 8 reviews the work on COVID-19 simulations, and the related work on COVID-19 hybrid modeling is reviewed in Section 9. Lastly, Section 10 further discusses the significant gaps and opportunities in modeling COVID-19.

This review aims to be *specific* to COVID-19 modeling so pure domain-specific research on its medicine, vaccine, biology and pathology is excluded; more *comprehensive* than the other references to cover problems and techniques from classic to present AI, data science and beyond; *unique* in summarizing the challenges of the COVID-19 disease, data and modeling; *structural* and *critical* by categorizing, comparing, criticizing and generalizing typical modeling methods tailored for COVID-19 modeling from various disciplines and areas; and *insightful* by extracting conclusive and contrastive (to other epidemics) findings about the virus and disease from the references. The review incorporates much discussion on the topics, opportunities and directions to tackle those issues which are rarely or poorly addressed or areas which remain open in the broad research landscape of modeling COVID-19.

However, this review also presents the various limitations and opportunities for further work. (1) As the scope and capacity of the review is limited, we do not cover the domain-specific literature on pure medical, biomedical and social science-oriented topics and methods without involving modeling methods. (2) There are over 10k references closely relevant to modeling COVID-19 and numerous specific modeling techniques from various disciplines and areas [38], which could not be fully covered or highlighted in detail in this review. (3) As discussed in the above, different from the narrowly-focused review papers in the area which highlight specific techniques and their relevant references, we only present those mostly used (useful) modeling techniques by summarizing their generalizable formulations. (4) This review does not answer many important questions concerning modelers, governments, policy-makers and domain experts, e.g., what has the modeling told us about the nature of COVID-19, which could be further highlighted in more purposeful reviews and analyses. (5) There are many challenging problems yet to be informed or addressed by the modeling progress, as discussed in Section 10. (6) There are increasingly more and newer references including preprints emerging online every day, which poses significant challenges for us to cover all up-to-date important references on modeling COVID-19.

2 COVID-19 CHARACTERISTICS AND COMPLEXITIES

In this section, we summarize the main characteristics and challenges of the COVID-19 disease, the data and the modeling, which are connected to the various modeling tasks and methods reviewed in this paper.

2.1 COVID-19 Disease Characteristics

Modeling COVID-19 is highly challenging because its sophisticated epidemiological, clinical and pathological characteristics are poorly understood [93, 96, 173]. Despite common epidemic clinical symptoms like fever and cough [105], SARS-CoV-2 and COVID-19 have many other sophisticated characteristics [167] that make them more mysterious, contagious and challenging for quantification, modeling and containment. We highlight a few of these below.

High contagiousness and rapid spread. The high contagiousness of SARS-CoV-2 is one of the most important factors driving the COVID-19 pandemic. In epidemiology, the *reproduction number* R_0 denotes the transmission ability of an epidemic or endemic. It is the expected

number of cases directly generated by one case in a population where all individuals are susceptible to infection [79]. If $R_0 > 1$, the epidemic will begin to transmit rapidly in the population, while $R_0 < 1$ indicates that the epidemic will gradually vanish and will not lead to a large-scale outbreak. Different computational methods have resulted in varying reproduction values of COVID-19 in different regions. For example, Sanche et al. [201] report a median R_0 value of 5.7 with a 95% confidence interval (CI) [3.8, 8.9] during the early stages of the epidemic in Wuhan China. Gatto et al. [80] estimate a generalized reproduction value of 3.60 (95% CI: 3.49 to 3.84) using the susceptible-exposed-infected-recovered (SEIR)-like transmission model in Italy. de Souza et al. [62] report a value of 3.1 (95% CI: 2.4 to 5.5) in Brazil. The review finds that the R_0 of COVID-19 may be larger than 3.0 in the initial stage, higher than that of SARS (1.7-1.9) and MERS (< 1) [180]. It is generally agreed that SARS-CoV-2 is more transmissible than severe acute respiratory syndrome conronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) although SARS-CoV-2 shares 79% of the genomic sequence identity with SARS-CoV and 50% with MERS-CoV [70, 96, 138, 179].

A varying incubation period. The *incubation period* of COVID-19, also known as the pre-symptomatic period, refers to the time from becoming infected by exposure to the virus and symptom onset. A median incubation period of approximately 5 days was reported in [124] for COVID-19, which is similar to SARS. In [173], the mean incubation period was found to range from 4 to 6 days, which is comparable to SARS (4.4 days) and MERS (5.5 days). Although an average incubation period of 5-6 days is reported in the literature, the actual incubation period may be as long as 14 days [124, 167, 264]. The widely varying COVID-19 incubation period and its uncertain value in a specific hotspot make case identification and infection control very difficult. Unlike SARS and MERS, COVID-19 infected individuals are already contagious during their incubation periods. As it is likely that they are unaware that they are infected and have no-to-mild symptoms during this period, they may easily be the unknown sources of widespread transmission. This has informed screening and control policies, e.g., mandatory 14-day quarantine and isolation, corresponding to the longest predicted incubation time.

A large number of asymptomatic and undocumented infections. It is clear that COVID-19 has a broad clinical spectrum which includes asymptomatic and mild illness [42, 128, 179]. However, the accurate number of asymptomatic and mild-symptomatic infections of both original and new-generations of viruses remains unknown. Asymptomatic infections may not be screened and diagnosed before symptom onset, leading to a large number of undocumented infections and the potential risk of contact with infected individuals [120]. The review in [30] reports that of those who tested positive in studies which were conducted in seven countries, the proportion who were asymptomatic ranged from 6% to 41%, while the study in [249] reports that 23% of those infected by COVID were asymptomatic. Buitrago-Garcia et al. [28] found that most people who are infected with COVID-19 do not remain asymptomatic throughout the course of infection, and only 20% of infections remain asymptomatic during follow-up, however this estimate requires further verification and study. Ravindra et al. [195] analyzed the possibility of different levels of asymptomatic transmission in the community and concluded that asymptomatic human transmission is relevant to the varying incubation periods between people and about 31% of all populations are asymptomatic, including familial clusters, adults, children, health care workers, and travelers. The study in [130] shows that a large percentage (86%) of infections are undocumented, about 80% of documented cases are due to transmission from undocumented cases, and the transmission rate of undocumented infections is about 55% of that of documented cases.

High mutation with mysterious strains and high contagion. The four major SARS-CoV-2 variants of concern such as B.1.1.7 (labeled Alpha by WHO) and B.1.351 (Beta) variants have higher transmissibility (B.1.1.7 has approximately 50% increased transmission) [186] and reproduction rate (B.1.1.7 has an increase reproduction rate of 1-1.4) [235], challenging existing vaccines, containment and mitigation methods. The recently identified variant B.1.617.2 (Delta) in India has even more sophisticated transmissibility and infectious characteristics. The identified variants of concern generally have increased transmissibility (20-50%), increased detrimental change of epidemiology, more severe virulence and disease presentation (e.g., increased hospitalizations or deaths), and result in the decreased effectiveness of public health and social measures, reduced effectiveness of available diagnostics, vaccines and therapeutics, increased diagnostic detection failures, and reduced neutralization by antibodies generated during previous infection or vaccination [229, 251].

Discussion. While the above summarizes the most recent understanding of SARS-CoV-2 and COVID-19 complexities, it is also noted that knowledge on the nature of the virus and its mutation is limited. Without knowing its origins, there is much misinformation about the virus, its contagion and the interventions required [199]. There is weak to no ground truth about the reality of its infection, symptoms shown in medical imaging, and mitigation and treatment measures. There have been no joint global pathological, epidemiological, biomedical and socioeconomic studies which provide a deep and systematic understanding of the COVID-19 virus and disease complexities, common knowledge, and ground truth.

2.2 COVID-19 Data Challenges

COVID-19 involves multisource, small, sparse and quality-inconsistent data [122][3]. Typical data sources and factors include (1) epidemiological factors (e.g., origin, incubation period, transmission rate, mortality, morbidity, and high to least vulnerable population, etc.); (2) daily new-infected-recovered-death case numbers, their reporting time and region of occurrence; (3) quarantine and mitigation measures and policies (e.g., social distancing and border control) relating to communities and individuals; (4) clinical, pathological and genomic data (e.g., symptoms, medical facilities, hospitalization records, medical history, medical imaging, pharmaceutical treatments, gene and protein sequences); (5) infective demographics (e.g., age, gender, race, cultural background, and habit); (6) social activities and mobility; (7) domain knowledge and precautionary guidance from authorities on the virus and disease; (8) seasonal and environmental factors (e.g., season, geographical location, temperature, humidity, and wind speed); (9) news, reports and social media discussions on coronavirus and COVID-19; and (10) fake news, rumor and misinformation. Such COVID-19 data are heterogeneously coded in character, text, number or image; in unordered, temporal/sequential or spatial modes; in static and dynamic forms; and with the characteristics as follows.

Despite the large volume of existing research, modeling COVID-19 is still in an early stage with many open issues, partially because of the significant complexities of COVID-19 data. The main characteristics of COVID-19 data are summarized below and impose a computational burden on the modeling of COVID-19.

Acyclic and short-range case numbers which are small in size. The publicly available data for COVID-19 modeling is limited. Except in rare scenarios such as in the US, most countries and regions report a short-range (2-3 months or even shorter such as local hotspot-based outbreaks), low-granularity (typically daily), and small-size (daily case numbers for a short period of time and a small cluster of the population) record of COVID-19 data. Such data

---

[3]COVID-19 modeling: https://datasciences.org/covid19-modeling/

is typically acyclic without obvious seasonal or periodical patterns as in influenza [56] and recurrent dengue epidemics in tropical countries [228].

Inaccurate statistics on COVID-19 cases. The reported new-infected-recovered-death case numbers are estimated much lower than their real number in most countries and regions. This may be due to many reasons, such as pre-symptomatic and asymptomatic infections, limited testing capability, nonstandard manual recording, different confirmation standards, an evolving understanding of the disease nature, and other subjective factors. The method of calculating case statistics may vary significantly from country to country; the actual figures in some countries and regions may even be unknown; and no clear differentiation is made between hotspot and country/region-based case reporting. The gaps between the infection reality and what has been documented may be more apparent in the first wave, in the early stage of outbreaks, and in some countries and regions [141]. As result, the actual infected number and infected regions of COVID-19 pandemic may be much bigger than those publicly reported.

Lack of reliable data particularly in an initial outbreak. The spread of an epidemic in its initial phase can be regarded as transmission under perfect conditions. In its initial phase, the intrinsic epidemiological characteristics of COVID-19, such as reproduction rate, transmission rate, recovery rate and mortality are closer to their true values. For example, the modeling results in [198] show a wide range of variations due to the lack of reliable data, especially at the beginning of an outbreak.

Lack of high-quality microlevel data. Data on COVID-19 cases, including daily infected cases, daily new cases, daily recovery cases, and daily death cases, is collected on a daily basis, while daily susceptible case numbers were also reported in Wuhan. However, macrolevel and low-dimensional data is far from comprehensive for inferring the complex transmission processes accurately and more fine-grained data with various aspects of features and high dimensions are needed. For example, during the initial phase of the Wuhan outbreak, the dissemination of SARS-CoV-2 was primarily determined by human mobility in Wuhan, however no empirical evidence on the effect of key geographic factors on local epidemic transmission was available [191]. The risk of COVID-19 death varies across various sociodemographic characteristics [66], including age, sex, civil status, individual disposable income, region of residence, and country of birth. More specific data is required to address the sociodemographic inequalities related to contracting the COVID-19 virus. To contain the spread of COVID-19, governments propose and initiate a series of similar to different NPIs. No quantitative evidence or systematic evaluation analyzes how these measures affect epidemic transmission, leading to challenges in inferring NPI-based COVID-19 transmission and mitigation.

Data incompleteness, inconsistencies, inequality and incomparability. Typically, it is difficult to find all-round information about a COVID-19 patient's infection source, demographics, behaviors, social activities (including mobility and in social media), clinical history, diagnoses and treatments, and resurgence if any. COVID-19 public data also presents strong inconsistencies and inequalities across reporting hotspots, countries and regions, updating frequencies and timelinesses, case confirmation standards, collection methods, and stages [198]. Data from different countries and areas may be unequal and incomparable due to their non-unified statistical criteria, confirmation standards, sampling and coverage methods, and health and medical conditions and protocols. These are also related to or affected by a person's race, living habits, and their applied mitigation policies, etc.

Other issues. Comparing public data available from different sources also reveals other issues like potential noise, bias and manipulation in some of the reported case numbers

(e.g., due to their nonuniform statistic standards or manual statistical mistakes), missing values (e.g., unreported on weekends and in the early stage of outbreaks), different categorization of cases and stages (e.g., some with susceptible and asymptomatic case numbers), misinformation, and lack of information and knowledge about resurgence and mutation.

Discussion. While increasing amounts of COVID-19 data are publicly available, they are in fact poor and limited in terms of quality, quantity, capability and capacity to discover deep insights about the nature of COVID-19, its interaction with external factors, and its effects. It is fundamental and urgent to acquire substantially larger and better-quality multisource COVID-19 data. This is crucial so that meaningful modeling can be robustly conducted and evaluated to reveal intrinsic knowledge and insights about the disease and to assist effective pandemic control.

2.3 COVID-19 Modeling Complexities

The COVID-19 pandemic is essentially an open complex system with significant system complexities [35, 247]. Examples are the hidden nature and strong uncertainty, self-organization, dynamics and evolution of the virus, disease and their developments and transmissions; their sophisticated interactions and relations to environments and context; the differentiated virus infections of individuals and communities; and the significant emergence of consequences and impacts on society in almost every part of the world. However, the publicly available small and limited COVID-19 data does not explicitly display a complete picture and a sufficient indication of the above complexities and intrinsic epidemiological attributes, transmission process, and cause-effect relations. It is thus challenging to undertake sound, robust, benchmarkable and generally useful modeling on such potential-limited data.

Achieving ambitious modeling objectives on low-quality small COVID-19 data. As discussed in Section 3.1, many business problems and objectives are expected to be addressed by modeling COVID-19. However, the strong constraints in COVID-19 public data discussed in Section 2.2 significantly limit this potential. Modelers have to carefully define learnable objectives, i.e., what can be learned from the data, acquire the essential and feasible data, or leverage data poverty by more powerful modeling approaches. For example, when a model is trained on a country's case numbers, its application to other countries may produce unfair results owing to their data inequalities. Another example is how to combine multisource but weakly connected data for meaningful high-potential analysis and results.

Undertaking complex modeling with limited to no domain knowledge and ground truth. The weak to no-firm knowledge and ground truth about COVID-19 and its medical confirmation and annotations and poor-quality data limit the capacity and richness of the hypotheses to be tested and modeled on the data. It is not surprising that rather simple and classic analytical and learning models are predominantly applied by medical and biological scientists to verify specific hypotheses, e.g., various SIR models, time-series regression, and traditional machine learning methods [49, 81, 207], which occupy the top-80 keywords-based methods in the 200k WHO-collected references. In contrast, statisticians and computer scientists tend to enforce overparameterized models, over-complicated hypotheses, or over-manipulated data, resulting in highly specific results and over- or under-fitting issues.

Challenges in addressing the COVID-19 disease and data complexities. The unique characteristics and complexities of the COVID-19 disease and data discussed in Sections 2.1 and 2.2 challenge the existing modeling methods including deep neural learning. Examples are generalized modeling of quality and quantity-limited COVID-19 data from different countries and regions and over evolving time periods [232], robust modeling of short-range, small-size

and incomplete-cycle data, and high-capacity modeling of mixture distributions with exponential growth [63], sub-exponential growth [141], discontinuous phase transition [232] and instant changes in case developments.

Disclosing complicated relations and interactions in weakly-coupled multisource data. COVID-19 is affiliated with many personal, social, health/medical, political and other factors, dispersedly reflected in explicitly or implicitly related multisource systems. The COVID-19 pandemic is formed and evolves as a dynamic social-technical process and the co-effects of multi-factor interplay. These multi-aspect factors are coupled strongly or weakly, locally or globally, explicitly or implicitly, subjectively or objectively, statically or dynamically, and essentially or accidentally in the virus and disease formation, development, influence, and evolution. Disclosing such sophisticated factor couplings and interactions is significantly challenging as they are not obvious or easily verifiable in observations. Therefore, modeling COVID-19 requires in-depth transdisciplinary cooperation between computer science, bioinformatics, virology, sociology and many other disciplines. A single factor alone cannot disclose the intrinsic and intricate nature of COVID-19 or explain the variability or shape the dynamics of this epidemic.

Discussion. The COVID-19 complexities result in significant modeling challenges, resulting from the data, the unclear epidemic transmission mechanisms and processes, and the entanglement between epidemic factors/observations and external objective (e.g. countermeasures) and subjective (e.g. people behavior changes) factors. COVID-19 modeling goes beyond the transformation and applications of powerful models such as overparameterized deep neural networks, SIR variants and hierarchical Bayesian networks on the highly limited and poorly coupled small COVID-19 data. Careful designs are necessary to address the specific COVID-19 characteristics and complexities of its data and disease, avoid under-/over-fitting, and focus on modeling the complexities in relation to their underlying nature and insight. Complicating models does not necessarily contribute to better or more actionable knowledge and intelligence about the COVID-19 disease and data [22, 34, 223].

3 COVID-19 MODELING LANDSCAPE

To address the aforementioned COVID-19 disease, problems, data and modeling challenges, we present a high-level landscape to categorize and connect the comprehensive objectives and techniques for modeling COVID-19.

3.1 Objectives of COVID-19 Modeling

Here, we summarize the main business problems and objectives in modeling COVID-19. The analysis of the WHO-collected literature [38] gives us a clear indication of the top business terms in over 200k references and 22k modeling-focused references. The top-ranked keywords include COVID-19 and coronavirus pandemic outbreak, spread, infection, transmission, factors, symptoms, characteristics, treatment, diagnosis, mortality, their risk and effects, as well as major data analysis and domain-specific research areas and methods. Below, we consolidate the main concerns and objectives of modeling COVID-19. Table 1 summarizes the associated modeling factors, modeling methods, and references.

Characterizing and predicting the COVID-19 epidemic dynamics and transmission. An imperative challenge is to understand the COVID-19 epidemic mechanisms, transmission process and dynamics, infer its epidemiological attributes, and understand how the virus spreads spatially and socially [1]. The majority of COVID-19 modeling tasks focus on exploring the source and spectrum of the COVID-19 infection, clinical and epidemiological characteristics, tracking its transmission routes, and forecasting case development trends

Table 1. Business Problems and Objectives of COVID-19 Modeling.

Objectives	Modeling factors	Approaches	References
Epidemic dynamics and transmission	Epidemiological factors (e.g., origins, incubation period, transmission rate, morbidity, mortality, and highly to least vulnerable population, etc.), daily new-infected-recovered-death case numbers and reporting time, side information about population, etc.	Regression, compartmental models, time/age-dependent compartmental models, probabilistic compartmental models, etc.	[5, 49, 57, 81, 177, 212, 250, 275]
Non-pharmaceutical intervention and policies	Intervention measures, quarantine and mitigation measures and policies on communities and individuals, epidemiological factors, daily new-infected-recovered-death case numbers and reporting time, social activities, communications	Regression, customized compartmental models, Bayesian hierarchical models, stochastic compartmental models, etc.	[5, 24, 63, 77, 81, 121, 141, 185, 221]
Diagnosis, identification and tracing	Clinical, pathological and genomic data (symptoms, medical facilities, hospitalization records, medical history, respiratory signals, medical imaging, physical and chemical measures, gene sequences, proteins), mobility and contacts, etc.	Regression, statistical learning, shallow (e.g. decision trees, random forest) and deep models, image and signal processing methods, etc.	[11, 17, 51, 54, 61, 112, 143, 149, 152, 240, 253, 268]
Treatment and pharmaceutical interventions	Clinical measures, hospitalization records, medical test records, drug selection, pharmaceutical treatments, ICU records, ventilator use, healthcare records, etc.	Classifiers, time-series methods, DNNs, etc.	[18, 133, 254, 260, 274]
Pathological and biomedical analysis	Genomic data, protein structures, pathogenic data, drug and vaccine info, immunization response, etc.	Classifiers, outlier detectors, genome analysis, protein analysis, DNNs, etc.	[7, 18, 111, 139, 169, 193, 205, 237, 256, 270]
Resurgence and mutation	Daily new-infected-recovered-death case numbers and reporting time, quarantine and mitigation measures and policies on communities and individuals, social activities and mobility, seasonal and environmental factors (e.g., season, geographical location, temperature, humidity, and wind speed)	Compartmental models, simulation models, compartmental models combined with regression, epidemic renormalisation group, etc.	[12, 31, 127, 137, 174]
Influence and impact	Quarantine and mitigation measures and policies on communities and individuals, domain knowledge and precaution guidance from authorities, social activities and mobility, demographics (e.g., age, gender, racist, cultural background, habit), related news, reports, social media discussions, and misinformation.	Statistical analysis, questionnaire methods, age-structured SIR/SEIR models, deep neural networks (e.g., BERT and LSTM), simulation models, etc.	[40, 119, 131, 208, 236]

and the peak number of infected cases and disease transmission [42, 254, 255]. They aim for

findings to understand the nature of the virus and disease and inform disease precaution, virus containment, mitigation campaigns, and medical resource planning, etc.

Modeling the resurgence and mutation. As the SARS-CoV-2 mutation and COVID-19 resurgence are highly uncertain and much more transmissible and infectious, we highlight their relevant research here. However, as our current understanding of the resurgence and mutation is very limited, COVID-19 may become another epidemic disease which stays with humans for a long time. The WHO-identified four variants of concern which have higher transmissibility, contagion and complexities [85, 86, 229, 251]. Imperative research is expected to quantify the resurgence conditions, control potential resurgences after lifting certain restrictions and reactivating businesses and activities [137, 174], distinguish the characteristics and containment measures between waves [8, 72, 84], and prepare for and predict resurgence, mutation and their responsive countermeasures [12].

Disease diagnosis, infection identification and contact tracing. Given the strong transmission and reproduction rates, high contagion, and sophisticated transmission routes and the unexpected resurgence of COVID-19 and its virus mutation, it is crucial to immediately identify and confirm exposed cases and trace their origins and contacts to proactively implement quarantine measures and contain their potential spread and outbreak [159]. This is particularly important during the varying incubation periods (usually 2 - 14 days) which are asymptomatic to mildly symptomatic yet highly contagious. In addition to chemical and clinical approaches, identifying COVID-19 by analyzing biomedical images, genomic sequences, symptoms, social activities, mobility and media communications is also essential [226].

Modeling the efficacy of medical treatment and pharmaceutical interventions. The general practices of timely and proper COVID-19 medical treatments, drug selection and pharmaceutical measures, and ICU and ventilation etc. play fundamental roles in fast recovery, mitigating severe symptoms and reducing the mortality rate of both the original and increasingly-mutated virus strains. However, the lack of best practices and standardized protocols and specifications of medical and pharmaceutical treatments on the respective virus variants in terms of patient's demographic and ethnic context and the wide dispersal of online misinformation of drug use may also contribute to global imbalance in containing COVID-19. Research is required to select and discover suitable drugs, which best match the patient's diagnosis and ethnic contexts with suitable medical treatments to mitigate critical conditions and mortality in a timely manner, etc. [18, 254, 260, 274].

Modeling the efficacy of non-pharmaceutical intervention and policies. Various NPIs, such as travel bans, border control, business and school shutdowns, public and private gathering restrictions, mask-wearing, and social distancing are often implemented to control the outbreak of COVID-19. Different governments tend to enforce them in varied combinations and levels and ease them within different timeframes and following different procedures, resulting in different outcomes. Limited research results have been reported to verify the effects of these measures and their combinations on containing the virus spread and case number development, the balance between enforcement levels and containment results, and the response sensitivity of the restrictions in relation to the population's ethnic context [24, 63, 221]. Limited results are available on the threshold and effects of COVID-19 vaccinations and herd immunity. More robust results will inform medical and public health policy-making on medication, business and society.

Understanding pathology and biomedical attributes for drug and vaccine development. By involving domain knowledge and techniques such as virology, pathogenesis, genomics and proteomics, pathological and biomedical analyses can be conducted on pathological test

results, gene sequences, protein sequences, physical and chemical properties of SARS-CoV-2, drug and vaccine information and their effects. Accordingly, it is necessary to conduct domain-driven analysis and model correlated drugs and vaccines with genomic and protein structures to select and develop COVID-19 drug and vaccines, to understand the drug-target interactions, and to diagnose and identify infection, understand virus mutation, etc. [18, 97, 193]. More research is needed on COVID-19 immunity responses, drug and vaccine development, and mutation intervention.

Modeling COVID-19 influence and impact. While the COVID-19 pandemic has changed the world and has had a significant and overwhelming influence on almost all aspects of life, society and the economy, quantifying its influence and impact has rarely been studied. COVID-19 negative impact modeling may include (1) economic impact on growth and restructuring [252]; (2) social impact on people's stress, psychology, emotions, behavior and mobility [175, 258]; and (3) transforming business processes and organizations, manufacturing, transport, logistics, and globalization [204, 234]. In contrast, it would also be interesting to model its 'opportunity' and influence on (1) enhancing the wellbeing and resilience of individuals, families, society and work-life balance [187]; (2) digitizing and transforming work, study, entertainment and shopping [215]; (3) restructuring supply-demand relations and supply chains for better immediate availability and to satisfy demand [64]; (4) promoting research and innovation on intervening in global black-swan disasters like COVID-19 and its impacts [269]; and (5) enhancing trust and development in science, medicine, vaccination and hygiene [182]. Other impact modeling tasks include analyzing the relations between the COVID-19 containment effect and socioeconomic level (e.g., income level particularly in relation to lower-income and disadvantaged groups), healthcare capacity and quality, government crisis management capabilities, citizen-government-cooperation, and public health and hygiene habits.

3.2 Categorization of COVID-19 Modeling

The flow of COVID-19 modeling has strong features such as: (1) multi-disciplinary techniques of mathematics and statistics, epidemiology, broad AI and data science including shallow and deep learning, and social science; (2) epidemiological methods to explore business problems and research areas; (3) domain, model and data-driven approaches consisting of various families of domain knowledge, models and methods which are widely applied in all business problems; (4) case studies and hypothesis tests to highlight the results of particular methods or modeling specific settings, scenarios or data.

The keyword-based analysis of the 22k WHO-collected modeling references in [38] shows that epidemiological modeling, mathematical and statistical modeling, artificial intelligence and data science, and simulation modeling play predominant roles in understanding, characterizing, simulating, analyzing and predicting COVID-19 issues. We thus categorize the research landscape of COVID-19 modeling into six: domain-driven modeling, mathematical/statistical modeling, data-driven learning, influence/impact modeling, simulation modeling, and hybrid methods in this review. Fig. 1 summarizes the transdisciplinary research landscape connecting the aforementioned six categories of modeling techniques and their respective modeling methods to major COVID-19 business problems and their modeling objectives in Section 3.1.

- *COVID-19 mathematical/statistical modeling*: developing and applying mathematical and statistical models such as time-series analysis (e.g., regression models and hazard and survival functions) and statistical models (e.g., descriptive analytics, statistical
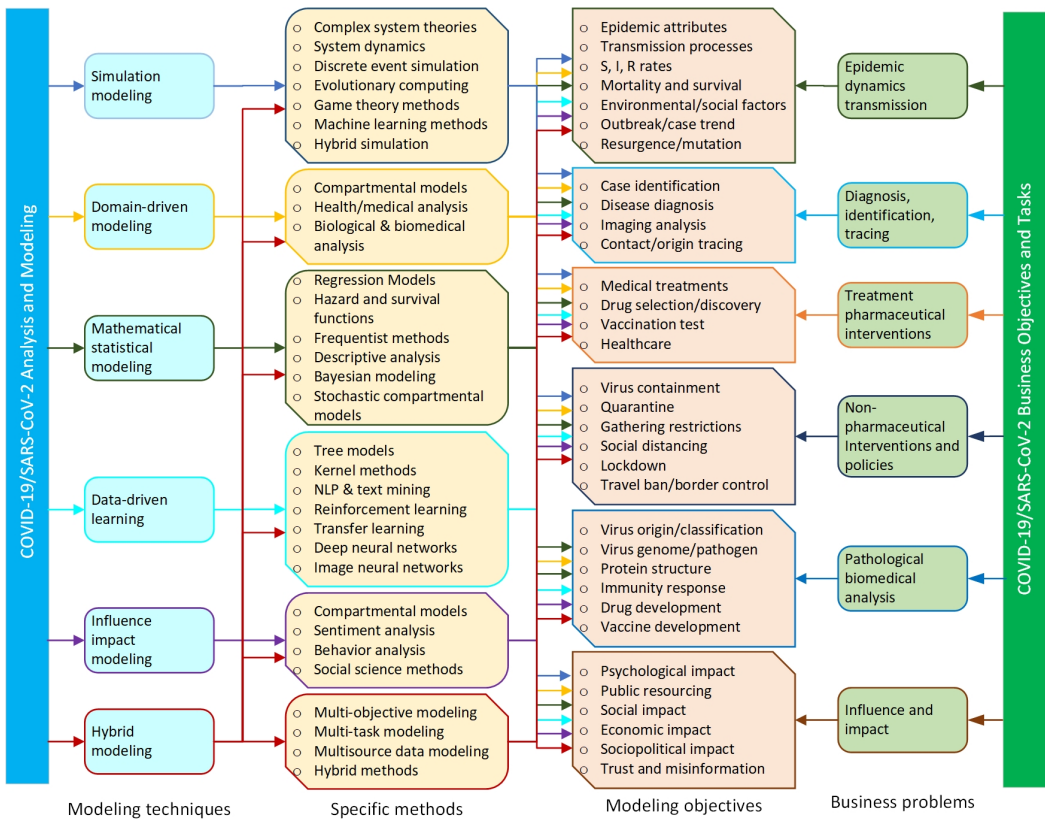
Fig. 1. The transdisciplinary research landscape of COVID-19 modeling: From right: business problems and modeling objectives to the left: modeling techniques and methods.

processes, latent factor models, temporal hierarchical Bayesian models, and stochastic compartmental models) to estimate COVID-19 transmission processes, symptom identification, disease diagnosis and treatment, sentiment analysis, misinformation analysis, and resurgence and mutation.

- *COVID-19 data-driven learning*: developing and applying data-driven classic (e.g., tree models such as random forests and decision trees, kernel methods such as support vector machines (SVMs), NLP and text analysis, and classic reinforcement learning) and deep (e.g., deep neural networks, transfer learning, deep reinforcement learning, and variational deep neural models) analytics and learning methods on COVID-19 data to characterize, represent, classify, and predict COVID-19 problems, such as case development, mortality and survival forecasting, medical imaging analysis, NPI effect estimation, and genomic analysis.

- *COVID-19 domain-driven modeling*: developing and applying domain knowledge and domain-specific models for COVID-19; examples are epidemiological compartmental models to characterize the COVID-19 epidemic transmission processes, dynamics, transmission and risk, and the influence of external factors on COVID-19 epidemics, resurgence and mutation; and medical, pathological and biomedical analysis for infection

diagnosis, case identification, patient risk and prognosis analysis, medical imaging-based diagnosis, pathological and treatment analysis, and drug development.
- *COVID-19 influence/impact modeling*: developing and applying methods to estimate and forecast the influence and impact of SARS-COV-2 variations and COVID-19 diseases and their interventions, treatments and vaccination on epidemic transmission dynamics, virus containment, disease treatment, public resources including healthcare systems, social systems, economy, and human psychological health and behaviors.
- *COVID-19 simulation modeling*: developing and applying simulation models such as theories of complex systems, agent-based simulation, discrete event analysis, evolutionary learning, game theories, and Monte-Carlo simulation to simulate the COVID-19 epidemic evolution and the effect of interventions and policies on the COVID-19 epidemic.
- *COVID-19 hybrid modeling*: hybridizing and ensembling multiple models to tackle multiple business problems and objectives, multiple tasks, and multisource data and those individual objectives, tasks and data sources that cannot be better understood by single approaches.

It is worth mentioning that each of the above modeling techniques and their specific methods may be applicable to address different business problems and modeling objectives, as shown in Fig. 1. Below, we review the progress of the above six categories of COVID-19 modeling by (1) summarizing the typical modeling techniques and (2) categorizing their typical applications in modeling diverse COVID-19 issues.

4 COVID-19 MATHEMATICAL MODELING

Mathematical and statistical models are overwhelmingly used to estimate and predict the transmission dynamics and reveal the truth of epidemic in a formalized and quantitative way. Accurate COVID-19 mathematical models are indispensable for the COVID-19 epidemic forecasting and decision making, amounting to 13k of 22k references on modeling COVID-19. Here, we review two sets of main mathematical methods: time-series analysis and statistical modeling, and their applications in COVID-19 modeling.

4.1 Time-series Analysis

4.1.1 Time-series modeling. We here focus on two typical methods that are predominantly customized for modeling COVID-19: regression models and hazard and survival functions.

Regression models. Typical regression models such as logistic regression and auto-regressive integrated moving average (ARIMA) variants are widely used in epidemic and COVID-19 modeling. *Logistic growth models* can model the number of COVID-19 infected cases. For a population of S with the infection rate β, the growth scale of infection number I can be modeled by

$$\partial_t I = \beta I 1 - IS \tag{1}$$

over time t. Accordingly, with historical COVID-19 cases of a place at a time period, an S-shaped curve can be derived to describe and forecast the growth distributions of the COVID-19 infections and the peak infected number by adjusting the constant rate β. Logistic models are weak to incapable of modeling other states of COVID-19 cases and there are many challenges, such as nonstationary characteristics discussed in Section 2.2, while some may be better modeled by more sophisticated regression models.

ARIMA and its variants also model the temporal movement of COVID-19 case numbers with more flexibility than logistic ones. For example, the number I of infected cases can be

modeled by $ARIMAp, d, q$ which factorizes the number into consecutive past numbers with errors:

$$I_t = \alpha_1 I_{t-1} \ \alpha_2 I_{t-2} \ \cdots \ \alpha_p I_{t-p} \ \varepsilon_t \ \theta_{t-1} \varepsilon_{t-1} \ \cdots \ \theta_{t-q} \varepsilon_{t-q} \ c \qquad (2)$$

over the number of time lags (order) of autoregression p, the order of moving average q, the degree of differencing d, and time t with a constant c. I_{t-p} refers to the infected cases at time $t - p$ with weight α_p, ε_{t-q} refers to the error between I_{t-q} and I_{t-q-1} with weight θ_{t-q}. Adjusting parameters like p, d and q can simulate/capture some of the time series characteristics (e.g., the process and trend of the infection series by p and q, seasonality by d, and volatile movement by the distribution of error terms). Similarly, ARIMA models can be used to simulate and forecast the number of recoveries and deaths in relation to COVID-19.

In addition, ARIMA and its variants can be integrated with other modeling methods to characterize other aspects of COVID-19 time series. For example, the wavelet decomposition of frequency-based nonstationary factors can model the oscillatory error terms of ARIMA-based modeling of COVID-19 infected cases [40]). Another example is to combine the decision tree method with regression to form a regression tree and identify mortality-sensitive COVID-19 factors [41].

Hazard and survival functions. Hazard functions and survival functions are often used to model the mortality and survival (recovery) rates of patients using time-to-event analysis. A hazard function models the mortality probability $ht|\mathbf{x}$ of a COVID-19 patient with the factor vector $\mathbf{x}$ ($\in \mathcal{R}^d$) of dying at discrete time t:

$$ht|\mathbf{x} = pT = t|T \geq t; \mathbf{x} \qquad (3)$$

On the contrary, the survival function models the probability $St|\mathbf{x}$ of surviving until time t:

$$St|\mathbf{x} = PT > t|\mathbf{x} \qquad (4)$$

In discrete time, $St = {}^t_{t_i=1} 1 - h_{t_i}$ where h_{t_i} is the mortality probability at time t_i. In continuous time, $St|\mathbf{x} = 1 - Ft|\mathbf{x}$ where $F\cdot$ is the cumulative distribution function until time t. For covariates $\mathbf{x}$ with their relations represented by an either linear or nonlinear function $f\cdot; \delta$ with parameters δ, the mortality rate of COVID-19 can be modeled by a Cox proportional hazard model [190]:

$$ht|\mathbf{x} = h_0 texpf\cdot; \delta \qquad (5)$$

where $h_0 t$ is the baseline hazard function, the function $f\cdot; \delta$ can be implemented by a linear function such as a linear transform or a nonlinear function such as a deep convolutional network. For example, in [203], $f\cdot; \delta$ is implemented by a shallow neural network with a leaky rectified linear unit-based activation of the input and then another tangent transformation. In the case of time-varying covariates $\mathbf{x}_t$, the above hazard, survival and transform functions should be time sensitive as well.

Typically, the performance of mathematical modeling is measured by metrics such as mean absolute error (MAE), root mean square error (RMSE), the improvement percentage index (IP), and symmetric mean absolute percentage error (sMAPE) in terms of certain levels of confidence intervals.

4.1.2 COVID-19 time-series modeling. Time-series analysis contributes the most (about 3k of the 22k WHO-listed references) to COVID-19 modeling. As shown in [38], regression models, linear regression, and logistic regression are mostly applied in COVID-19 modeling. Many linear and nonlinear, univariate, bivariate and multivariate analysis methods have been intensively applied for the regression and trend forecasting of new, susceptible, infectious, recovered and death case numbers. Popular methods include linear regression models such

as ARIMA and GARCH [213, 217], logistic growth regression [243], COX regression [203], multivariate and polynomial regression [6, 46, 89], generalized linear model and visual analysis [153], support vector regression (SVR) [87] [196], regression trees [41], hazard and survival functions [203], and more modern LSTM networks. In addition, temporal interpolation methods such as best fit cubic, exponential decay and Lagrange interpolation, spatial interpolation methods such as inverse distance weighting, smoothing methods such as moving average, and spatio-temporal interpolation [33] are applicable to fit and forecast COVID-19 case time series. We illustrate a few tasks below: COVID-19 epidemic distributions, case number and trend forecasting, and COVID-19 factor and risk analysis.

COVID-19 case number and trend forecasting and epidemic distributions. Regression-centered time series analysis has been widely applied to forecast case number developments and trends. For COVID-19 prediction, Singh et al. [213] apply ARIMA to predict the COVID-19 spread trajectories for the top 15 countries with confirmed cases and conclude that ARIMA with a weight to adjust the past case numbers and the errors has the ability to correct model prediction and is better than regression and exponential models for prediction. However, ARIMA lacks flexible support for volatility and in-between changes during the prediction periods [213]. Gupal et al. [89] adopt polynomial regression to predict the number of confirmed cases in India. Almeshal et al. [9] utilize logistic growth regression to fit the actual infected cases and the growth of infections per day. Wang et al. [243] model the cap value of the epidemic trend of COVID-19 case data using a logistic model. With the cap value, they derive the epidemic curve by adapting time series prediction. To find the best regression model for case forecasting, Ribeiro et al. [196] explore and compare the predictive capacity of the most widely-used regression models including ARIMA, cubist regression (CUBIST), random forest, ridge regression, SVR, and stacking-ensemble learning models. They conclude that SVR and stacking ensemble are the most suitable for the short-term COVID-19 case forecasting in Brazil. In addition, linear regression with Shannon diversity index and Lloyd's index are applied to analyze the relations between the meta-population crowdedness in city and rural areas and the epidemic length and attack rate [191].

COVID-19-specific factor and risk analysis. Time-series analysis may be used to (1) analyze the influence of specific and contextual factors on COVID-19 infections and COVID-19 epidemic developments including infection, transmission, outbreak, hospitalization, and recovery, e.g., on COVID-19 survival, mortality and recovery; and (2) analyze the influence and impact of external and contextual factors of COVID-19 outbreak on the population, health, society and the economy, case developments and containment. For example, to investigate the potential risk factors associated with fatal outcomes from COVID-19, Schwab et al. [203] present an early warning system assessing COVID-19 related mortality risk with a variation of the Cox proportional hazard regression model. Chen et al. [48] adapt the Cox regression model to analyze the clinical features and laboratory findings of hospitalized patients. Charkraborty et al. [41] design the wavelet transform optimal regression tree (RT) model, which combines various factors including case estimates, epidemiological characteristics and healthcare facilities to assess the risk of COVID-19. The advantage of RT is that it has a built-in variable selection mechanism from high dimensional variable space and can model arbitrary decision boundaries.

Correlation analysis between COVID-19 epidemic dynamics and external factors. Much research has been conducted on analyzing the relationships between COVID-19 transmission and dynamics and external and contextual factors. For example, Cox proportional hazard regression models are used to analyze high risk sociodemographic factors such as gender, individual income, education level and marital status that may be associated with a patient's

death [66], and logistic regression models are applied to analyze the relations between COVID-19 (or SARI with unknown aetiology) and socioeconomic status (per-capita income) [62]. To reveal the impact of meteorological factors, Chen et al. [46] examine the relationships between meteorological variables (i.e., temperature, humidity, wind speed and visibility) and the severity of the outbreak indicated by the confirmed case numbers using the polynomial regression method; while Liu et al. [134] fit the generalized linear models (GLM) with negative binomial distribution to estimate the city-specific effects of meteorological factors on confirmed case counts. In [183], Loess regression does not show an obvious relation between the COVID-19 reproduction number, weather factors (humidity and temperature) and human mobility. Lastly, linear models including linear regression, Lasso regression, ridge regression, elastic net, least angle regression, Lasso least angle regression, orthogonal matching pursuit, Bayesian ridge, automatic relevance determination, passive aggressive regressor, random sample consensus, TheilSen regressor and Huber regressor are applied to analyze the potential influence of weather conditions on the spread of coronavirus [142].

Discussion. Time-series methods excel at characterizing sequential transmission processes and temporal case movements and trends. They lack the capability to involve other multi-source factors and disclose deep insights into why case numbers evolve in a certain way and how to intervene in the infection, treatment and recovery.

4.2 Statistical Modeling

Statistical learning, in particular Bayesian models, play a critical role in stochastic epidemic and infectious disease modeling [23]. It takes generative stochastic processes to model epidemic contagion in epidemic modeling [10, 166]. In contrast to compartmental models, statistical models involve prior knowledge about an epidemic disease and their results have confidence levels corresponding to distinct assumptions (i.e., possible mitigation strategies), which better interpret and more flexibly model COVID-19 complexities. Below, we summarize typical statistical models and their applications in COVID-19 statistical modeling.

4.2.1 Statistical models. Statistical models are widely applied to COVID-19 modeling tasks including (1) simulating and validating the state distributions and transitions of COVID-19 infected individuals over time, (2) modeling latent and random factors affiliated with the COVID-19 epidemic processes, movements and interactions, (3) forecasting short-to-long-term transmission dynamics, (4) evaluating the effect of non-pharmaceutical interventions (NPI), and (5) estimating the impact of COVID-19 such as on socioeconomic aspects. Typical methods include descriptive analytics, Bayesian hierarchical models, probabilistic compartmental models, and probabilistic deep learning. Below, we introduce some common statistical settings and corresponding statistical models in both frequentist and Bayesian families that are often applied in COVID-19 statistical modeling.

Taking a stochastic (vs. deterministic) state transition assumption, various statistical processes can be assumed to simulate and estimate the state-specific counts (case numbers) and the probability of state transitions (e.g., between infections and deaths) during the COVID-19 spread. The stochastic processes and states (e.g., its infection and mortality) of a COVID-19 outbreak are influenced by various explicit and latent factors. Examples of explicit (observable) factors include a person's demographics (e.g., age and race), health conditions (e.g., disease history and hygienic conditions), social activities (e.g., working environment and social contacts), and the containment actions (e.g., quarantined or not) taken by the person. Latent factors may include the person's psychological attitude toward

cooperation (or conflict) with containment, health resilience strength to coronavirus and the containment influence on the outcome (e.g., infected or deceased).

Fig. 2(a) illustrates a general graphical model of the temporal hierarchical Bayesian modeling of COVID-19 case numbers for estimation and forecasting. The reported case number Y_t (e.g., death toll or infected cases) at time t can be estimated by $\tilde{y}_t$, which is inferred from the documented (declared) infections i_t^K and removed (e.g., recovered and deceased) rate κ_t. The documented infection number i_t^K is inferred from the infected population i_t and the test rate ρ_t. i_t is inferred from the exposed population e_t and the infection rate β_t, e_t is determined by its exposed rate ϵ_t. Further, we assume the removed rate κ_t is influenced by various medical treatments α, determined by auxiliary variables including socioeconomic condition λ_2, the treatment effectiveness ψ, and the public health quality ω. The infection rate β_t is determined by NPIs ζ, which are further influenced by the NPI execution rate τ and the socioeconomic factor λ_1. The priors of the corresponding parameters are a, b, c, d_1, d_2, f, g, h and j, which may follow specific assumptions.

For the statistical settings and hypotheses, typical statistical distributions of COVID-19 state-specific counts are applied to (1) *infection modeling*, e.g., by assuming a Bernoulli process (Bn, ε with the probability ε of exposure to infections over n contacts) and then a Poisson process at points of infections with exponentially-distributed infectious periods ($Pois\beta$ with the rate β referring to the infection rate within the infectious period); (2) *mortality modeling*, e.g., by assuming a negative binomial distribution ($NB\mu, \sigma$) or a Poisson distribution ($Pois\gamma$ with the rate γ parameterized on the mortality rate γ_D and population N). Further, the basic reproduction number R_0 may be estimated by $N\varepsilon\beta\beta \gamma$, the infections will be under control if R_0 is less than a given threshold (e.g., 1).
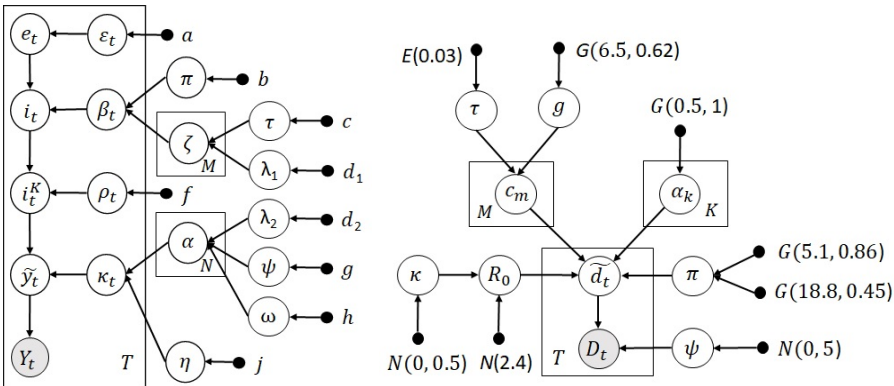
(a) A general graphical model of statis- (b) The graphic model for the method proposed in [77]
tical COVID-19 modeling

Fig. 2. A general and a specific hierarchical Bayesian model for COVID-19.

The standard deviation:	$\kappa \sim \text{Normal}0, 0.5$	(6a)
The initial reproduction number:	$R_{0,m} \sim \text{Normal}2.4, |\kappa|$	(6b)
The intervention impact:	$\alpha_k \sim \text{Gamma}0.5, 1$	(6c)
The time-varying reproduction number:	$R_{t,m} = R_{0,m}\exp-\sum_{k=1}^{6} \alpha_k I_{k,t,m}$	(6d)
The distribution rate:	$\tau \sim \text{Exponential}0.03$	(6e)
The 6 sequential days of infections:	$c_{1,m}, ..., c_{6,m} \sim \text{Exponential}\tau$	(6f)
The daily serial interval:	$g \sim \text{Gamma}6.5, 0.62$	(6g)
The number of infections:	$c_{t,m} = R_{t,m}\sum_{\tau=0}^{t-1} c_{\tau,m}g_{t-\tau}$	(6h)
The time from infection to death:	$\pi_m \sim \text{ifr}_m\text{Gamma}5.1, 0.86\ \text{Gamma}18.8, 0.45$	(6i)
The variance latent variable:	$\psi \sim \text{Normal}0, 5$	(6j)
The expected number of deaths:	$d_t = \sum_{\tau=0}^{t-1} c_{\tau,m}\pi_{t-\tau,m}$	(6k)
The observed daily deaths:	$D_t \sim \text{Negative Binomial}d_{t,m}, d_{t,m}\ \dfrac{d_{t,m}^2}{\psi}$	(6l)

The above hierarchical statistical model in Fig. 2(a) can be customized to estimate and forecast COVID-19 case numbers in terms of specific hypotheses, settings and conditions. For example, Fig. 2(b) shows the graphic model for the hierarchical model proposed in [77] to estimate the death number D from its inferred variable $\tilde{d}_t$ and inferred from auxiliary variable ψ. The inferred death number $\tilde{d}_t$ is sampled from the basic reproduction number R_0 with a normally-distributed prior $Normal2.4, |\kappa|$ parameterized by its variance variable κ and the probability of infected death π determined by two Gamma priors. In addition, $\tilde{d}_t$ is also influenced by the number of new infections c with two latent variables, the distribution rate of the Exponential distribution τ and the daily serial interval g and a variable α as a parameter of the reproduction rate. Fig. 2(b) also shows the prior distributions of the auxiliary variables, for example, assuming the variable describing the time from infection to death τ following an exponential prior $Exponential0.03$. The hierarchical statistical model in Fig. 2(b) to estimate the death number can be described by the following equations.

Another major set of COVID-19 statistical modeling incorporates statistical hypotheses and settings into other epidemic models such as compartmental models to approximate some state distributions or estimate some parameters. A typical application reformulates SIR-based models as a system of stochastic differential equations, e.g., by assuming Gamma-distributed probability density of the exposed E and infected I states in Section 6.1. Lastly, modeling the influence of mitigation strategies on the COVID-19 case numbers is also a typical statistical modeling problem.

4.2.2 COVID-19 Statistical Modeling. The contagion of an epidemic like COVID-19 is complex and uncertain. Statistical or probabilistic modeling naturally captures this uncertainty around epidemics better than other models. In COVID-19, for example, hierarchical Bayesian distributions with hidden states and parameters are used to model the causal relationships in their transmission [77, 162], and probabilistic compartmental models [91, 168, 275] integrate the transmission mechanisms of epidemics with the statistics of observed case data. Below, we summarize the relevant applications of descriptive analytics, Bayesian statistical modeling

and stochastic compartmental modeling of the COVID-19 epidemic statistics, epidemic processes, and the influence of external factors such as NPIs on the epidemic. Table 2 further summarizes various applications of COVID-19 statistical modeling.

COVID-19 descriptive analytics. Descriptive analytics are the starting point of COVID-19 statistical analysis, which are typically seen in non-modeling-focused references and communities. Typically, simple statistics such as the mean, deviation, trend and change of COVID-19 case numbers are calculated and compared. For example, the statistics of asymptomatic infectives are reported in [120]. In [19], change point analysis detects a change in the exponential rise of infected cases and Pearson's correlation between the change and lockdown implemented across risky zones. In addition, case statistics may be calculated in terms of specific scenarios, e.g., a population's mobility [100] or workplace [15].

Bayesian statistical modeling of COVID-19 epidemic processes. Bayesian statistical modeling can model stochastic COVID-19 epidemic processes, specific factors that may influence the COVID-19 epidemic process, causality, partially-observed data (e.g., under-reported infections or deaths), and other uncertainties. For example, stochastic processes are adopted to model conventional epidemic contagion [10, 166]. Niehus et al. [162] use a Bayesian statistical model to estimate the relative capacity of detecting imported cases of COVID-19 by assuming the observed case count to follow a Poisson distribution and the expected case count to be linearly proportional to daily air travel volume. To capture the complex relations in the COVID-19 pandemic, the causal relationship in the transmission process can be modeled by hierarchical Bayesian distributions [77, 162]. In [71], a special case of the continuous-time Markov population process, i.e., a partially-observable pure birth process, assumes a binomial distribution of partial observations of infected cases and estimates the future actual values of infections and the unreported percentage of infections in the population.

Bayesian statistical modeling of external factors on COVID-19 epidemic. Another important application is to model the influence of external factors on COVID-19 epidemic dynamics. For example, Flaxman et al. [77] infer the impact of NPIs including case isolation, educational institution closure, banning mass gatherings and/or public events and social distancing (including local and national lockdowns) in 11 European countries and estimate the course of COVID-19 by back-calculating infections from observed deaths by fitting a semi-mechanistic Bayesian hierarchical model with an infection-to-onset distribution and an onset-to-death distribution. In addition, case numbers, especially deaths, their model also jointly estimates the effect sizes of interventions.

Stochastic compartmental modeling of COVID-19 epidemic. Stochastic compartmental models can simulate stochastic hypotheses of specific aspects (e.g., probability of a state-based population or of a state transition) of the COVID-19 epidemiological process and the stochastic influence of external interventions on the COVID-19 epidemic process. Such probabilistic compartmental models integrate the transmission mechanisms of epidemics with the characteristics of observed case data [63, 91, 168, 275]. For example, in [239], a COVID-19 transmission tree is sampled from the genomic data with Markov chain Monte Carlo (MCMC)-based Bayesian inference under an epidemiological model, the parameters of the offspring distribution in this transmission tree are then inferred, and the model infers the person-to-person transmission in an early outbreak. Based on probabilistic compartmental modeling, Zhou et al. [275] develop a semiparametric Bayesian probabilistic extension of the classical SIR model, called BaySIR, with time-varying epidemiological parameters to infer the COVID-19 transmission dynamics by considering the undocumented and documented infections and estimates the disease transmission rate by a Gaussian process prior and the

removal rate by a gamma prior. To estimate the all-cause mortality effect of the pandemic, Kontis et al. [117] apply an ensemble of 16 statistical models (autoregressive with holiday and seasonal terms) on the vital statistics data for a comparable quantification of the weekly mortality effects of the first wave of COVID-19 and an estimation of the expected deaths in the absence of the pandemic. Other similar stochastic SIR models can also be found such as by assuming a Poisson time-dependent process on infection and reproduction [94], a beta distribution of infected and removed cases [241], and a Poisson distribution of susceptible, exposed, documented infected and undocumented infected populations in a city [130].

Statistical influence modeling of COVID-19 interventions and policies. Apart from modeling the transmission dynamics or forecasting case counts, Bayesian statistical models are also applied in some other areas, e.g., to estimate the state transition distributions by applying certain assumptions such as of the susceptible-to-infected (i.e., the infection rate) or infected-to-death (mortality rate) transition. For example, Cheng et al. [50] use a Bayesian dynamic item-response theory model to produce a statistically valid index for tracking the government response to COVID-19 policies. Dehning et al. [63] combine the established SIR model with Bayesian parameter inference with MCMC sampling to analyze the time dependence of the effective growth rate of new infections and to reveal the effectiveness of interventions. With the inferred central epidemiological parameters, they sample from the parameter distribution to evolve the SIR model equations and thus forecast future disease development. In [241], a basic SIR model is modified by adding different types of time-varying quarantine strategies such as government-imposed mass isolation policies and micro-inspection measures at the community level to establish a method of calibrating cases of under-reported infections.

Discussion. Statistical modeling and a Bayesian statistical framework allow us to elicit informative priors for parameters that are difficult to estimate due to the lack of data reflecting the clinical characteristics of COVID-19, offer coherent uncertainty quantification of the parameter estimates, and capture nonlinear and non-monotonic relationships without the need for specific parametric assumptions [275]. Compared with compartmental models, statistical models usually converge at different confidence levels for different assumptions (i.e., possible mitigation strategies), providing better interpretability and flexibility for characterizing the COVID-19 characteristics and complexities discussed in Section 2. However, the related work on COVID-19 statistical modeling is limited in terms of addressing COVID-19-specific characteristics and complexities, e.g., asymptomatic effect, the couplings between mitigation measures and case numbers, and the time-evolving and nonstationary case movement.

5 COVID-19 DATA-DRIVEN LEARNING

This section reviews the related work on data-driven discovery, i.e., applying classic (shallow) and deep machine learning methods, AI and data science techniques on COVID-19 data, to discover interesting knowledge and insights through characterizing, representing, analyzing, classifying and predicting COVID-19 problems.

5.1 Shallow and Deep Learning

Classic *shallow machine learning* methods have been predominantly applied to COVID-19 classification, prediction and simulation, as shown by the WHO-based literature statistics in [38]. Typical shallow learning methods include artificial neural networks (ANN), SVM, decision trees, Markov chain models, random forest, reinforcement learning, and transfer learning. These tools are easy to understand and implement and they are more applicable than other sophisticated methods (e.g., deep models and complex compartmental models)

Table 2. Examples of COVID-19 Mathematical and Statistical Learning.

Objectives	Approaches	Data
Infection diagnosis	Descriptive analytics, Bayesian models, continuous-time Markov processes [26, 71, 120, 162, 226]	Case numbers, demographics, biomedical test, medical imaging, sensor data, etc.
Transmission processes	Bayesian inference, stochastic compartmental models, state-space model, MCMC [91, 94, 130, 168, 239, 275]	Case numbers, demographics, genomic data, external factors, etc.
Medical treatment	Descriptive analytics, Bayesian models [117]	Health/medical data, case numbers, etc.
NPI evaluation	Bayesian models, temporal and hierarchical Bayesian model, stochastic compartmental models, compartmental models with Bayesian inference [5, 24, 50, 63, 77]	NPI policies, case numbers, external data (e.g., social activities), etc.
Sentiment and emotion impact	Descriptive analytics, latent models for sentiment/topic modeling, time-series analysis like regression variants [175, 187, 230, 258]	Questionnaire data, social media data, external factors like wellbeing, etc.
Social, economic and workforce influence	Descriptive analytics, time-series analysis, numerical methods, stochastic compartmental models [15, 64, 100, 109, 118, 129, 154, 215, 234, 236]	Case numbers, data related to economy, trade, supply chain, logistics, social activities, workforce, technology, transport, mobility, sustainability and public resources, etc.
Misinformation	Descriptive analytics, time-series models, numerical methods, statistical language models [4, 126, 199]	Fact data, online texts, social media, case numbers, etc.

for the often small COVID-19 data. They are well explained in the relevant literature (e.g., [47, 157, 160]) and interested readers can refer to them and other textbooks for technical details. Though different machine learning methods may be built on their respective learning paradigms [36], their main learning tasks and processes for COVID-19 modeling are similar, including (1) selecting discriminative features $\mathbf{x}$, (2) designing a model f (e.g., a random forest classifier) to predict the target y: $\tilde{y}\mathbf{x} = f\theta, \mathbf{x}, \mathbf{b}$ with parameters θ and bias term $\mathbf{b}$, and (3) optimizing the model to fit the COVID-19 data by defining and optimizing an objective function $\mathcal{L} = \arg\min_\theta y - \tilde{y}\mathbf{x}$ for the goodness of fit between expected $\tilde{y}$ and actual y target (e.g., infective or diseased case numbers).

Deep learning as represented by deep neural networks is a more advanced COVID-19 modeling typically favored by computing researchers. Typical models applied in COVID-19 modeling include (1) convolutional neural networks (CNN) and their extensions in particular for images such as ImageNet and ResNet; (2) sequential networks such as LSTM, recurrent neural networks (RNN), memory networks and their variants; (3) textual neural networks such as BIRT, Transformer and their variants; (4) unsupervised neural networks such as autoencoders and generative adversarial networks (GAN); and (5) other neural learning mechanisms such as attention networks.

Typical approaches for COVID-19 deep modeling can be represented by a general deep interaction and prediction framework as follows. It models (1) temporal dependencies over sequential *case* ($\mathbf{x}$, which may consist of categories of case numbers s, i and r or their rates) evolution, (2) interactions and influence between external containment *actions* ($\mathbf{a}$, which may consist of various control measures such as masking and social distancing) and case developments, and (3) the influence of personal *context* ($\mathbf{c}$, which may consist of demographic and health circumstances and symptomatic features on COVID-19 infections) over time t.

As COVID-19 case developments are sequential and stochastic to be influenced by many external factors, the framework combines autoencoders for the influence of unknown and stochastic asymptomatic and unreported case dynamics on reported numbers $\mathbf{x}$, RNN for sequential evolution of case numbers, control measures and personal context, and contextual attention for exterior containment strategies applied on case control to model complex interactions between various sources of underlying and control factors in COVID-19 sequential developments.
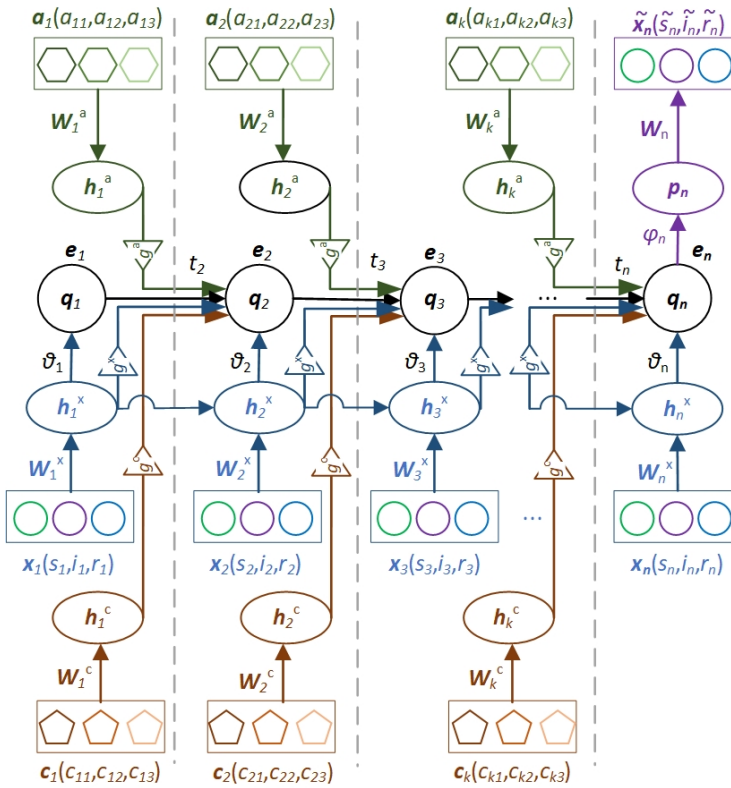

Fig. 3. A time-varying case-action-context neural interaction network for COVID-19 deep sequential modeling.

Fig. 3 illustrates a deep sequential case-action-context interaction network for this purpose. The formulation of the key variables and their interactions are shown below. In practice, networks $\mathbf{h}$ can be based on an LSTM, RNN, Transformer or other deep networks, the gating function unit g can be implemented by a gated recurrent unit (GRU) or other gating functions to determine the influence of control actions/context on case number movement. The transformation from input case vector $\mathbf{x}$ to action and context ($\mathbf{a}$ and $\mathbf{c}$) adjusted case representation $\mathbf{e}$ through network q_θ can be treated as an encoder, while the estimation (reconstruction) of $\mathbf{x}_t$ from $\mathbf{e}$ by network p_φ is a decoding or prediction process.

$$\mathbf{h}_t^x \sim h^x \mathbf{x}_t, \mathbf{W}_t^x, \mathbf{b}_t^x | \mathbf{h}_{t-1}^x \tag{7a}$$

$$\mathbf{h}_t^c \sim h^c \mathbf{c}_t, \mathbf{W}_t^c, \mathbf{b}_t^c \tag{7b}$$

$$\mathbf{h}_t^a \sim h^a \mathbf{a}_t, \mathbf{W}_t^a, \mathbf{b}_t^a \tag{7c}$$

$$\mathbf{g}_t^x \sim g \mathbf{x}_t, \mathbf{h}_t \tag{7d}$$

$$\mathbf{e}_t \sim q_\theta \mathbf{e}_t | \mathbf{e}_{t-1}, \mathbf{h}_t^x, \mathbf{h}_{t-1}^x, \mathbf{g}_t^x, \mathbf{g}_t^c, \mathbf{g}_t^a \tag{7e}$$

$$\tilde{\mathbf{x}}_t \sim p_\varphi \tilde{\mathbf{x}}_t | \mathbf{e}_t, \mathbf{h}_t^x \tag{7f}$$

The interaction and prediction network in Fig. 3 can be implemented in terms of an autoencoder (where q and p refer to encoding and decoding networks, e.g., [101]) or LSTM/RNN-based prediction (with q for representation and p for estimating the next input, e.g., [206]) framework. Accordingly, the objective function can be defined in terms of the discrepancy $\mathcal{J}$ between $\mathbf{x}_t$ and $\tilde{\mathbf{x}}_t$ (i.e., $\arg\min_{\theta,\varphi} \mathcal{J}\mathbf{x}_t, \tilde{\mathbf{x}}_t$) or the KL-divergence ($\mathcal{D}$) with loss $\mathcal{L}$ (where $\mathbf{h}_t$ and $\mathbf{h}_{t-1}$ refer to the representations of input $\mathbf{x}$ interacting with actions $\mathbf{a}$ under the context $\mathbf{c}$ through gating g integration).

$$\mathcal{L}\theta, \varphi; \mathbf{x}_t \sim -\mathcal{D}q_\theta \mathbf{e}_t | \mathbf{h}_t, \mathbf{h}_{t-1} || p_\varphi \mathbf{e}_t \; \mathcal{E}_{q_\theta} \mathbf{e}_t | \mathbf{h}_t, \mathbf{h}_{t-1} log p_\varphi \mathbf{h}_t, \mathbf{h}_{t-1} | \mathbf{e}_t \tag{8}$$

5.2 COVID-19 Shallow Learning

Here, *COVID-19 shallow learning* refers to the application of general shallow or classic machine learning methods to the analytics and modeling of COVID-19 problems and data. It forms the second popular set of modeling methods (about 4k of 22k WHO-listed references) that model COVID-19 outbreak, risk, transmission, uncertainty, anomalies, complexities, classification, variation, and prediction and more specifically case forecasting, medical diagnostics, contact tracing, and drug development [106]. General machine learning methods including ANN, tree models such as decision trees and random forest, kernel methods like SVM, transfer learning, NLP and text mining methods, evolutionary computing like genetic algorithms and fuzzy set, and reinforcement learning are mostly applied in addressing the above COVID-19 tasks by medical, biomedical, computing and social scientists [151, 194, 207], as discussed below.

Machine learning for COVID-19 outbreak prediction and risk assessment. Typical classifiers like ANN, SVM, decision trees, random forest, regression trees, least absolute shrinkage and selection operator (LASSO), and self-organizing maps are applied to forecast COVID-19 spread and outbreak and their coverage, patterns, growth and trends; estimate and forecast the confirmed, recovered and death case numbers or the transmission and mortality rates; and cluster infected cases and groups, etc. For example, in [108], logistic regression, decision trees, random forest and SVM are applied to estimate the growth trend and containment sign on the data consisting of factors about health infrastructure, environment, intervention policies and infection cases with accuracy between 76.2% and 92.9%. Evolutionary computing such as genetic algorithm, particle swarm optimization, and gray wolf optimizer forecast COVID-19 infections [161, 200, 224].

Machine learning for COVID-19 diagnosis on clinical attributes. The machine learning of COVID-19 clinical reports such as blood test results can assist in diagnosis. For example, in [114], clinic attributes and patient demographic data are extracted by term frequency/inverse document frequency (TF/IDF), bag of words (BOW) and report length from textual clinic reports. The extracted features are then classified in terms of COVID, acute respiratory distress syndrome (ARDS), SARS and both COVID and ARDS by SVM, multinomial naïve

Bayes, logistic regression, decision tree, random forest, bagging, Adaboost, and stochastic gradient boosting, reporting an accuracy of 96.2% using multinomial naıve Bayes and logistic regression. In [25], hematochemical values are extracted from routine blood exam-based clinic attributes, which are then classified into positive or negative COVID-19 infections by decision trees, extremely randomized trees, KNN, logistic regression, naive Bayes, random forest, and SVM. It reports an accuracy of 82% to 86%. The work in [253] applies random forest to identify COVID-19 infections.

Machine learning for COVID-19 diagnosis on respiratory data. Machine learning can be conducted on COVID-19 patient's respiratory data such as lung ultrasound waves and breathing and coughing signals to extract respiratory behavioral patterns and anomalies. For example, logistic regression, gradient boosting trees and SVMs distinguish COVID-19 infections from asthmatic or healthy people on the Android app-based collection of coughs and breathing sounds and symptoms with AUC at 80% [26].

Machine learning for COVID-19 diagnosis on medical imaging. A very intensive application of classic machine learning methods is to screen COVID-19 infections on CT, chest X-ray (CXR) or PET images. For example, in [43], the majority voting-based ensemble of SVM, decision tree, KNN, naive Bayes and ANN is applied to classify normal, pneumonia and COVID-19-infected patients on CXR images with an accuracy of 98% and AUC of 97.7%. In [74], the simple applications of SVM, naive Bayes, random forest and JRip on CT images screen COVID-19 diseases with a reported accuracy of 96.07% by naive Bayes combined with random forest and JRip, in comparison with 94.11% by CNN.

Machine learning for COVID-19 diagnosis on latent features. Further, shallow learners are applied to detect and diagnose COVID-19 infections on latent features learned by shallow to deep representation models on COVID-19 medical images. For example, in [107], ANN-based latent representation learning captures latent features from gray, texture, histogram, number, intensity, surface and volume features in CT images, then classifiers including SVM, logistic regression, Gaussian naive Bayes, KNN and ANN are applied to differentiate COVID-19 infections from community-acquired pneumonia with 95.5% accuracy reported. In [170], latent features are extracted from CXR and CT images to form a gray level co-occurrence matrix (GLCM), local binary gray level co-occurrence matrix (LBGLCM), gray level-run length matrix (GLRLM) and segmentation-based fractal texture analysis (SFTA)-based features, which are then oversampled by the synthetic minority over-sampling technique (SMOTE) and further selected by a stacked autoencoder (sAE) and principal component analysis (PCA), before SVM is applied to achieve 94.23% accuracy. In [222], MobileNetV2 and SqueezeNet extract features from CXR images, which are then processed by social mimic optimization to classify coronavirus, pneumonia, and normal images with 99.27% accuracy by SVM. Lastly, in [225], a residual exemplar local binary pattern (ResExLBP)-based method extracts features from CXR images, which are then selected by an iterative relief-based method before decision trees, linear discriminant, SVM, KNN and subspace discriminant are applied on the selected features to detect COVID-19 infection with an accuracy of 99.69% to 100.0%.

Modeling the influence of external factors on COVID-19. Various machine learning tasks are undertaken to analyze the relation and influence of external and contextual factors on COVID-19 epidemic attributes. For example, ensemble methods including random forest, extra trees regressor, AdaBoost, gradient boosting regressor, extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), CatBoost regressor, kernel ridge, SVM, KNN, MLP and decision trees indicate potential association between COVID-19 mortality and weather data [142].

Machine learning-driven drug and vaccine development for COVID-19. Machine learning methods are applied to analyze the drug-target interactions, drug selection, and the effectiveness of drugs and vaccines on containing COVID-19. For example, machine learning methods including XGBoost, random forest, MLP, SVM and logistic regression are used to screen thousands of hypothetical antibody sequences and select nine stable antibodies that potentially inhibit SARS-CoV-2 [111, 139].

5.3 COVID-19 Deep Learning

Deep learning has been intensively applied to modeling COVID-19 as discussed in Section 5.1, with about 2k of 22k references on modeling COVID-19. Typical applications involve COVID-19 data on daily infection case numbers, health and clinic records, hospital transactions, medical imaging, respiratory signals, genomic and protein sequences, and exterior data such as infective demographics, social media communications, news and textual information, etc. Below, we first discuss a common application of deep learning for COVID-19 epidemic description and forecasting, and then briefly review other applications.

Deep learning of the COVID-19 epidemic. Deep neural networks are intensively applied to characterize and forecast COVID-19 epidemic outbreak, dynamics and transmission. Examples are predicting the peak confirmed numbers and peak occurrence dates, forecasting daily confirmed, diseased and recovered case numbers, and forecasting N-day (e.g., $N = \{7, 14, 10, 30, 60 \; days\}$) infected/confirmed, recovered and death case numbers (or their transmission/mortality rates) through modeling short-range temporal dependencies in case numbers by applying LSTM, stacked LSTM, Bi-LSTM, convolutional LSTM-like RNNs, and GRU [65, 206]. Other work models the transmission dynamics and predicts daily infections of COVID-19 using a variational autoencoder (VAE), encoder-decoder LSTM or LSTM with encoder and Transformer [116] and modified auto-encoder [178], GAN and their variants, tracks its outbreak [99], predicts the outbreak size by encoding quarantine policies as the strength function in a deep neural network [60], estimates global transmission dynamics using a modified autoencoder [98], predicts epidemic size and lasting time, and combines medical information with local weather data to predict the risk level of a country by a shallow LSTM model [172]. In [265], a comparative analysis shows that VAE outperforms simple RNN, LSTM, BiLSTM and GRU in forecasting COVID-19 new and recovered cases.

Broad deep COVID-19 learning. In addition, we highlight several other typical application areas of COVID-19 deep learning:

(1) Characterizing symptoms of coronavirus infections, e.g., by pretrained neural networks (e.g., [202]), with more discussion in Section 6.2.1;
(2) Analyzing health and medical records, blood sample-based test reports, and respiratory sounds and signals for diagnosis and treatment e.g. by CNN, LSTM and GRU [194], with more discussion in Section 6.2.2;
(3) Analyzing medical imaging for diagnosis, quarantine and treatment by convolutional neural networks (CNN, e.g., ImageNet and ResNet), GAN and their mutations [103, 194], with more discussion in Section 6.2.3;
(4) Analyzing COVID-19 genomic and protein sequence and interaction analysis by RNN, CNN and their variants for drug and vaccine development, tracing infection sources, and analyzing virus structures and evolution, with more details in Section 6.2.4;
(5) Repurposing and developing drugs and vaccines by generative autoencoders, generative tensorial reinforcement learning and generative adversarial networks [270] for generative chemistry discovery;

(6) Analyzing COVID-19 impact on sentiment and emotion by RNN, Transformer-based NLP neural models and their derivatives [67, 131];

(7) Characterizing the COVID-19 infodemic by NLP and text mining including misinformation identification [208], enhancing epidemic modeling using social media data [115], and analyzing the COVID-19 research progress and topic evolution [269];

(8) Other topics such as analyzing the influence and effect of countermeasures, e.g., the effect of quarantine policies on outbreak using DNNs [60], with more discussion in Section 7.

Table 3 illustrates some typical applications of shallow and deep learning methods for modeling COVID-19. More discussion on COVID-19 deep learning can be found in Section 6.2.

Discussions. Most of the existing studies on shallow and deep COVID-19 modeling directly apply the existing shallow machine learning methods and deep neural networks on COVID-19 data, as shown in reviews like [133, 205, 240]. Our literature review also shows that deep neural models are widely applicable to COVID-19 modeling tasks, which are unnecessarily overwhelmingly applied to all possibilities and significantly outperform time-series forecasters and shallows machine learners. In fact, sometimes, deep models may even lose their advantage over traditional modelers such as ensembles, as shown in Table 3, and Table 5.

6 COVID-19 DOMAIN-DRIVEN MODELING

As a complex social-technical issue, COVID-19 modeling brings many specific challenges and research questions from the relevant domains and for domain-specific research communities. In this section, we focus on two major and mostly relevant domains of COVID-19: epidemic modeling, and medical and biomedical analysis.

6.1 COVID-19 Epidemic Modeling

6.1.1 Epidemiological compartmental models. Epidemiological modeling portrays the state-space, interaction processes and dynamics of an epidemic in terms of its macroscopic population, states and behaviors. Compartmental models are widely used in characterizing COVID-19 epidemiology by incorporating epidemic knowledge and compartmental hypotheses into imitating the multi-state COVID-19 population transitions. An individual in the COVID-19 epidemic sits at one state (compartment) at a time-point and may transit this state to another at a state transmission rate. The individuals of the closed population are respectively labeled per their compartments and migrate across compartments during the COVID-19 epidemic process, which are modeled by (ordinary) differential equations.

By consolidating various COVID-19 epidemiological characteristics, hypotheses and compartmental models, Fig. 4 illustrates a typical COVID-19 state-space and evolution system with major (thick) and minor (thin) states and state transition paths that can be sequentially categorized into four phases:

- Susceptible (S): Individuals (S) are susceptible to infection under free (uncontained or unrestrained, S_F at uncontained rate θ_F) or contained (restrained, $S_{\bar{F}}$ at containment rate $\theta_{\bar{F}}$) conditions at a respective transmission rate α_F or $\alpha_{\bar{F}}$;
- Exposed (E): Free or contained susceptibles are exposed to infection from those who are infected but in the incubation period (which could be as long as 14 days), and may be noninfectious and free (E_F) or infectious and contained ($E_{\bar{F}}$) at a respective exposure rate ε_F or $\varepsilon_{\bar{F}}$;

Table 3. Examples of COVID-19 Shallow and Deep Learning[a].

Objectives	Approaches	Data
Transmission & external factor impact	Shallow learners like SVM, ANN, decision tree, random forest and ensemble methods [108], evolutionary computing methods such as particle swarm optimization [161, 200, 224], DNN variants such as LSTM, GRU, VAE, GAN and BiLSTM, etc. [60, 65, 99, 116, 178, 206, 265]	Epidemic case numbers and external data such as meteorological data, environmental data (e.g., humidity), social activity and mobility data, etc.
Infection diagnosis	Shallow learners [25, 26, 43, 253], CNN and RNN variants like LSTM and GRU, and pretrained CNN-based image nets like ResNet, MobileNetV2 and SqueezeNet, etc. [74, 104, 222, 225], text analysis models [114]	Pathological and clinical records, respiratory signals (e.g. coughing and breathing signals and patterns in ultrasound or thermal video), computed tomography (CT) and CXR images, etc.
Mortality and survival analysis	Shallow learners like SVM, ANN, decision tree, regression tree, random forest and ensemble methods like XGBoost [41, 203, 203], CNNs, pretrained CNN-based image nets, RNN variants like LSTM and GRU [65, 206, 266]	Medical imaging including CT and CXR images, clinical records, patient demographics, case numbers, external data, etc.
Medical treatment	Shallow machine learning methods and DNNs, etc. [18, 254, 260, 274]	Health/medical records, pharmaceutical treatments, ICU data, etc.
Genomic and protein analysis, drug/vaccine development	Shallow classifiers like SVM and ensembles, frequent pattern and sequence analysis methods, CNN variants, RNN variants, attention networks, GAN, autoencoders, reinforcement learning, NLP models like Transformer, etc. [7, 18, 97, 111, 145, 146, 156, 270, 271]	Genomic data, proteomic data, drug-target interactions, molecular reactions, etc.
Resurgence and mutation	Shallow learners like linear discriminant, SVM, KNN and subspace discriminant, combining classifiers with compartmental models, DNN variants, sequence analysis[b], etc. [12, 193]	Resurgence case numbers, virus strain genome and protein sequences, NPI data, external data, etc.
NPI evaluation	Various Bayesian models[c], combining compartmental models with classifiers or estimators, DNNs[b], etc. [73, 80]	Case numbers, NPI policies, external data, etc.
Sentiment and emotion impact	NLP models like LDA and topic models and DNN variants like BERT and Transformer variants, etc. [14, 67, 131, 131, 150, 158, 245]	Social media data, news feeds, Q/A data, external factors, etc.
Socioeconomic influence	Relation (e.g., correlation and causality) analysis[b], topic modeling by NLP models [269]	Social, economic and workforce activities, case numbers, etc.
Misinformation analysis	Classic NLP models, correlation analysis, shallow learners, outlier detectors, DNN variants like BERT and Transformer mutations [126, 147, 208]	Social media, online texts, Q/A data, news feeds, etc.

[a] See Table 5 for deep COVID-19 medical imaging analysis; [b] such methods are applicable but not much work is reported in the literature; [c] See Table 2 for NPI effect modeling.

- Infective (I): Those exposed become infectious and may be detected (registered/documented and known to medical authorities, I_K) or undetected (unreported/undocumented and unknown to medical management, $I_{\bar{K}}$); also, some initially undetected infectives may be further detected and converted to detected infectives at rate $\beta_K^{\bar{K}}$; some documented
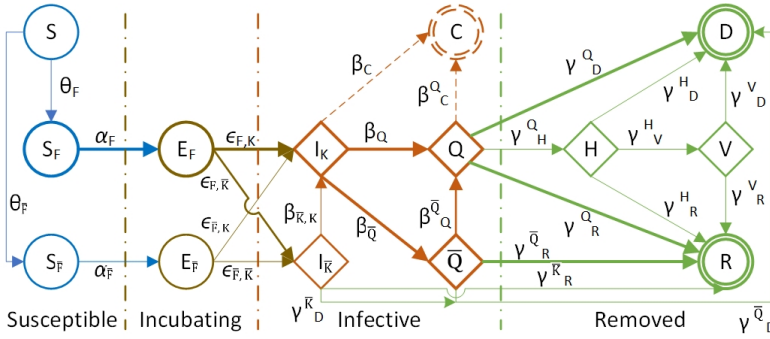
Fig. 4. COVID-19 epidemiological compartmental modeling. The COVID-19 epidemiological process can be categorized into four major phases: susceptible (S), exposed/incubating (E), infective (I) and removed (R), and other optional states. A *circle* denotes the initial states, a *diamond* denotes the optional states, and a *double circle* denotes the final states. *Blue* indicates the noninfectious states, *red* the infectious states, and *green* the states with patients removed or to be removed from the infection. The *thick* lines denote the main states and state transition paths while the *thin* lines are minor ones which may be ignored in modeling.

infectives may be symptomatic and quarantined (Q) at quarantine rate β_Q while others may be asymptomatic and unquarantined ($\bar{Q}$) at unquarantined rate $\beta_{\bar{Q}}$; there may be some rare cases (at rate β_C) who carry the virus and infection for a long time with or without symptoms, called lasting *carriers* (C); in addition, some initially asymptomatic cases may transfer to symptomatic and quarantined at rate $\beta_Q^{\bar{Q}}$;

- Removed (R): Unquarantined infectives may recover at recovery rate $\gamma_R^{\bar{Q}}$ or die at mortality rate $\gamma_D^{\bar{Q}}$ respectively, the same for quarantined infectives at rate γ_R^Q or γ_D^Q and unknown/undetected infectives at rate $\gamma_R^{\bar{K}}$ or $\gamma_D^{\bar{K}}$; some quarantined infectives may present acute symptoms even with life threat, who are then hospitalized (H) at rate γ_H^Q or even further ventilated (V) at rate γ_V^H; hospitalized infectives may recover or die at rate γ_R^H or γ_D^H, the same for ventilated at rate γ_R^V or γ_D^V.

In practice, the above COVID-19 state-space may be too complicated to model and not all states are characterizable by the available data. Accordingly, a focus is on those main states and their transitions when the corresponding data is available, for example, susceptible, exposed, infectious (which consists of both detected and undetected), recovered, and diseased. Below, we illustrate the differential equations of the states S, E, I, Q, $\bar{Q}$, R and D which are the main states of a closed COVID-19 population. Here, (1) S, E, I, Q, $\bar{Q}$, R and D represent the fraction of the population at each state; (2) S and E are impacted by containment measures at the containment rate θ and I at θ_I who are contained and recovered; and (3) the state transitions take place at the rates shown in Fig. 4.

$$\partial_t S = -\alpha SI - \theta S \tag{9a}$$

$$\partial_t E = \alpha SI - \epsilon E - \theta E \tag{9b}$$

$$\partial_t I = \epsilon E - \beta_Q I - \beta_{\bar{Q}} I - \theta_I I \tag{9c}$$

$$\partial_t Q = \beta_Q I \ \beta_{\bar{Q}}^Q \bar{Q} - \gamma_D^Q D - \gamma_R^Q R \tag{9d}$$

$$\partial_t \bar{Q} = \beta_{\bar{Q}} I - \beta_Q^{\bar{Q}} \bar{Q} - \gamma_R^{\bar{Q}} \bar{Q} - \gamma_D^{\bar{Q}} \bar{Q} \tag{9e}$$

$$\partial_t R = \gamma_R^Q R \ \gamma_R^{\bar{Q}} \bar{Q} \ \theta S \ \theta E \ \theta_I I \tag{9f}$$

$$\partial_t D = \gamma_D^Q Q \ \gamma_D^{\bar{Q}} \bar{Q} \tag{9g}$$

$$S \ E \ I \ Q \ \bar{Q} \ R \ D = 1 \tag{9h}$$

6.1.2 COVID-19 epidemiological modeling. We summarize and discuss the related work on the major tasks of COVID-19 epidemiological modeling, and highlight the work on modeling COVID-19 epidemic transmission processes, dynamics, external factor influence, and resurgence and mutation.

COVID-19 epidemiological modeling tasks. Epidemiological models dominate COVID-19 modeling (about 3.5k publications of the 22k reported in the WHO literature) by epidemic researchers and computing scientists through the expansion or hybridization with other models such as statistical models and machine learning methods. COVID-19 compartmental modeling aims to answer several epidemiological problems: (1) the growth (spread) of COVID-19 and its case number movements at different epidemiological states to forecast case numbers in the next days or periods; (2) the basic reproduction rate R_0 that informs the contagion and transmission level and control strategies; (3) the sensitivity and effect of control measures on infection containment and case movements; and (4) the sensitivity and effect of strategies for herd immunity and mass vaccination. Accordingly, various compartmental models are customized to cater for specific assumptions, settings and conditions of modeling COVID-19, as discussed in Section 6.1.2. For (1), with historical case numbers of a country or region and the initial settings of hyperparameters, we can estimate the parameters and further predict the number over time, e.g., the number of infections and deaths in a country or city. Regarding (2), with the state-space shown in Fig. 4, to resolve these differential equations, we first obtain the population projection matrix A corresponding to all states and their transition probabilities. The projection matrix A can be converted to a state transition matrix T (where each element T_{ij} is the probability of an individual transferring from state i at time t to state j at time j 1) and a fertility (reproductive) matrix F (where an element F_{ij} refers to the reproduced number of i-state offsprings of an individual at state j), i.e., $A = T \ F$. Further, we can calculate the fundamental matrix N: $N = I - T^{-1}$ with identity matrix I to represent the expected time spent in each state and that to death. Then, we can obtain another matrix R: $R = FN$ with each entry referring to the expected lifetime production number of i-state offspring by an individual at stage j [39, 211]. Its dominant eigenvalue is the net reproduction rate R_0. With regard to (3), since control measures such as social distancing and lockdown may influence the growth of case numbers and reproduction and transmission rates, we can analyze the sensitivity of adjusting related parameters on the case numbers and rates. To explore the opportunities for herd immunity and mass vaccination in (4), the herd immunity rate and vaccination rate are expected to be greater than $1 - \frac{1}{R_0}$ to eradicate the disease.

In addition to these major problems, below, we further discuss two applications of epidemiological modeling in COVID-19: modeling its transmission (which is also the most explored area) and resurgence and mutation (which is a recent challenge). More discussion on modeling the NPI effect on COVID-19 transmission and epidemic is in Section 7.1.

Modeling COVID-19 epidemic transmission process. Studies on modeling COVID-19 epidemic transmission mainly focus on evaluating the epidemiological attributes (e.g., infection rate, recovery rate, mortality, reproduction number, etc.), predicting the infection and death counts, and revealing the transmission, spread and outbreak trends under experimental or real-world scenarios. As illustrated in Table 4, various compartmental models are available to characterize COVID-19. For example, the SIDARTHE compartmental model considers eight stages of infection: susceptible (S), infected (I), diagnosed (D), ailing (A), recognized (R), threatened (T), healed (H) and extinct (E) to predict the course of the epidemic and to plan an effective control strategy [81]. A new compartment is introduced to the classic SIR model to quantify those who are symptomatic, quarantined infecteds [141]. Further, a stochastic SHARUCD model framework contains seven compartments: susceptible (S), severe cases prone to hospitalization (H), mild, sub-clinical or asymptomatic (A), recovered (R), patients admitted to the intensive care units (U), and the recorded cumulative positive cases (C), which include all new positive cases for each class of H, A, U, R, and deceased (D) [5]. In addition, several models involve new compartments to represent asymptomatic features to mild symptoms [5, 250] and undocumented cases [130].

Modeling COVID-19 epidemic dynamics, transmission and risk. Classic compartmental models assume constant transmission and recovery rates between state transitions. This assumption is taken in many SIR variants tailored for COVID-19, which cannot capture the disease characteristics in Section 2.1. To cater for COVID-19-specific characteristics especially when mitigation measures are involved, the classic susceptible-infectious-recovered (SIR) [110] and susceptible-exposed-infectious-recovered (SEIR) models [13], which were applied to modeling other epidemics like measles and Ebola, are tailored for COVID-19. Since COVID-19 transmission contains more states, especially with interventions, SIR/SEIR models are extended by adding customized compartments like quarantine, protected, asymptomatic and immune [5, 57, 81, 141, 250]. Accordingly, to capture the evolving COVID-19 epidemiological attributes including time-variant infection, mortality and recovery rates, time-dependent compartmental models are proposed. For example, a time-dependent SIR model adapts the change of infectious disease control and prevention laws as city lockdowns are imposed and traffic halt with the control parameters infection rate β and recovery rate γ modeled as time-variant variables [49]. Dynamical modeling is also considered in temporal SIR models with temporal susceptible, insusceptible, exposed, infectious, quarantined, recovered and closed (or death) cases in [177]. An early-stage study of a dynamic SEIR model estimates the epidemic peak and size, and an LSTM further forecasts its trend after taking into account public monitoring and detection policies [261].

Modeling the influence of external factors on the COVID-19 epidemics. COVID-19 epidemic dynamics reflect the time-varying states, state transition rates, and their vulnerability to contextual and external factors such as a person's ethnicity and public health conditions and social contacts and networking [135]. To depict the influence of external factors, more complex compartmental models involve the relevant side information (e.g., NPIs, demographic features such as age stratification and heterogeneity, and social activities such as population mobility) into their state transitions. Examples include an age-sensitive SIR model [51] which integrates known age-interaction contact patterns into the examination of potential effects of age-heterogeneous mitigations on an epidemic in a COVID-19-like parameter regime,

an age-structured SIR model with social contact matrices and Bayesian imputation [212], and an age-structured susceptible-exposed-infectious-recovered-dead (SEIRD) model that identifies no significant susceptibility difference between age groups [165]. More about the NPI influence on the COVID-19 epidemic is in Section 7.1. In addition, environmental factors, especially humidity and temperature, may affect COVID-19 virus survival and the epidemic's transmission [27, 164, 242, 257] despite inconsistent conclusions. In [59], variational mode decomposition decomposes COVID-19 case time series into multiple components and then a Bayesian regression neural network, cubist regression, KNN, quantile random forest and support vector regression (SVR) are combined to forecast six-day-ahead case movements by involving climatic exogenous variables.

Modeling COVID-19 resurgence and mutation. Our current understanding of COVID-19 resurgence and mutation is very limited while the British, South African, Indian and other newly-emergent mutations show higher contagion and complexities [85, 86]. COVID-19 may indeed become another epidemic disease which remains with humans for a long time. Imperative research is expected to quantify the virus mutation and disease resurgence conditions, forecast and control potential resurgences and future waves after lifting certain mitigation restrictions and reactivating businesses and social activities [137, 174], distinguish the epidemiological characteristics, age sensitivity, and intervention and containment measures between waves [8, 84], compare the epidemiological wave patterns between countries experiencing mutations and resurgences and compare COVID-19 wave patterns with influenza wave patterns [72], predict resurgences and mutations (e.g., by estimating the daily confirmed case growth when relaxing interstate movement, mobility and contact restrictions and social distancing by SEIR-expanded modeling) and prepare for countermeasures on future waves [12]. Limited research results are available in the literature on the above broad issues. For example, a comparative analysis in [21] shows the differences in the second COVID-19 wave in Europe in Italy and indicates the different causing strategies taken by them in implementing facemasks, social distancing, business closures and reopenings. In [32], building on fitting the first wave data, an epidemic renormalisation group approach further simulates the dynamics of disease transmission and spreading across European countries over weeks by modeling the European border control effects and social distancing in each country. In [127], an SIR model estimates the scenarios of incurring a potential second wave in China and the potential case fatality rate if containment measures such as travel ban and viral reintroduction from overseas importation are relaxed for certain durations in a population with a certain epidemic effect size and cumulative count after the first wave. In [174], an SEIR model incorporates social distancing to model the mechanism (closure releasing) of forming the second wave, the epidemiological conditions (ranges of transmission rate and the inverse of the average infectious duration) for triggering the second and third waves, and the socioeconomic (economic loss due to lockdown) and intervention (novel social behavior spread) factors on case numbers. In [137], a revised stochastic SEIR model estimates different resurgence scenarios reflected on infections when applying time-decaying immunity, lockdown release, or increasing implementation of social distancing and other individual NPIs.

Discussion. COVID-19 compartmental models excel at modeling epidemiological hypotheses, processes and factors with domain knowledge and interpretation. Such models often assume constant state-space transitions, capture average behaviors and the contagion of a closed population, and are sensitive to initial states and parameters. Challenges and opportunities exist in expanding its traditional frameworks to address the specific COVID-19 complexities and challenges in Section 2. Examples are time-varying, non-IID dynamics and

complex couplings between interior and exterior factors related to COVID-19 populations, management groups and contexts. Other important issues include understanding how vaccination and specific vaccines affect coronavirus mutation and discovering the relationships between interior and exterior factors and the resurgence and mutations.

Table 4. Examples of COVID-19 Epidemiological Modeling.

Objectives	Factors and Settings	Data
Epidemic transition and spread	SIR variants like SIDARTHE with eight phases: susceptible, infected, diagnosed, ailing, recognized, threatened, healed and extinct as well as severe symptoms [81], variational mode decomposition with shallow regressors [59], etc.	Case numbers, external data, etc.
Epidemic dynamics	Time-dependent compartment transmissions, time-varying state transition rates [49, 177]	Case numbers, reporting time information, etc.
Asymptomatic transmission	Asymptomatic to mild symptoms [250], undocumented [130], SHARUCD differing mild and asymptomatic from severe infections [5], undocumented cases [130], epidemiological interventions with serological tests, age-dependent and asymptomatic settings [250], etc.	Case numbers, reporting information, symptoms, demographics, etc.
External factor's epidemic influence	Age-sensitive SIR model [51], age-structured SIR model with social contacts [212], age-structured SEIRD [165], public monitoring and detection policies [261], ethnicity, public health conditions and social contacts [135], environmental factors [27, 164, 242, 257]	Case numbers, demographics, health conditions, social activities, environmental factors, etc.
NPI influence	Lockdown and social distancing [49], lockdown [5], quarantine [57], symptomatic and quarantined infecteds [141], self-protection and quarantine [177], etc.	Case numbers, NPIs, health conditions, test results, demographics, etc.
Resurgence	Second waves [12], wave difference [72, 84], reopening business and social activities [174], time-decaying immunity and easing lockdown and social distancing [137], age sensitivity [8], NPI influence on future waves [12], travel ban and virus importation [127]	Case numbers, multi-wave data, NPI and external data, etc.
Herd immunity	Compartmental model for simulating 'shield immunity' in a population [250]	Case numbers, serological tests, etc.

6.2 COVID-19 Medical and Biomedical Analyses

COVID-19 medical and biomedical analyses reveal the intrinsic and intricate characteristics, patterns and outliernesses of SARS-CoV-2 virus and COVID-19 disease. A wide range of research issues may benefit from such analyses, including but not limited to: COVID-19 infection diagnosis, prognosis and treatment, virology and pathogenesis analysis, potential therapeutics development (e.g., drug repurposing and vaccine development), genomic similarity analysis and sourcing, and contact tracing. In this section, we summarize the medical and biomedical modeling of COVID-19 infection diagnosis and case identification, risk and prognosis analysis, medical imaging analysis, pathological and treatment analysis and drug development.

6.2.1 COVID-19 infection diagnosis, test and case identification. Given the high transmission and reproduction rates, high contagion, and sophisticated and unclear transmission routes of COVID-19 and its virus strains such as the Delta strain, it is crucial to immediately identify and confirm exposed cases, test positive or negative infections, identify the infected virus variant types, and trace their origins and contacts so as to timely and proactively implement

appropriate quarantine measures and contain their potential spread and outbreak [159]. This is particularly important during the varying incubation periods which are often asymptomatic to mildly symptomatic yet highly contagious particularly for the virus variants. The SARS-CoV-2 diagnosis and test methods include (1) chemical and clinical methods, typically nucleic acid-based molecular diagnosis and antibody-based serological detection; (2) medical imaging-driven analysis, such as symptom inspection from CXR and CT images; (3) clinical diagnoses and tests like respiratory signal analysis, such as on the abnormal patterns of the lung's ultrasound waves and coughing and breathing signals; and (4) other noninvasive methods such as by involving SARS-CoV-2 and its disease data and external data [45, 47]. Data-driven discovery also plays an increasingly important role in improving COVID-19 diagnosis. Due to the virus and disease complexities, alternative and complementary to the chemical and clinical diagnosis approaches, COVID-19 identification [226] can benefit from analyzing biomedical images, genomic analysis, symptom identification and discrimination, and external data including social contacts, social activities, mobility and media communications, etc. by data-driven discovery [36].

- *Nucleic acid-based diagnosis test* (NAT) [3, 76, 263] refers to various molecular diagnosis test methods, including non-isothermal amplification (e.g., the real-time reverse transcription polymerase chain reaction (RT-PCR) test, which is the golden standard of COVID-19 diagnosis), isothermal amplification (e.g., CRISPR-based), and sequencing-based tests. Such methods may benefit from modeling techniques including gene and protein sequence analysis and drug-target and virus-host interaction analysis. It is highly sensitive and usable for large-scale operations, but it is expensive as typically it is done using specific test materials and in labs, and is less accurate as it is subject to the varied quality and quantity of specimen collections. The challenges are to reduce its false-negative and false-positive rates supplemented by other diagnosis tools and develop scalable fast test tools.

- *Antibody-based serological diagnosis* [125, 132, 176] is to detect anti-SARS-CoV-2 immunoglobulins i.e. the antibodies produced in response to COVID-19 infections by validating the specificity and sensitivity of chemiluminescent immunoassays, enzyme-linked immunosorbent assays and lateral flow immunoassays against SARS-CoV-2. It is an alternative or complement to NATs for acute infection diagnosis with easier and cheaper operations at any time. It, however, may produce poor-performing results which are unreliable for decision-making, it may take time to get the results, and it might be difficult for early large-scale diagnosis. There is an urgent need to develop more accurate serological test methods and tools. Machine learning methods such as CNNs could improve test performance e.g. by analyzing the test results, involving external data on patient demographics and clinical results, and integrating various test results [144].

- *Clinical diagnosis and analysis* involves clinical reports, domain knowledge and clinicians in identifying COVID-19-specific symptoms, indications and infections, differentiating them from other similar diseases such as influenza, and confirming positive, negative, severe or fatal conditions. Such diagnoses are conducted by blood tests, cough sound judgment, breathing pattern detection, and external factors by involving external data, etc. AI, machine learning and analytics methods are increasingly being used to classify COVID-19 from other diseases, predict infections, recovery and mortality rates, numbers or timing, etc. [25, 26, 114]. For further discussion, see Section 5.2.

- *Clinical medical imaging analysis* for COVID-19 inspection on COVID-19-sensitive medical images, typically by DNN-based image analysis, can complement the afore-mentioned chemical and medical methods by detecting abnormal and discriminative symptoms and patterns sensitive to COVID-19 in patient's CXR and CT images. Both typical deep and shallow learning methods are widely applied, which also present inconsistencies and biases in their applications, experiments, results and actionability [197]. For further discussion, see Section 6.2.3.
- *Data-driven prediction* on COVID-19 related data such as blood test results, respiratory signals, and external data that may indicate symptoms, patterns or anomalies of COVID-19 infections. Shallow and deep learning and mathematical modeling methods are applied to classify the symptom types, differentiate COVID-19 infections from other diseases, or detect outliers that may indicate COVID-19 infections. For example, in [253], a random forest algorithm-driven assistant discrimination tool extracts 11 top-ranking clinically available blood indices from 49 blood test samples to identify COVID-19 infectives from suspected patients. In [202], computer audition is used to recognize COVID-19 patients under different semantics such as breathing, dry/wet coughing or sneezing, and speech during colds, etc. AI4COVID-19 [102] combines the deep domain knowledge of medical experts with smart phones to record cough/sound signals as the input data to identify suspect COVID-19 infections with 92.8% accuracy reported. In [152], a shallow LSTM model combines medical information and local weather data to predict the risk level of the country.

Discussion. A comparison of the diagnosis methods is shown in Table 6. As commented in various reviews [20, 132, 148, 218, 231], COVID-19 diagnosis and tests still suffer from various limitations and challenges. The issues include concerns about result quality, implementation scalability, actionability for determining isolation and quarantine strategies, and trustfulness of accepting medical findings as general clinical specifications. An increasing number of studies appear promising by incorporating advanced data science and AI techniques to complement medical and chemical test approaches and tools, to integratively enhance preanalytical and postanalytical test results, and strengthen the interpretability and actionability of the results for clinicians, microbiological staff and public health authorities.

6.2.2 COVID-19 patient risk and prognosis analysis.
COVID-19 patient risk assessment identifies the risk factors and parameters associated with patient infections, disease severity, and recovery or fatality to support accurate and efficient prognosis, resource planning, treatment planning, and intensive care prediction. This is crucial for early interventions before patients progress to more severe illness stages. Moreover, risk and prognosis prediction for patients can help with effective health and medical resource allocation when intense monitoring, such as that involving ICU and ventilation and more urgent medical interventions are needed and prioritized. Machine learning models and data-driven discovery can also play a vital role in such risk factor analysis and scoring, prediction, prioritization and planning of prognostic and hospitalization resources and facilities, treatment and discharge planning, and the influence and relation analysis between COVID-19 infection and disease conditions and the external environment and context (e.g., weather conditions and socioeconomic statuses).

Techniques including mathematical models, and shallow and deep learners are applicable on health records, medical images, and external data. For example, LightGBM and Cox proportional-hazard (CoxPH) regression models incorporate quantitative lung-lesion features and clinical parameters (e.g., age, albumin, blood oxygen saturation, CRP) for prognosis prediction [268], their results showing that lesion features are the most significant contributors

Table 5. Examples of Deep COVID-19 Medical Imaging Analysis.

Method	Task	Data	Performance
Domain extension transfer learning with pretrained CNN [17]	Diagnosis	Italian Society of Medical Radiology and Interventional (25 cases)[a]; Radiopaedia.org (20 cases)[b]; COVID-19 images (180 cases) [55]; A Spanish hospital (80 cases)[c]	Overall accuracy 90.13% $\pm$ 0.14
Shallow CNN [152]	Diagnosis	COVID-19 images [55] (321 cases); Kaggle non-COVID-19 CXR images (5856 cases)[d]	Highest accuracy 99.69%, sensitivity 1.0, AUC 0.9995
CNN-based truncated InceptionNet [61]	Diagnosis	COVID-19 images [55] (162 cases); Kaggle CXR images (5863 cases); Tuberculosis CXR images[e]	Accuracy 99.96% (AUC of 1.0) in classifying COVID-19 cases from combined pneumonia and healthy cases; accuracy 99.92% (AUC of 0.99) in classifying COVID-19 cases from combined pneumonia, tuberculosis and healthy CXRs
DarkCovidNet [171] built on Darknet-19	Diagnosis	COVID-19 images [55] (127 cases); CXR images [246]	Accuracy 98.08% for binary classes and 87.02% for multi-class cases
CoroNet built on Xception pre-trained on ImageNet [113]	Diagnosis	COVID-19 images [55]; Kaggle CXR images	Accuracy 89.6%, precision 93% and recall 98.2% for COVID vs pneumonia bacterial, pneumonia viral and normal
COVID-CAPS based on capsule network [2]	Diagnosis	COVID-19 images [55], Kaggle CXR images	Accuracy 95.7%, sensitivity 90%, specificity 95.8%, and AUC 0.97
VGG16 and transfer learning [163]	Diagnosis	COVID-19 images [55], RSNA Pneumonia Detection Challenge data[f]	Accuracy 83.6% for COVID-19 pneumonia vs non-COVID-19 pneumonia and healthy; sensitivity 90% for COVID-19 pneumonia
COVID-Net [240]	Diagnosis	COVID-19 images [55], COVID-19 CXR Dataset Initiative[g], ActualMed COVID-19 CXR Dataset Initiative[h], RSNA data, COVID-19 radiography data[i]	Accuracy 93.3%, sensitivity 91.0%, positive predictive value 98.9%
Inf-Net with decoder and attention [55]	Segment	COVID-19 CT Segmentation[j], COVID-19 CT Collection	Dice similarity coefficient 0.739, Sensitivity 0.725, Specificity 0.960
DeepPneumonia built on ResNet-50 [214]	Diagnosis	Private data	Sensitivity 0.93, AUC 0.99; AUC 0.95 and sensitivity 0.96 for COVID-19 vs. bacteria pneumonia-infections

[a]https://www.sirm.org/category/senza-categoria/covid-19/;
[b]https://radiopaedia.org/search?utf8=%E2%9C%93&q=covid&scope=all&lang=us;
[c]https://twitter.com/ChestImaging/status/1243928581983670272;
[d]https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia;
[e]https://ceb.nlm.nih.gov/tuberculosis-chest-X-ray-image-data-sets/;
[f]https://www.kaggle.com/c/rsna-pneumonia-detection-challenge;
[g]https://github.com/agchung/Figure1-COVID-chest-xray-dataset;
[h]https://github.com/agchung/Actualmed-COVID-chest-xray-dataset;
[i]https://www.kaggle.com/tawsifurrahman/covid19-radiography-database;
[j]https://medicalsegmentation.com/covid19/.

in clinical prognosis estimation. Supervised classifiers like XGBoost are applied on electronic health records to predict the survival and mortality rates of severe COVID-19 infectious

Table 6. Comparison of COVID-19 Infection Diagnosis Methods.

Methods	Pros	Cons	Data
Nucleic acid-based diagnosis test [263]	High sensitivity, suitable for large-scale operation	Preliminary assessment by technicians, professional data analysis, expensive, less accurate, false-negative or false-positive results	Nasal, nasopharyngeal or oropharyngeal swab, aspiration, saliva or wash specimens
Serological diagnosis [132]	Easy and cheap to implement, no requirement of experts	Unstable performance, time-inefficient, unscalable for early diagnosis	Serum or plasma samples
Clinical diagnosis analysis [25]	Diagnosis from mixed clinical reports and tests, on-demand, verifiable by domain experts	Require professional tools and domain knowledge	Blood and respiratory test samples, etc.
Medical imaging inspection [197]	Fast and automated detection, data-driven analysis	Need trained experts, costly in labeling and early detection, train data scarcity	CT and CXR images
Data-driven prediction [47]	Algorithmic prediction by data-driven analytics and learning on data relevant to the COVID-19 diagnosis	Biases from data and predictors	Any relevant data including clinical test results and genomic/protein sequences

patients [203, 259] for the detection, early intervention and potential reduction of mortality of high-risk patients. In [189], logistic regression and random forest are used to model CT radiomics on features extracted from pneumonia lesions to predict feasible and accurate COVID-19 patient hospital stay, which can be treated as one of the prognostic indicators. Further, shallow and deep machine learning methods are applied to screen COVID-19 infections on respiratory data including lung ultrasound waves, coughing and breathing signals. For example, in [104], a bidirectional GRU network with attention differentiates COVID-19 infections from normal on face-based videos captured by RGB-Infrared Sensors with 83.69% accuracy. Lastly, external data can be involved for risk analysis; e.g., the work in [142] analyzes the association between weather conditions and COVID-19 confirmed cases and mortality.

6.2.3 COVID-19 medical imaging analysis. A rapidly growing body of research literature on COVID-19 medical image processing is available, which involves both shallow and deep learning methods especially pretrained CNN-based image nets in learning tasks such as feature extraction, region of interest (ROI) segmentation, infection region/object detection, and disease/symptom diagnosis and classification, etc. Typical COVID-19 medical imaging data includes CXR and CT images of lung (lobes or segments), lesion, trachea and bronchus. The most commonly used DNNs are pretrained or customized CNN, GAN, VGG, Inception, Xception, ResNet, DenseNet and their variants [103, 209].

Further, CNN-based transfer learning models, deep transfer learning and GAN are applied on CXR images to detect COVID-19 pneumonia and its segmentation and severity [54, 112, 149]. On chest CT images, CNNs like ResNet, DenseNet and VGG16 and the inception transfer model are applied to classify COVID-19 infected patients and detect and localize COVID-19 pneumonia and infection regions [11, 184, 214, 244, 272].

The application of DNNs in COVID-19 medical imaging analysis show significant performance advantages. For example, several references report close-to-perfect prediction performance of pretrained DNNs on CXR images (e.g., achieving accuracy and F-score 100

[136], AUC 100 [188] and 99.97 [140], and accuracy and F-score 98 [155]), in contrast to the lower performance of customized networks on CT images (e.g., with accuracy 99.68 [90], AUC 99.4 [74] and F-score 94 [74, 214]). The highly promising medical imaging analysis results provide strong evidence and support to further case confirmation, medical treatment, hospitalization resource planning, and quarantine, etc.

Table 5 illustrates various DNNs applied on medical imaging for COVID-19 screening and abnormal infection region segmentation, etc. For example, various CNNs such as shallow CNN, truncated InceptionNet, VGG19, MobileNet v2, Xception, ResNet18, ResNet50, SqueezeNet, DenseNet-121, COVIDX-Net with seven different architectures of deep CNN models, GoogleNet, AlexNet and capsule networks [2, 17, 61, 92, 152, 240] are applied to analyze CXR images for screening COVID-19 patients, assisting in their diagnosis, quarantine and treatments, and differentiating COVID-19 infections from normal, pneumonia-bacterial and pneumonia-viral infections.

6.2.4   COVID-19 pathological and treatment analysis and drug development. The modeling of COVID-19 pathology and treatment aims to characterize virus origin and spread, infection sources, pathological findings, immune responses, and drug and vaccine development, etc. The formulation of molecular mechanisms and pathological characteristics underlying viral infection can inform the development of specific anti-coronavirus therapeutics and prophylactics, which disclose the structures, functions and antigenicity of SARS-CoV-2 spike glycoprotein [237]. The pathological findings pave the way to design vaccines against the coronavirus and its mutations. For example, the higher capacity of membrane fusion of SARS-CoV-2 compared with SARS-CoV is shown in [256], suggesting the fusion machinery of SARS-CoV-2 as an important target of developing coronavirus fusion inhibitors. Further, human angiotensin coverting enzyme 2 (hACE2) may be the receptor for SARS-CoV-2 [169] informing drug and vaccine development for SARS-Cov-2. In [238], a structural framework for understanding coronavirus neutralization by human antibodies can help understand the human immune response upon coronavirus infection and activate coronavirus membrane fusion. The kinetics of immune responses to mild-to-moderate COVID-19 discloses clinical and virological features [220]. Data-driven analytics are applied in COVID-19 virology, pathogenesis, genomics and proteomics and collecting pathological testing results, gene sequences, protein sequences, physical and chemical properties of SARS-CoV-2, drug information and its effect, together with their domain knowledge. This plays an important role in discovering and exploring feasible drugs and treatments, drug discovery, drug repurposing, and correlating drugs with protein structures for COVID-19 drug selection and development. For example, a pre-trained MT-DTI (molecule transformer-drug target interaction) deep learning model based on the self-attention mechanism identifies commercially available antiviral drugs by finding useful information in drug-target interaction tasks [18]. The GAN-based drug discovery pipeline generates novel potential compounds targeting the SARS-CoV-2 main protease in [271]. In [270], 28 machine learning methods including generative autoencoders, generative adversarial networks, genetic algorithms, and language models generate molecular structures and representations on top of generative chemistry pipelines and optimize them with reinforcement learning to design novel drug-like inhibitors of SARS-CoV-2. Further, multitask DNN screens candidate biological products [97]. In [145, 146], CNN-enabled CRISPR-based surveillance supports a rapid design of nucleic acid detection assays.

For genome and protein analysis, frequent sequential pattern mining identifies frequent patterns of nucleotide bases, predicts nucleotide base(s) from their previous ones, and identifies the genome sequence locations where nucleotide bases are changed [156]. In [7], a

bidirectional RNN classifies and predicts the interactions between COVID-19 non-structural proteins and between the SARS-COV-2 virus proteins and other human proteins with an accuracy of 97.76%.

Classic and deep machine learning methods such as classifiers SVM and XGBoost, sequence analysis, multi-task learning, deep RNNs, reinforcement learning such as deep Q-learning network, and NLP models are applied to SARS-COV-2 therapy discovery, drug discovery, and vaccine discovery [111]. Examples are the rule-based filtering and selection of COVID-19 molecular mechanisms and targets; virtual screening of protein-based repurposed drug combinations; identifying the links between human proteins and SARS-COV-2 proteins; developing new broad-spectrum antivirals, and molecular docking; identifying functional RNA structural elements; discovering vaccines such as predicting potential epitopes for SARS-COV-2 and vaccine peptides by LSTM and RNNs, and analyzing protein interactions, molecular reactions by neural NLP models such as Transformer variants. Table 7 briefly illustrates the applications of modeling in supporting COVID-19 treatments and drug and vaccine development.

Discussion. Most of the literature on COVID-19 medical and biomedical analytics directly applies existing mathematical models, shallow and pretrained deep models. There are gaps and opportunities in characterizing COVID-19-specific characteristics and domain knowledge into tailored modeling and training deep neural networks on usually small and quality-limited COVID-19 data and involving multimodal COVID-19 data to discover more informative medical and biomedical insights.

Table 7. Examples of Modeling for COVID-19 Treatment and Drug/Vaccine Development.

Objective	Approaches	Data
Treatment	Data-driven diagnosis-informed treatment e.g. pathological analysis, medical imaging analysis, immune reaction, genomic and proteomic analysis [103, 103, 209, 268]	Pathological, clinical, virological, genomic, proteomic data
Drug development	Correlating drugs with protein structures and molecule transformer for drug-target interactions [18], DNNs like GANs and multitask DNNs for drug discovery [97, 271], machine learning and language models to generate molecular structures and drug-like inhibitors [270]	Virological, genomic, proteomic data
Vaccine development	Sequence analysis and sequential modeling like LSTM and RNN variants and NLP models like Transformer variants for functional RNA structures, vaccine epitopes and peptides, protein interactions and molecular reactions [111]	Genomic and proteomic data

7 COVID-19 INFLUENCE AND IMPACT MODELING

COVID-19 has had an unprecedented and overwhelming influence and impact on all aspects of our life, society and economy, posing significant health, economic, environmental and social challenges to the entire world and human population [40]. Over 3k references of the 22k literature involve the topic of influence and impact modeling. In this section, we review and summarize the modeling and analysis methods and results on many broad areas affected by SARS-CoV-2 and COVID-19. These include the modeling of the effect of COVID-19-sensitive NPIs and the COVID-19 healthcare, psychological, economic and social influence and impact.

7.1 Modeling COVID-19 Intervention and Policy Effect

On one hand, pharmaceutical measures, drug and vaccine development play fundamental roles [274]. On the other, to control the outbreaks of COVID-19 and its further influence on various aspects of life, governments adopt various NPIs such as travel restrictions, border control, business and school shutdown, public and private gathering restrictions, mask-wearing, and social distancing. For example, travel bans and lockdown are issued to decrease cross-boarder population movement; social distancing and shutdowns minimize contacts and community spread; schooling closures and teleworking reduce indoor gatherings and workplace infections. Although these control measures flatten the curve, they also undoubtedly change the regular mobility and activities of the population, normal business and economic operations, and the usual practices of our daily businesses.

A critical modeling issue is to characterize, estimate and predict how such NPIs influence COVID-19 epidemic dynamics, infection spread, case development, and population structure including deceased, medical resource and treatment allocation, and human, economic and business activities. Accordingly, various modeling tasks involve epidemiological, statistical and social science modeling methods and their hybridization (typically stochastic compartmental models) to evaluate and estimate the effects, typically by aligning the NPIs with case numbers for correlation and dependency modeling. Below, we summarize a few aspects of NPI influence.

Modeling the effect of NPIs on COVID-19 epidemic dynamics. This typically models the correlations between COVID-19 cases and NPIs, the NPI influence on COVID-19 epidemic factors including transmission rate and case numbers, and the NPI influence on improving recovery rates and lowering death rates. Various SIR and statistical modeling variants evaluate the effects of such control measures and their combinations on containing the virus spread and controlling infection transmission (e.g., per transmission rate) and estimate the corresponding scenarios (distributions) of case number development [24, 63, 221]. For example, in [177], a generalized SEIR model includes the self-protection and quarantine measures to interpret the publicly released case numbers and forecast their trend in China. The effect of control measures, including city lockdowns and travel bans implemented in the first 50 days in Wuhan and their effect on controlling its outbreak across China in terms of infection case numbers estimated by an SEIR model before and after the controls is described in [221].

Often, various NPIs are jointly implemented to contain a COVID-19 epidemic. It may be reasonable that multiple NPIs cooperatively reduce the epidemic effective reproduction number [24, 77, 121, 185]. In [24], a temporal Bayesian hierarchical model incorporates auxiliary variables describing the temporal implementation of NPIs, which infers the effectiveness of individually (estimated 13% to 42% reduction of reproduction number) and conjunctionally (77% reduction of reproduction number) implementing NPIs such as staying-at-home, business closures, shutting down educational institutions and limiting gathering sizes in terms of their influence on the reproduction number. In [77], a hierarchical Bayesian model infers the impact and effectiveness of NPI (including case isolation, school closure, mass gathering ban, social distancing) on the infections, reproduction number R_0, effect sizes of population, and death tolls in 11 European countries and suggests continued interventions to keep the epidemic under control.

Modeling NPI influence on public resources including healthcare systems. The implementation of NPIs affects the demand, priority and effectiveness of anti-pandemic public health resources and the planning and operations of healthcare systems. For example, in

[73], an SEIR model and a polynomial regressor simulates the effect of early detection, isolation, treatment, adequate medical supplies, hospitalization and therapeutic strategy on COVID-19 transmission, in addition to estimating the reproductive number and confirmed case dynamics. The SIDARTHE model [81] simulates possible scenarios and the necessity of implementing countermeasures such as lockdowns and social distancing together with population-wide testing and contact tracing to rapidly control the pandemic. The SHARUCD model [5] predicts the COVID-19 transmission response (in terms of infection cases, growth rate and reproduction number) to the control measures including partial lockdown, social distancing and home quarantining and differentiates asymptomatic and mild-symptomatic from severe infections, which could inform the prioritization of healthcare supplies and resources.

Modeling NPI influence on human activities. This explores the relations between COVID-19 NPIs and human mobility, travel, and social and online activities. For example, in [118], the alignment between human mobility and case number development in Wuhan and China presents the effect of travel restrictions on case reduction and COVID-19 spread. In [109], a simple SEIR model analyzes the tracing contacts in UK social network data, estimates the scenarios of COVID-19 infection control and subsequent untraced cases and infections, and shows the efficacy of close contact tracing in identifying secondary infections. In [80], MCMC parameter estimation and a metacommunity Susceptible–Exposed–Infected–Recovered (SEIR)-like disease transmission model shows the need for planning emergency containment measures such as restrictions on human mobility and interactions to control COVID-19 outbreak (by 42% to 49% transmission reduction). In [83], mobile phone data is collected and analyzed to inform COVID-19 epidemiologically relevant behaviors and response to interventions. Weitz et al. [250] develop and analyze an epidemiological intervention model that leverages serological tests to identify and deploy recovered individuals as focal points for sustaining safer interactions by interaction substitution, developing the so-called 'shield immunity' at the population scale. It is shown that the change of contact patterns could dramatically decrease the probability of infections and reduce the transmission rate of COVID-19 [75, 123, 267].

Discussion. The many diverse applications of SIR-based modeling of COVID-19 invention and policy effects enable an epidemiological explanation. Such methods assume each NPI independently acts on case movement. This leaves open issues including characterizing the effectiveness of individual NPIs by assuming they are coupled with each other and cooperatively contribute to flatten the curves; and exploring the interactions between NPIs, case development, and external factors including people's behaviors and environmental factors without disentangling them (opposite to the method of DNNs-based decoupled, homogeneous and independent representations and learning).

7.2 Modeling COVID-19 Psychological Impact

A common concern is the influence of COVID-19 on individual and public psychological and mental health [258]. Typical tasks are to characterize, classify and predict social-media-based individual and public emotion and sentiment and their mental health. These may be sensitive to the COVID-19 outbreak, health and medical mitigation, NPI measures, government governance, public healthcare system performance, vaccine, resurgence and coronavirus mutations, and the 'new normal' including working from home and online education, etc. The data involved are from social media and networks such as Twitter, Facebook, Wechat, Weibo, YouTube, Instagram and Reddit; online news feeds, discussion boards, blogs and Q/A; and instant messaging such as mobile messaging and apps.

Negative sentiments [158, 245], opinion and topic trends, online hate speech [233], psychological stress, men's and women's worries [230], responsive emotions [95] and behaviors and events [14] can be characterized, clustered or classified on short and long texts by simply applying NLP processing techniques. Examples are extracting TF-IDF and part-of-speech features, shallow NLP and text analysis models including BOW and latent Dirichlet allocation (LDA), and neural text modelers including DNN variants such as BioBERT, SciBERT and Transformer variants on the word, sentence or corpus level. For example, in [258], the preferred reporting items for systematic reviews and meta-analyses guidelines are used to review the COVID-19 impact on public mental health, disclosing the extent of symptoms and risk factors associated with anxiety (6.33% to 50.9%), depression (14.6% to 48.3%), posttraumatic stress disorder (7% to 53.8%), psychological distress (34.43% to 38%) and stress (8.1% to 81.9%) in the surveyed population of 8 countries.

Discussion. The existing modeling of COVID-sensitive psychological influence often misses psychological knowledge because it is purely driven by data; the analytical results are based on a cohort of infected people owing to its anonymous nature; no work is reported on fusing various sources of data including online misinformation to infer the predominant drivers of specific mental stress such as vaccination hesitation; and the targeted analysis of specific mental issues in vulnerable groups, such as COVID-driven teenage suicide and racism.

7.3 Modeling COVID-19 Economic Impact

The COVID-19 pandemic has incurred overwhelming and devastating impact on regional and global economy and business activities including trade, tourism, education exchange, logistics, supply chain, workforce and employment. It seems that no economy on the interconnected globe is immune from the negative consequences of COVID-19 [52]. A critical modeling task is to quantify how COVID-19 influences various aspects of the economy and businesses, how to manage and balance COVID-19 control measures (including NPIs and vaccination rollouts) and government relief and recovery programs, and how to sustain and recover business and economic activities without seriously suffering from uncontrollable outbreaks and resurgences for better sustainability in the COVID new normal.

Modeling the COVID-19 impact on economic growth. A rapidly growing body of research investigates the heterogeneous, non-linear and uncertain macroeconomic effects of COVID-19 across regions and sectors in individual countries, as well as on a global scale. It is estimated that COVID-19 and SARS-CoV-2 may cause over 2% monthly GDP loss and a 50% to 70% decline in tourism [40]. In [181], a sectoral macroeconomic model analyzes the short-term effects of intervention measures such as lockdown, social distancing and business reopening on economic outcomes such as production network, supply and demand, inventory dynamics, unemployment and consumption and estimates their influence on the relations between reproduction number and GDP. The study in [236] illustrates the relations between a country's income levels, public healthcare availability and capacity and the COVID-19 infected patient's demography and social patterns in low- to middle-income countries.

Modeling the COVID-19 impact on workforce and sustainability. COVID-19 drives the new normal of working, including a hybrid work mode, cloud-based enterprise operations, the shift from centralized infrastructures (including IT) to cloud-based ICT and home-based workplaces, and new ways of ensuring sustainability including engaging and supporting clients through online operations and services and AI-enabled cost-effective planning, production, logistics and services. In [15], the descriptive statistics of the daily activities of Baidu developers show the positive and negative impacts of working from home on developer productivity, particularly on large and collaborative projects. The survey conducted in [154]

shows the various impacts of COVID-19-driven work from home on the scientific workforce, including the time spent on work, parenting distraction, and impact on laboratory-based projects. The analysis in [129] in Australia shows the impact of government welfare support responses to COVID-19-infected people and businesses on mitigating potential unemployment, poverty and income inequality and the sustainability of such support measures.

Discussion. The existing modeling objectives, tasks and methods are highly preliminary, specific and limited. Expectations include macro-, meso- and micro-level modeling of economic impact by involving their economic-financial variables and activities, contrastive analysis with similar historical events and periods, and data-driven discovery of insights for a sustainable tradeoff between mitigation and economic growth in the new normal, to name a few.

7.4 Modeling COVID-19 Social Impact

The COVID-19 pandemic has had significant impact on public health, welfare, social, political and cultural systems, including restricting human activities, affecting people's well-being, causing an overwhelming burden on public health systems, reshaping sociopolitical systems, and disturbing social regularity such as incurring online information disorder. This section reviews the relevant modeling work on such social impacts.

Modeling the COVID-19 influence on human behaviors. In addition to the COVID-19-sensitive NPI influence on human activities as discussed in Section 7.1, SARS-CoV-2 and COVID-19 have fundamentally reshaped people's social activities and habits. For example, Baidu-based daily transportation behaviors and simple statistics were collected which show high-level mobility patterns such as visiting venues, origins, destinations, distances, and transport time during the COVID-19 epidemic in China [100]. In [83], large-scale mobile phone data such as call detail records, GPS locations, Bluetooth data and contact tracing apps are collected and analyzed by off-the-shelf tools to extract statistic metrics and patterns of behaviors, mobility and interactions. The results may inform population behaviors, individual contacts, movement paths and mobility patterns, and networking, in addition to evaluating the effectiveness of NPIs and informing COVID-19 responses such as contact tracing. In [95], social media data from Sina Weibo, the Baidu search engine, and 29 Ali e-commerce marketplaces were collected and analyzed using keyword-based linguistic inquiries and statistics like word frequencies and Spearman's rank correlation coefficient analysis. Keywords are extracted to show people's behavioral responses to COVID-19 outbreaks, public awareness and attention to COVID-19 protection measures, concerns about misinformation and rumors about ineffective treatments, and the correlation between risk perception and negative emotions.

Modeling the COVID-19 influence on public health systems. The sudden COVID-19 endemic or pandemic and its mysterious resurgence has resulted in the imperative, nonscheduled and overwhelming rationing demand on healthcare and medical professionals, public health and medical resources and supplies including oxygen, hospital beds and facilities, ICU facilities and ventilators, medical waste processing equipment, hygiene protection equipment such as medical masks and sanitization chemicals, and intervention materials and devices. How to plan, prioritize, ration and manage these resources, assess their supply/demand and effects to prioritized hotspots and regions and optimize their reorganization per local and global needs and population-wide well-being are some challenging issues to model and optimize. In [69], recommendations are made to allocate medical resources to both COVID-19 and non-COVID-19 patients to maximize benefits, prioritize health workers, avoid a first-come, first-served approach, in a way that is evidence-based and involves science and research.

Modeling the COVID-19 influence on sociopolitical systems. The COVID-19 influence on social and political systems is unprecedented. This influence extends to the confidence and trust in existing sociopolitical systems such as public and moral values, national interests, social welfare systems, human services, political relations, globalization, scientific exchange and collaborations, science-driven epidemic mitigation policies and strategies, and the impact on social governance and disaster management. For example, in [29], an identity fusion theory-based online sampling and a moral foundations theory-based computer simulation show the correlations between nationalism, religiosity, and anti-immigrant sentiment from a socio-cognitive perspective during the COVID-19 pandemic in Europe. The surveys undertaken in [119] show that the scientific uncertainty of COVID-19-oriented modeling and findings affect the public and political trust in science-based policy making in the US and suggest more careful science communications. The work in [210] evaluates the impact of COVID-19 on globalization and global health, in particular, mobility, trade, travel, event management, food and agriculture, and a pandemic vulnerability index quantitatively measures the potential impact on global health and the countries most impacted.

Modeling misinformation and disorder in the COVID-19 infodemic. The COVID-19 infodemic has been accompanied by a large volume of misinformation (partially or entirely inaccurate or misleading information), biased (polarized), questionable or unverified information, rumor and propaganda. Such information is harmful for correctly understanding, recognizing, intervening, and preventing the COVID-19 pandemic. Its diffusion is usually fast, its spread is often wide, and its impact is typically devastating. Modeling the COVID-19 misinformation and information disorder involves tasks such as detecting and ranking misinformation, classifying them, undertaking fact checks and cross-references, tracing their sources and transmission paths, discovering their diffusion and propagation networks and paths, and estimating their effects on the COVID-19 epidemic spread and control. For example, in [53], skip-gram is used to represent the words collected from Twitter, Instagram, YouTube, Reddit and Gab; the converted vector representations are then clustered by partitioning them around medoids and cosine distance-based similarity analysis to extract the topics of concern. An SIR model is then applied to estimate the basic reproduction number of the social media-based COVID-19 infodemic. A comparative analysis then estimates and compares the platform-dependent interaction patterns, information spread (w.r.t. reproduction rate), questionable and reliable information sourcing and differentiation, and rumor amplification across the above platforms. In [248], SVM classifies credible and misinformation from Twitter texts and a correlation analysis shows the predominant credible information on wearing masks and social distancing can lead their misinformation with a time lag. In [4], bivariate (ANOVA) and multivariate logistic regression identifies similar belief profiles of political orientation, religious commitment, and trust in science in survey-based narratives and compares the profiles of those who are disinformed or conspiratorial with scientific narratives. Further, the statistics on Weibo tweets show the COVID-19 misinformation evolution related to topics and events such as city lockdowns, cures, preventive measures, school reopening and foreign countries, the bias involving cures and preventive methods, and sentiment evolution such as fear of specific topics [126]. The work in [147] applies SVM, logistic regression and BERT to classify COVID-19 misinformation and counter-misinformation tweets, characterizes the type, spread and textual properties of counter-misinformation, and extracts the user characteristics of the citizens involved.

Discussion. Typical research on COVID-19 influence and impact modeling only involves local and regional COVID-19 data and their affected objects, hence the resultant conclusions are limited in the ability to indicate their applicability to general practice and broad pandemic

control. More robust results are expected to inform medical and public health policy-making on medication, business and society. No-to-rare outcomes are available on how NPIs influence the threshold and effects of COVID-19 vaccinations and herd immunity and on how to balance NPIs and economic and social revivification. It is difficult to find actionable evidence and guidelines on what policies should be implemented and what tradeoff is appropriate in balancing a COVID-19 outbreak and containing resurgence with economic and social business recovery.

Table 8. Examples of COVID-19 Influence and Impact Modeling.

Aspects	Objectives	Approaches	Data
NPI effect	on epidemic dynamics	SIR variants, statistical models, Bayesian hierarchical models, etc. [24, 63, 77, 121, 177, 185, 221]	COVID-19 case data, NPIs
	on public resources	SIR variants, statistical models, polynomial regressors, etc. [5, 73, 81]	Case data, NPIs, public resource (incl. healthcare) data
	on human behaviors	SIR variants, statistical models e.g. MCMC, relation modeling, etc. [75, 80, 83, 109, 118, 123, 250, 267]	Case data, NPIs, human activities (incl. mobile phone data and mobility), etc.
Psychological influence	on individual mental health	Psychology, systematic reviews, classic and neural NLP models e.g. BOW, LDA, SciBERT, Transformer variants, etc. [14, 158, 230, 245, 258]	Identity, social media data, news feeds, Q/A, surveys, instant messaging, behaviors, NPIs, etc.
	on public mental health	Psychology, systematic reviews, classic and neural NLP models e.g. BOW, LDA, SciBERT, Transformer variants, etc. [95, 158, 233, 245, 258]	Social media data, news feeds, Q/A, surveys, instant messaging, public emotion, activities and events, NPIs, etc.
	on mental health	Psychology, systematic reviews and meta-analyses, classic and neural NLP models, statistics, etc. [258]	Social media data, questionnaires, instant messaging, behavior and events, NPIs, etc.
Economic impact	on economic growth	Time series analysis, descriptive analytics, macroeconomic modes, relational models, etc. [40, 181, 236]	Economic data, case data, NPIs, etc.
	on workforce and sustainability	Descriptive analytics, time-series analysis, relational models, etc. [15, 129, 154]	Work and sustainability-related data, performance, employment, surveys, social welfare data, etc.
Social impact	on human behaviors	Descriptive analytics, pattern analysis, social media/network analysis, NLP models, etc. [83, 95, 100]	Public, online and household activities, gathering, mobility data, mobile phone data, social media data, etc.
	on public health systems	Descriptive analytics, relational models, etc. [69]	Public health and medical data, public hygiene data, case data, etc.
	on misinformation	Classifiers, classic and neural NLP models, social media/network analysis, sentiment/topic modeling, time-series analysis, outlier detection, etc. [4, 53, 126, 147, 248]	Social media data, news feeds, Q/A, cross/fact-check, etc.
	on sociopolitical systems	Descriptive analytics, sociopolitical methods, survey analysis, etc. [29, 119, 210]	Social and political data, case data, surveys, questionnaires, sociopolitical events, etc.

8 COVID-19 SIMULATION MODELING

Despite being a small focus (over 1.5k of the 22k publications on modeling), simulation is an essential means to understand, imitate, replicate and test the working mechanisms, the epidemic transmission processes, the evolution and mutation of COVID-19 and its virus SARS-CoV-2, the interactions and self-organization between factors, the effect of mitigation measures and various interior and contextual factors, and resource planning and optimization such as healthcare resource allocation. Typical simulation methods include dynamic systems, state-space modeling, discrete event simulation, agent-based modeling, reinforcement learning, Monte-Carlo simulation, and hybrid simulation [58]. Below, we summarize the relevant work on simulating the COVID-19 epidemic evolution and the effect of interventions and policies on COVID-19 epidemic development.

Simulating the COVID-19 epidemic evolution. One important but unclear question is how does the COVID-19 evolve over time in the community. What-if analyses can be applied to estimate infection case numbers and their evolution under various hypotheses tests [275]. Typical methods include SIR variants, statistical and mathematical models, e.g., introducing control measure-sensitive variables into such models to estimate their effects on infections, reproduction number, transmission rate, and outbreak control after implementing or relaxing certain interventions. For example, in [78], composite Monte Carlo simulation conducts the what-if analysis of future COVID-19 epidemic development possibilities on top of the estimation made by a polynomial neural network on COVID-19 cases, then fuzzy rule induction outputs decision rules to inform epidemic growth and control. In [82], an agent-based simulation system simulates a COVID-19 patient's demographic, mobility and infectious disease state (susceptible, exposed, seriously-infected, critically-infected, recovered, immune and dead) information and their dynamic interactions between each other (agents, i.e., people in epidemiology) in certain environments (home, public transport stations, and other places of interest), and evaluates the effect of adjusting individual and social distancing (separation) on epidemics (e.g., numbers in each state).

Simulating the policy effects on the COVID-19 epidemic. Another important task is to simulate how interventions, interior and external factors, and other policies and control measures of interest influence the dynamics of the COVID-19 epidemic. For example, a discrete-time and stochastic agent-based simulation system (Australian Census-based Epidemic Model) [44] incorporates 24 million software agents, where each agent mimics an Australian individual in terms of their demographics, occupation, immunity and susceptibility to COVID-19, contact rates in their social contexts, interactions, commuting and mobility patterns, and other aspects, which are informed by census data from the Australian government. The system evaluates various scenarios by adjusting the level of restrictions on case isolation, home quarantine, international air travel, social distancing and school closures and their effects on COVID-19 pandemic consequences in terms of the reproductive number, the generation period, the growth rate of cumulative cases, and the infection rate for children. The simulation provides evidence to help the Government understand how COVID-19 is transmitted and what policies should be implemented to control COVID-19 in Australia. In [141], an SIR model is extended by adding variables reflecting symptomatic infections and the quarantine of susceptibles, which then estimates the case development distribution as subexponential after implementing the quarantine. In [262], an attributed heterogeneous information network incorporates the representations of external information about the COVID-19 disease features, the population's demographic features, mobility and

public perception of sentiment into a GAN model, which then assesses the hierarchical community-level risks of COVID-19 to inform interventions and minimize disruptions.

Discussion. Although we mention many aspects and questions that could be (better) addressed by simulation, very limited research is available in this direction. In addition to the above two aspects closely relevant to COVID-19 epidemic dynamics, other important topics include simulating the mutation and resurgence of the coronavirus and COVID-19 in communities with different social, ethnic and economic conditions; the influence of individual and compound COVID-19-sensitive policies on social, economic and psychological aspects; and the tradeoff between the strength and width of mitigation strategies and their impact on society and the economy.

9 COVID-19 HYBRID MODELING

Hybrid COVID-19 modeling can be categorized into the following families: (1) multi-objective modeling: to address multiple problems and multiple business and modeling objectives at the same time, such as jointly understanding COVID-19 epidemic dynamics and the corresponding effective NPI policies; (2) multi-task modeling: to handle multiple modeling tasks, e.g., simultaneously forecasting daily confirmed, death and recovered case numbers; (3) multisource (multimodal etc.) COVID-19 data modeling: to involve multiple sources of internal and external data for modeling, e.g., supplementing environmental and demographic data with case numbers and complementing case numbers with medical imaging and social mobility data; (4) hybrid methods for COVID-19 modeling: typically by sequentializing (i.e., multi-phase) or parallelizing multiple tasks or methods from different disciplines and areas, e.g., integrating statistical methods, shallow or deep learning methods, and evolutionary computing methods into compartmental models; and (5) hybrid modeling with multi-methods from various disciplines on multisource COVID-19 data for multi-objective or multi-task modeling.

COVID-19 multi-objective modeling is commonly seen in COVID-19 modeling, as shown in Sections 4-8, where, multiple business problems and learning objectives are involved in one research or case study. Examples are forecasting COVID-19 transmission and its sensitivity to external factors such as the patients' age groups, hygiene habits and environmental factors; modeling the influence of NPIs and people's ethnic conditions on case movements; modeling the influence of NPIs on both case trends and public psychological health; and survival/mortality rate estimation and the influence analysis of dependent factors such as the patients' health conditions. Typical methods include multivariate analysis, probabilistic compartmental models, simulation systems, multi-objective evolutionary learning methods, and DNN variants. For example, in [183], a regression model estimates the relations between reproduction number and environment factors and human movements. In [63], Bayesian inference of an SIR model infers the effect of various interventions on new infections. In [174], an SEIR models the relations between case trends and epidemic conditions, socioeconomic effect, and interventions. In [258], systematic reviews and meta-analyses review the work on the relations between COVID-19 symptom severity, risk factors and public emotions.

COVID-19 multisource data modeling serves various purposes such as predicting COVID-19 epidemic spread and transmission, medical diagnosis and treatment, and government and community interventions by combining data from respective modalities, sources or views. Examples of multisource data are combining COVID-19 case numbers with NPI data; people's demographics, health conditions, mobility, social and business activities, social networking and media information; health and medical records, diagnosis information, treatments, pharmaceutical interventions, and pathological tests; social and public activities

and events, economic data, and sociopolitical data; and online, social media and mobile apps-based messaging, news, Q/A, and discussion groups. Typical methods include data fusion-based learning, mixed representations-based learning, clustering and classification on mixed data types, DNN variants, etc. [107]. For example, a novel variational-LSTM autoencoder model in [101] predicts the coronavirus spread in various countries by integrating historical confirmed case numbers with urban factors (about location, urban population, population density, and fertility rate) and governmental measures and responses (school, workplace and public transport closures, public events cancellation, contact tracing, public information campaigns, international travel controls, fiscal measures, and investment in health care and vaccines). In [142], COVID-19 case numbers and weather data are combined to analyze the correlation between COVID-19 confirmed cases and mortality and weather factors. NLP methods can extract and analyze related news, which are then input to LSTM networks to update the infection rate in a susceptible–infected epidemic model [273], which shows to beat the susceptible–infected epidemic model and its combination with LSTM. In [216], coupling LSTM with an epidemic model forecasts COVID-19 spread on case data, population density and mobility.

COVID-19 hybrid methods integrate various methods for single or multiple-objective/task/source learning. In addition to ensemble learning by integrating the results from multiple learners such as ensemble trees and XGBoost, often multiple methods are sequentially involved to learn specific tasks or data over phases; other common tasks are to integrate compartmental models with other methods such as statistical models, classifiers and DNNs for the improved forecasting of COVID-19 epidemic dynamics and attributes. For example, in [266], a hybrid model predicts the infected and death cases by integrating a genetic algorithm to optimize infection rates and integrating LSTM for parameter optimization into a modified susceptible-infected-quarantined-recovered (SIQR) epidemic model. In [41], a regression tree combined with wavelet transform predicts COVID-19 outbreak and assesses its risk on case numbers. In [16], a baseline method generates a granular ranking (discrimination) of severe respiratory infection or sepsis on the medical records of the general population, then a decision-tree-based gradient boosting model adjusts the former predicted results in subpopulations by aligning it with the published aggregate fatality rates. In addition to the aforementioned methods, other methods and tasks e.g. for innovative pandemic responses are available in the literature. Examples are automated primary care tools to alleviate the shortage of healthcare workers [219], expert systems and chatbots for symptom detection and lessening the mental health burden [150], IoT and smart connecting tools to prevent outbreaks, remotely monitoring patients, and prompting enforcement of guidelines and administrative orders to contain future outbreaks [88].

Discussion. Though various methods of hybridization have been summarized in this section, the relevant research is not systematic, comprehensive, or substantial. This observation applies to hybrid data, hybrid tasks, and hybrid methods. The complex characteristics and challenges of both the virus and disease and of modeling their problems and data, as discussed in Sections 2.2, 2.1 and 2.3 , are substantial. Though overwhelming efforts have been made in modeling COVID-19, the above complexities require significant novel developments through synergizing problems, data, and modeling techniques.

10 DISCUSSION AND OPPORTUNITIES

In the above review of each category of COVID-19 modeling techniques, a brief discussion has been provided on the main limitations, gaps and opportunities in those areas. Here, we

expand this specific discussion to broad major gaps in the research on modeling COVID-19. Further, we discuss various open issues and opportunities for future research.

10.1 Modeling Gap Analyses

Two major aspects of modeling gaps include: the gaps in understanding the virus and disease nature, and the gaps in modeling their complexities.

10.1.1 Gaps in understanding the problem nature. Since the virus is new and unique, we have limited knowledge on all aspects of the SARS-CoV-2 virus and COVID-19, such as virus characteristics, epidemiological attributes and dynamics, socioeconomic influence, and virus mutations, and so on. Specifically, our poor understanding of the intrinsic and intricate pathological, biomedical and epidemiological attributes of the evolving SARS-CoV-2 and COVID-19 systems limits the modeling attempts and contributions. As a result, our understanding of the virus and disease is still *insufficient* without substantial knowledge and comprehensive evidence on the system complexities; it is *biased* to specific data, conditions or settings; it is *shallow* without deep insights into the virus and disease nature; and it is *partial* without a full picture of the SARS-CoV-2 and COVID-19 complexities, in understanding the COVID complex systems and their data complexities [35, 247].

To address these issues, the modeling has to start with building a comprehensive understanding of the virus nature and the fundamental complexities of the COVID-19 complex systems. Of the many questions to explore, we highlight the following important unknowns, which require cross-disciplinary scientific explorations by integrating medical science, virology, bio-medicine, and data-driven discovery.

- *The hidden nature of SARS-CoV-2 and COVID-19*: How does the virus interact with human and animal hosts? How does the virus genetic system look like? What are the epidemiological attributes of the virus and the disease characteristics after infections under different contexts, e.g., demographics, community (population) scenarios, ethnics, seasonality, and weather conditions? What are the high risk factors or high risk factor combinations of infection and mortality? What causes the different levels of symptom onset and differs asymptomatic and mild symptomatic infections from severe symptomatic infections?
- *The mysterious mutation mechanisms of SARS-CoV-2*: What genomic and pathological factors determine the virus transform from one generation to another and over time? What genomic and pathological mechanisms drive the variations? Why the genetic variants differ from regions to regions and between population ethnics?
- *The influence of external factors on virus spread and evolution*: How does the virus evolve under different ethnic, environmental and intervention (including pharmaceutical and non-pharmaceutical) contexts? How do external factors such as demographics, ethnics, environment, and healthcare quality influence the virus transformation? How do personal hygiene, public health systems, public activities, population mobility and daily commuting affect the virus spread and evolution?
- *The virus adaption to vaccine and drug*: What are the relations between key factors such as the various vaccines and drugs available for treating COVID-19, the increasing virus mutants and their more contagious new strains, the widespread delta strain outbreaks, and unpredictable resurgence? How do the virus variants adapt to the vaccines and drugs? How does the vaccination affect the virus evolution?
- *The herd immunity vs. zero tolerance for the virus*: What is the new COVID normalcy, i.e., should a 'zero tolerance for the virus' be the target and eventually remove the

SARS-COV-2 like we did for SARS or shall 'quotidian existence' be the new normal such that humans live with the virus as influenza? For the former, what is the herd immunity threshold for a manageable normal of living with the virus? Where is the manageable risk level of balancing the vaccination rate, public health system capacity, ethnic and community conditions, and acceptable daily numbers of infections and deaths? How does the regional inequality of vaccinations and public health systems affect the global recovery?

10.1.2 Gaps in modeling the system complexities. The modeling gaps come from both a poor understanding of the virus and disease nature and the limitations in modeling their characteristics and complexities. On one hand, even though massive efforts have been made in modeling COVID-19, the existing modeling work is still in its early stage. The weaknesses and limitations of the existing work lie in

- an average description of the population-wise coronavirus and the disease's epidemiological characteristics and observations after applying mitigation and control measures, no fine-grained and microlevel analysis and findings are available;
- a direct application of existing (even very simple and classic) modeling methods without COVID-specific and optimal modeling mechanisms, typically by applying overparameterized or independently pretrained deep neural models or complex statistical and compartment models on low-quality and often small COVID data;
- simple data-driven modeling purely motivated by applying advanced models (typically deep models) on COVID-19 data without a deep incorporation of domain and external knowledge and factors; and
- a purposeful design without a comprehensive design or exploration of the multi-faceted COVID-19 characteristics and challenges in one framework or system.

On the other hand, the general applications of existing methods also present unsuitability and incapability in tackling the complexities of the complex virus and disease. Table 9 compares the major modeling methods and their pros and cons in modeling COVID-19. Consequently, it is common that the existing models and their modeling results

- often only reflect a specific population or cohort-based average estimation or hypothesis of epidemic transmission, losing a personalized applicability to individual cases or scenarios, making it difficult to undertake personalized treatment;
- are too specific to expand to other countries and scenarios, hard to reproduce and transfer to other regions without (significant) changes, making it unsuitable for broad applications;
- over- or under-fit the given data and hypothesis settings, they are difficult to validate in a fine-grained way and have weak robustness or generalization for a general but deep understanding of the problems; and
- lack the ability and capacity to disclose the intrinsic nature and general insights about the SARS-CoV-2 virus, COVID-19 disease, and their interventions.

10.2 Opportunities for AI and Data Science

There are enormous opportunities and future directions in modeling COVID-19, including (1) fundamentally characterizing the system complexities, (2) addressing the aforementioned limitations of existing work, and (3) exploring new directions and alternatives. These are particularly valid for AI, data science and machine learning, which play a dominating role in the data-driven COVID-19 modeling.

Table 9. Comparison of COVID-19 Modeling Methods.

Methods	Pros	Cons
Time-series analysis	Temporal representations and interaction modelings of periodic and aperiodic components, relations and trends of COVID-19 cases at different states (e.g., new, susceptible, infected, recovered and death) and external temporal factors	Weak modeling power involving other rich factors (e.g., demographics and clinical attributes) and complex data characteristics (e.g., nonstationarity) and discovering the insight of COVID-19 driving forces and interventions
General machine learning	Multifaceted factor and relation analysis, outlier detection, profiling, classification, prediction and impact analysis for disease diagnosis and case detection on small and poor-quality COVID-19 data, etc.	Poor modeling of weak but complex interactions, couplings, high-dimensional dependencies, heterogeneity, nonstationarity and other data challenges in multisource COVID-19 data
Statistical modeling	Modeling distributional dynamics, uncertainty and dependency with analytical explanation and parameter settings	Requires informative prior knowledge, high modeling and computation complexity on poor-quality COVID data
Epidemiological modeling	Built on epidemic knowledge, straightforward but domain-friendly and explainable hypothesis test, strong characterization of infection processes, state transitions, and parameter selection	Captures complex epidemic transmission characteristics, factors, causal relations and processes in COVID-19 developments; hypothesis of homogeneous disease transmissions
Deep learning	Performs well with large and complex COVID-19 data (e.g., medical imaging)-based case and disease prediction and identification with annotated samples; pretrained model easily adaptable to new tasks	Requires annotated ground-truth of COVID-19 learning targets, easy to overfit small COVID-19 data, vulnerable results, poor interpretability, high computational cost
Simulation	Imitates and replicates complex COVID-19 mechanisms and processes, cost-effective, reproducible and risk-averse, manually controllable for purposeful test and optimization	Proper knowledge and hypotheses about COVID-19 transmission and factor interactions, high experimental complexity, inactionable for evolving and random real-life scenarios
Hybrid methods	Flexible and powerful in selecting and combining small COVID-19 multisource data and relevant multi-methods on demand for combined COVID-19 learning tasks and data	Understands constituents for their best ensemble to address specific COVID-19 challenges with appropriate design complexity, less flexible in combination optimization and explanability

10.2.1 Characterizing the system complexities. To discover the mysteries of the COVID virus and disease, the most important opportunities come from understanding their nature and system characteristics and complexities, as discussed in Sections 2.1 and 10.1.1. Combining the domain-driven and data-driven thinking and techniques, there are various directions in characterizing the problem nature and system complexities:

- extracting, representing and distinguishing observable and latent factors and metrics to describe the epidemiological, biological (genomic), medical (clinical and pathological) and social attributes, liveliness and dynamic processes of the virus, virus mutations, the disease and its variants from other similar viruses and diseases;

- identifying and characterizing external entities and factors (e.g., drugs, vaccines, ethnics, environment) and how they interact with the virus and disease and influence their evolution;
- characterizing and simulating the diversified (e.g., explicit vs. implicit, global vs. local, domain-specific vs. general) interactions and relations between the above-extracted explicit and implicit internal and external factors and their dynamics;
- quantifying and simulating the virus and disease's system dynamics and genetic mechanisms (e.g., self-organization, genomic expression, genetic crossover and mutation, interaction and adaptation with external environment) in terms of temporal, dependent variables and major transformations;
- simulating and quantifying the virus parasitism, interactions, adaptation and evolution with human, animal and living hosts in a large scale.

10.2.2 Enhancing COVID-19 modeling. To address the modeling gaps in Section 10.1.2 and those rarely and poorly explored areas and challenges in Section 2.2, we here highlight the following major directions.

Rarely to poorly addressed areas. First, opportunities to focus on the areas rarely or poorly addressed in the existing COVID-19 modeling include: (1) characterizing the effective NPIS on the variants of the SARS-CoV-2 virus and comparing them with those on the original strains; (2) quantifying the effects of COVID-19 vaccines, pharmaceutical and NPI interventions on the infection control, mobility, mental health, society and the economy, e.g., the efficacy of vaccinated population percentage on herd immunity, and the effect of variable close-contact interactions and individual actions on epidemic de-escalating; (3) balancing the NPI strength and the socioeconomic recovery, e.g., modeling the effect of full vs partial business close-downs and border control on virus confinement at different stages and for different sectors, and characterizing the effect of increasing daily commuting and workforce movement vs working-from-home and telecommuting on the virus confinement; (4) capturing the temporally evolving interplay and interactions between virus propagation and external interventions; and (5) modeling target problems by systemically coupling relevant multisource data and multiple modeling techniques, e.g., by involving pathogen-related, societal, environmental and racist factors and the disparities between developing and developed countries, age groups, and races.

Hybrid modeling. Second, the hybridization of relevant data and techniques offers significant opportunities to improve and expand the existing modeling capacity and results. Examples include integrating (1) coarse-grained and fine-grained modeling, e.g., epidemic modeling by SIR variants to inform further specific NPI's effect analysis; (2) static and dynamic modeling, e.g., from population-based static epidemic modeling to specific NPI-varying and time-varying case forecasting; (3) observable and hidden factors and relations, e.g., multisource-based attributed modeling with deep abstraction and representation of interactions between the multisource factors; (4) local-to-micro-level and global-to-macro-level factors, e.g., involving patient clinical and demographic records with their environmental and socioeconomic context in survival and mortality prediction and medical resource planning; and (5) domain, data and models for domain-specific, interpretable, evidence-based and actionable findings. These typically involve compound modeling objectives, multisource data, and multi-method ensembles.

Enhanced COVID-19 modeling. Third, another set of new opportunities is to undertake sequential or multi-phase modeling, such as (1) from coarse-grained to fine-grained modeling: e.g., applying epidemic models like SIR and SEIR on COVID-19 in the initial stage and

then modeling the impact of NPIs, the mobility and behaviour change of a population on epidemic dynamics; (2) from static to dynamic modeling: e.g., testing constant epidemic parameters and then time-varying settings such as NPI-sensitive varying parameters; and (3) from core to contextual factors: e.g., modeling epidemic processes on case data and then involving pathogen-related, societal and environment (like temperature and humidity) variables to model their influence on epidemic movements.

Lastly, alternative opportunities exist by (1) developing COVID-19-specific modeling methods, benchmarks and evaluation measures to address the virus and disease's challenges and their data challenges for an intrinsic interpretation of the virus and disease nature and dynamics; (2) trans-disciplinarily integrating the relevant domain knowledge and hypotheses from biomedical science, pathology, epidemiology, statistics and computing science to address multifaceted challenges in the virus, disease, data and modeling and to form a comprehensive understanding of the virus and disease; (3) defining multifaceted modeling objectives and tasks to directly address comprehensive epidemiological, clinical, social, economic or political concerns and their challenges in one framework; and (4) ethical and explainable COVID-19 modeling with privacy-preserving and distributed heterogeneous information integration, augmentation, representation and learning by utilizing personal computing devices (e.g., smart phones) and cloud analytics.

10.2.3 Exploring new opportunities. In addition to many specific perspectives, such as hybridizing modeling objectives, data and methods in Section 10.2.2 and addressing the shortcomings in Section 10.1.2, we here highlight some other opportunities that may particularly benefit (from) AI, data science and machine learning advances.

Quantifying the virus nature and complexities. An imperative yet challenging task for the AI, data science and machine learning communities is to 'quantify' the nature and complexities of the virus and disease and address the fundamental questions on the virus nature and complexities raised in Section 10.1.1. Building on multi-disciplinary knowledge such as on epidemiology, genetic computing and theories of complex systems, large-scale agent-based epidemic simulation systems are demanding to test and improve genetic, clinical and epidemiological hypotheses and knowledge about the virus and characterize the virus' genetic evolution mechanisms.

- *Large-scale COVID epidemiological dynamics*: to obtain quantitative results and verification on questions in Section 10.1.1, such as how does a virus evolve, cross-over and mutate; how do billions of coronaviruses interact, compete, and transform over time; and how do environmental factors affect the virus life and genetic evolution.
- *Large-scale human-virus interactions*: to characterize the experiments in relation to questions such as how does the full population of a country interact with the virus under their varied demographic profiles, hygiene protection habits, health conditions, vaccination conditions, mobility settings, etc. by mimicking their physical census data and circumstances in the real world; what is the vaccination threshold to build the herd immunity for a country by considering their specific circumstances; and to compare the simulation results with the reality of various waves of COVID-19 epidemic occurred in the country.
- *Large-scale intervention influence on human-virus interactions*: to quantify and evaluate how does the residents in a country respond to various intervention policies and restrictions on public and household activities over time; how does enforcing or relaxing interventions and restrictions affect the virus spread, infection numbers, and public

heath system quality; and what vaccination and intervention preconditions make business reopening possible, etc.

Data-driven discovery of COVID mysteries. There is increasing and comprehensive sources of COVID-19 data available publicly and through private providers. Data-driven discovery on this COVID-19 data can substantially leverage other domain-specific research on COVID to disclose the mysteries of COVID.

- *COVID data genomics*: forming the data genomics of COVID for a person, country, community or task by automatically extracting and fusing all possibly relevant data, e.g., contacts, personal health, mobility, clinical reports, exposure to infected people, and household and public activities in a privacy-preserving manner.
- *COVID data augmentation*: developing new techniques to address the various data quality issues embedded in the data, as discussed in Section 2.2 and novel augmented analytics and learning methods to directly learn from poor quality COVID data.
- *All-purpose representation of COVID attributes*: learning the representations on all-relevant COVID data that can be used to describe the full profile of COVID and support diverse learning objectives and tasks in an ethical and privacy-preserving manner.
- *Automated COVID screening and diagnosis*: developing techniques and systems to automatically detect, screen, predict and alert potential infection of the virus and disease on the COVID data genomics.
- *Virus detection and interaction modeling*: developing personal IoT assistants and sensors to detect the virus, trace its movement and its origin and visualize the 'COVID net' showing its propagation paths, interactions and networking with other viruses and hosts.
- *COVID knowledge graph*: generating knowledge graph showing the ontology about the virus; ontological connections between concepts on the virus; relations between knowledge on the virus and its protection, intervention, treatment and influence; and important highlights such as new knowledge discovered and misinformation detected.
- *COVID safety and risk management*: developing systems and tools (including mobile apps) for personal and organizational daily management of their COVID safety and risk, e.g., COVID-safe physical and emotional health management, mobility planning, risk estimation and alerting, infection tests, immunity estimation, and compliance management.
- *Metasynthetic COVID decision-support systems*: developing evidence-based decision support systems to fuse real-time and relevant big data, simulate and replay the outbreaks, estimate NPI effects, discover evidence from data and modeling, engage domain experts in the modeling and optimization processes, generate recommendations for decision-making, and support the data-driven analytics and management of severe disasters and emergencies.

11 CONCLUDING REMARKS

The COVID-19 pandemic's short-to-long-term influence and impact on public health (both physical and mental health), human daily life, global society, economy and politics is unprecedented, lasting, evolving yet quantified and verified. This review paints a comprehensive picture of the field of COVID-19 modeling. The multidisciplinary methods including mathematical modeling, AI, data science and deep learning on COVID-19 data have deepened our understanding of the SARS-CoV-2 virus and its COVID-19 disease's complexities and

nature; contributed to characterizing their propagation, evaluating and assisting in the effect of preventive and control measures, detecting COVID-19 infections, predicting next outbreaks, and estimating the COVID-19 influence and impact on psychological, economic and social aspects.

The review also highlights the important demands and significant gaps in deeply and systemically characterizing COVID-19-related problems and complexities; and developing effective, interpretable and actionable models to characterize, measure, imitate, evaluate and predict broad-based challenges and problems and to proactively and effectively intervene in them. Such COVID-19 modeling research proposes many significant challenges and opportunities to the multidisciplinary modeling communities in the next decade. These include not only immediately gaining intrinsic knowledge and proactive insight about the evolving coronavirus and its disease outbreak, infection, transmission, influence and intervention; but also preparing to tackle future global health, financial, economic, security-related and other black-swan events and disasters.

ACKNOWLEDGMENTS

REFERENCES

[1] Aniruddha Adiga, Jiangzhuo Chen, Madhav Marathe, Henning Mortveit, Srinivasan Venkatramanan, and Anil Vullikanti. 2020. Data-driven modeling for different stages of pandemic response. *Journal of the Indian Institute of Science* (2020), 1–15.
[2] Parnian Afshar, Shahin Heidarian, Farnoosh Naderkhani, Anastasia Oikonomou, Konstantinos N Plataniotis, and Arash Mohammadi. 2020. Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images. *Pattern Recognition Letters* 138 (2020), 638–643.
[3] Adeel Afzal. 2020. Molecular diagnostic technologies for COVID-19: Limitations and challenges. *Journal of Advanced Research* 26 (2020), 149–159.
[4] Jon Agley and Yunyu Xiao. 2021. Misinformation about COVID-19: evidence for differential latent profiles and a strong association with trust in science. *BMC Public Health* 21, 89 (2021).
[5] Maira Aguiar, Eduardo Millán Ortuondo, Joseba Bidaurrazaga Van-Dierdonck, Javier Mar, and Nico Stollenwerk. 2020. Modelling COVID 19 in the Basque Country from introduction to control measure response. *Scientific reports* 10, 1 (2020), 1–16.
[6] Amir Ahmad, Sunita Garhwal, Santosh Kumar Ray, Gagan Kumar, Sharaf Jameel Malebary, and Omar Mohammed Barukab. 2020. The number of confirmed cases of covid-19 by using machine learning: Methods and challenges. *Archives of Computational Methods in Engineering* (2020), 1–9.
[7] Talha Burak Alakus and Ibrahim Turkoglu. 2021. A Novel Protein Mapping Method for Predicting the Protein Interactions in COVID-19 Disease by Deep Learning. *Interdisciplinary Sciences: Computational Life Sciences* 13, 1 (2021), 44–60.
[8] Alberto Aleta and Yamir Moreno. 2020. Age differential analysis of COVID-19 second wave in Europe reveals highest incidence among young adults. *medRxiv* (2020).
[9] Abdullah M Almeshal, Abdulla I Almazrouee, Mohammad R Alenizi, and Saleh N Alhajeri. 2020. Forecasting the spread of COVID-19 in Kuwait using compartmental and logistic regression models. *Applied Sciences* 10, 10 (2020), 3402.
[10] Hakan Andersson and Tom Britton. 2012. *Stochastic epidemic models and their statistical analysis*. Vol. 151. Springer Science & Business Media.
[11] Ioannis D Apostolopoulos and Tzani A Mpesiana. 2020. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine* 43, 2 (2020), 635–640.

[12] Ashwin Aravindakshan, Jörn Boehnke, Ehsan Gholami, and Ashutosh Nayak. 2020. Preparing for a future COVID-19 wave: insights and limitations from a data-driven evaluation of non-pharmaceutical interventions in Germany. *Scientific reports* 10, 1 (2020), 1–14.

[13] Joan L Aron and Ira B Schwartz. 1984. Seasonality and period-doubling bifurcations in an epidemic model. *Journal of theoretical biology* 110, 4 (1984), 665–679.

[14] Meysam Asgari-Chenaghlu, Narjes Nikzad-Khasmakhi, and Shervin Minaee. 2020. Covid-Transformer: Detecting COVID-19 Trending Topics on Twitter Using Universal Sentence Encoder. *CoRR* abs/2009.03947 (2020).

[15] Lingfeng Bao, Tao Li, Xin Xia, Kaiyu Zhu, Hui Li, and Xiaohu Yang. 2021. How does Working from Home Affect Developer Productivity? – A Case Study of Baidu During COVID-19 Pandemic. *arXiv* (2021), 1–23. arXiv:2005.13167

[16] Noam Barda, Dan Riesel, Amichay Akriv, Joseph Levy, Uriah Finkel, Gal Yona, Daniel Greenfeld, Shimon Sheiba, Jonathan Somer, Eitan Bachmat, Guy N Rothblum, Uri Shalit, Doron Netzer, Ran Balicer, and Noa Dagan. 2020. Developing a COVID-19 mortality risk prediction model when individual-level data are not available. *Nature Communications* 11, 4439 (2020), 1–9. Issue 1.

[17] Sanhita Basu, Sushmita Mitra, and Nilanjan Saha. 2020. Deep learning for screening covid-19 using chest x-ray images. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. 2521–2527.

[18] Bo Ram Beck, Bonggun Shin, Yoonjung Choi, Sungsoo Park, and Keunsoo Kang. 2020. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. *Computational and structural biotechnology journal* 18 (2020), 784–790.

[19] Hemant Bherwani, Saima Anjum, Suman Kumar, Sneha Gautam, Ankit Gupta, Himanshu Kumbhare, Avneesh Anshul, and Rakesh Kumar. 2020. Understanding COVID-19 transmission through Bayesian probabilistic modeling and GIS-based Voronoi approach: a policy perspective. *Environment, Development and Sustainability* (2020), 1–19.

[20] Matthew J. Binnicker. 2020. Challenges and Controversies to Testing for COVID-19. *Journal of Clinical Microbiology* 58 (2020), e01695–20. Issue 11.

[21] E. Bontempi. 2021. The Europe second wave of COVID-19 infection and the Italy "strange" situation. *Environmental Research* 193, 110476 (2021), 1–7.

[22] Adam L Booth, Elizabeth Abels, and Peter McCaffrey. 2020. Development of a prognostic model for mortality in COVID-19 infection using machine learning. *Modern Pathology* (2020), 1–10.

[23] Fred Brauer, Carlos Castillo-Chavez, and Zhilan Feng. 2019. *Mathematical Models in Epidemiology* (1st ed. ed.). Springer.

[24] Jan M Brauner, Sören Mindermann, Mrinank Sharma, David Johnston, John Salvatier, Tomáš Gavenčiak, Anna B Stephenson, Gavin Leech, George Altman, Vladimir Mikulik, Alexander John Norman, Joshua Teperowski Monrad, Tamay Besiroglu, Hong Ge, Meghan A. Hartwick, Yee Whye Teh, Leonid Chindelevitch, Yarin Gal, and Jan Kulveit. 2021. Inferring the effectiveness of government interventions against COVID-19. *Science* 371, eabd9338 (2021). Issue 6531.

[25] Davide Brinati, Andrea Campagner, Davide Ferrari, Massimo Locatelli, Giuseppe Banfi, and Federico Cabitza. 2020. Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: A Feasibility Study. *Journal of medical systems* 44, 135 (2020), 1–12. Issue 8.

[26] Chloë Brown, Jagmohan Chauhan, Andreas Grammenos, Jing Han, Apinan Hasthanasombat, Dimitris Spathis, Tong Xia, Pietro Cicuta, and Cecilia Mascolo. 2020. Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data. In *KDD'20*. 3474–3484.

[27] Jin Bu, Dong-Dong Peng, Hui Xiao, Qian Yue, Yan Han, Yu Lin, Gang Hu, and Jing Chen. 2020. Analysis of meteorological conditions and prediction of epidemic trend of 2019-nCoV infection in 2020. *MedRxiv* (2020).

[28] Diana Buitrago-Garcia, Dianne Egli-Gany, Michel J Counotte, Stefanie Hossmann, Hira Imeri, Aziz Mert Ipekci, Georgia Salanti, and Nicola Low. 2020. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. *PLoS medicine* 17, 9 (2020), e1003346.

[29] Josh Bullock, Justin E. Lane, Igor Miklousic, and F. LeRon Shults. 2020. Modelling Threat Causation for Religiosity and Nationalism in Europe. *CoRR* abs/2009.09425 (2020).

[30] Oyungerel Byambasuren, Magnolia Cardona, Katy Bell, Justin Clark, Mary-Louise McLaws, and Paul Glasziou. 2020. Estimating the Extent of True Asymptomatic COVID-19 and Its Potential for Community Transmission: Systematic Review and Meta-Analysis. *Journal of the Association of Medical Microbiology and Infectious Disease Canada* 5 (2020), 223–234. Issue 4.

[31] Giacomo Cacciapaglia, Corentin Cot, and Francesco Sannino. 2020. Second wave COVID-19 pandemics in Europe: a temporal playbook. *Scientific reports* 10, 1 (2020), 1–8.

[32] Giacomo Cacciapaglia, Corentin Cot, and Francesco Sannino. 2020. Second wave COVID-19 pandemics in Europe: A temporal playbook. *Scientific Reports* 10, 15514 (2020).

[33] Junzhe Cai and Peter Z. Revesz. 2020. A novel spatio-temporal interpolation algorithm and its application to the COVID-19 pandemic. In *IDEAS'2020*, Bipin C. Desai and Wan-Sup Cho (Eds.). 9:1–9:10.

[34] Longbing Cao. 2013. Combined mining: Analyzing object and pattern relations for discovering and constructing complex yet actionable patterns. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* 3, 2 (2013), 140–155.

[35] Longbing Cao. 2015. *Metasynthetic Computing and Engineering of Complex Systems*. Springer.

[36] Longbing Cao. 2018. *Data Science Thinking: The Next Scientific, Technological and Economic Revolution*. Springer International Publishing.

[37] Longbing Cao and et al. 2021. What does COVID-19 modeling tell us about the pandemic? *Technical Report* (2021).

[38] Longbing Cao, Wenfeng Hou, and Qing Liu. 2021. How have global scientists responded to modelling COVID-19? *Technical Report* (2021).

[39] Hal Caswell. 2001. *Matrix population models* (2ed ed.). Sinauer.

[40] Indranil Chakraborty and Prasenjit Maity. 2020. COVID-19 outbreak: Migration, effects on society, global environment and prevention. *Science of the Total Environment* (2020), 138882.

[41] Tanujit Chakraborty and Indrajit Ghosh. 2020. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. *Chaos, Solitons & Fractals* 135 (2020), 109850.

[42] Jasper Fuk-Woo Chan, Shuofeng Yuan, Kin-Hang Kok, Kelvin Kai-Wang To, Hin Chu, Jin Yang, Fanfan Xing, Jieling Liu, Cyril Chik-Yan Yip, Rosana Wing-Shan Poon, Hoi-Wah Tsoi, Simon Kam-Fai Lo, Kwok-Hung Chan, Vincent Kwok-Man Poon, Wan-Mui Chan, Jonathan Daniel Ip, Jian-Piao Cai, Vincent Chi-Chung Cheng, Honglin Chen, Christopher Kim-Ming Hui, and Kwok-Yung Yuen. 2020. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *The Lancet* 395, 10223 (2020), 514–523.

[43] Tej Bahadur Chandra, Kesari Verma, Bikesh Kumar Singh, Deepak Jain, and Satyabhuwan Singh Netam. 2021. Coronavirus disease (COVID-19) detection in Chest X-Ray images using majority voting based classifier ensemble. *Expert Syst. Appl.* 165 (2021), 113909.

[44] 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. *Nature communications* 11, 1 (2020), 1–13.

[45] Cindy H. Chau, Jonathan D. Strope, and William D. Figg. 2020. COVID-19 Clinical Diagnostics and Testing Technology. *Pharmacotherapy* 40 (2020), 857–868. Issue 8.

[46] Biqing Chen, Hao Liang, Xiaomin Yuan, Yingying Hu, Miao Xu, Yating Zhao, Binfen Zhang, Fang Tian, and Xuejun Zhu. 2020. Roles of meteorological conditions in COVID-19 transmission on a worldwide scale. *MedRxiv* (2020).

[47] Jianguo Chen, Kenli Li, Zhaolei Zhang, Keqin Li, and Philip S Yu. 2020. A survey on applications of artificial intelligence in fighting against covid-19. *arXiv:2007.02202* (2020).

[48] Ruchong Chen, Wenhua Liang, Mei Jiang, Weijie Guan, Chen Zhan, Tao Wang, Chunli Tang, Ling Sang, Jiaxing Liu, Zhengyi Ni, Yu Hu, Lei Liu, Hong Shan, Chunliang Lei, Yixiang Peng, Li Wei, Yong Liu, Yahua Hu, Peng Peng, Jianming Wang, Jiyang Liu, Zhong Chen, Gang Li, Zhijian Zheng, Shaoqin Qiu, Jie Luo, Changjiang Ye, Shaoyong Zhu, Xiaoqing Liu, Linling Cheng, Feng Ye, Jinping Zheng, Nuofu Zhang, Yimin Li, Jianxing He, Shiyue Li, and Nanshan Zhong. 2020. Risk factors of fatal outcome in hospitalized subjects with coronavirus disease 2019 from a nationwide analysis in China. *Chest* 158, 1 (2020), 97–105.

[49] Yi-Cheng Chen, Ping-En Lu, Cheng-Shang Chang, and Tzu-Hsuan Liu. 2020. A time-dependent SIR model for COVID-19 with undetectable infected persons. *IEEE Transactions on Network Science and Engineering* 7, 4 (2020), 3279–3294.

[50] Cindy Cheng, Joan Barceló, Allison Spencer Hartnett, Robert Kubinec, and Luca Messerschmidt. 2020. COVID-19 government response event dataset (CoronaNet v. 1.0). *Nature human behaviour* 4, 7 (2020), 756–768.

[51] Maria Chikina and Wesley Pegden. 2020. Modeling strict age-targeted mitigation strategies for COVID-19. *PloS one* 15, 7 (2020), e0236237.

[52] Alexander Chudik, Kamiar Mohaddes, Mohammad Hashem Pesaran, M Raissi, and A Rebucci. 2020. Economic consequences of Covid-19: A counterfactual multi-country analysis. http://www.publicdebtnet.org/pdm/.content/Report/Report-00728.html. (2020).

[53] Matteo Cinelli, Walter Quattrociocchi, Alessandro Galeazzi, Carlo Michele Valensise, Emanuele Brugnoli, Ana Lucía Schmidt, Paola Zola, Fabiana Zollo, and Antonio Scala. 2020. The COVID-19 Social Media Infodemic. *CoRR* abs/2003.05004 (2020).

[54] 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, e9448 (2020). Issue 7.

[55] 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:2006.11988* (2020).

[56] Pietro Coletti, Chiara Poletto, Clément Turbelin, Thierry Blanchon, and Vittoria Colizza. 2018. Shifting patterns of seasonal influenza epidemics. *Scientific reports* 8, 1 (2018), 1–12.

[57] Nuno Crokidakis. 2020. Modeling the early evolution of the COVID-19 in Brazil: Results from a Susceptible-Infectious-Quarantined-Recovered (SIQR) model. *International Journal of Modern Physics C (IJMPC)* 31, 10 (2020), 1–7.

[58] Christine SM Currie, John W Fowler, Kathy Kotiadis, Thomas Monks, Bhakti Stephan Onggo, Duncan A Robertson, and Antuela A Tako. 2020. How simulation modelling can help reduce the impact of COVID-19. *Journal of Simulation* 14, 2 (2020), 83–97.

[59] Ramon Gomes da Silva, Matheus Henrique Dal Molin Ribeiro, Viviana Cocco Mariani, and Leandro dos Santos Coelho. 2020. Forecasting Brazilian and American COVID-19 cases based on artificial intelligence coupled with climatic exogenous variables. *Chaos Solitons Fractals* 139, 110027 (2020).

[60] Raj Dandekar and George Barbastathis. 2020. Neural Network aided quarantine control model estimation of COVID spread in Wuhan, China. *arXiv:2003.09403* (2020).

[61] Dipayan Das, KC Santosh, and Umapada Pal. 2020. Truncated inception net: COVID-19 outbreak screening using chest X-rays. *Physical and engineering sciences in medicine* 43, 3 (2020), 915–925.

[62] William Marciel de Souza, Lewis Fletcher Buss, Darlan da Silva Candido, Jean-Paul Carrera, Sabrina Li, Alexander E. Zarebski, Rafael Henrique Moraes Pereira, Carlos A. Prete Jr, Andreza Aruska de Souza-Santos, Kris V. Parag, Maria Carolina T. D. Belotti, Maria F. Vincenti-Gonzalez, Janey Messina, Flavia Cristina da Silva Sales, Pamela dos Santos Andrade, Vítor Heloiz Nascimento, Fabio Ghilardi, Leandro Abade, Bernardo Gutierrez, Moritz U. G. Kraemer, Carlos K. V. Braga, Renato Santana Aguiar, Neal Alexander, Philippe Mayaud, Oliver J. Brady, Izabel Marcilio, Nelson Gouveia, Guangdi Li, Adriana Tami, Silvano Barbosa de Oliveira, Victor Bertollo Gomes Porto, Fabiana Ganem, Walquiria Aparecida Ferreira de Almeida, Francieli Fontana Sutile Tardetti Fantinato, Eduardo Marques Macário, Wanderson Kleber de Oliveira, Mauricio L. Nogueira, Oliver G. Pybus, Chieh-Hsi Wu, Julio Croda, Ester C. Sabino, and Nuno Rodrigues Faria. 2020. Epidemiological and clinical characteristics of the early phase of the COVID-19 epidemic in Brazil. *Nature Human Behavior* 4 (2020), 856–865.

[63] 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).

[64] R Maria del Rio-Chanona, Penny Mealy, Anton Pichler, Francois Lafond, and J Doyne Farmer. 2020. Supply and demand shocks in the COVID-19 pandemic: An industry and occupation perspective. *Oxford Review of Economic Policy* 36, Supplement_1 (2020), S94–S137.

[65] Jayanthi Devaraj, Rajvikram Madurai Elavarasan, Rishi Pugazhendhi, G.M. Shafiullah, Sumathi Ganesan, Ajay Kaarthic Jeysree, Irfan Ahmad Khan, and Eklas Hossaine. 2021. Forecasting of COVID-19 cases using deep learning models: Is it reliable and practically significant? *Results in Physics* 21, 103817 (2021), 1–25.

[66] Sven Drefahl, Matthew Wallace, Eleonora Mussino, Siddartha Aradhya, Martin Kolk, Maria Brandén, Bo Malmberg, and Gunnar Andersson. 2020. A population-based cohort study of socio-demographic risk factors for COVID-19 deaths in Sweden. *Nature communications* 11, 1 (2020), 1–7.

[67] Viet Duong, Phu Pham, Tongyu Yang, Yu Wang, and Jiebo Luo. 2020. The ivory tower lost: How college students respond differently than the general public to the COVID-19 pandemic. *arXiv:2004.09968* (2020).

[68] Michael Eisenstein. 2018. Infection forecasts powered by big data. *Nature* 555, 7695 (2018).

[69] Ezekiel J. Emanuel, Govind Persad, Ross Upshur, Beatriz Thome, Michael Parker, Aaron Glickman, Cathy Zhang, Connor Boyle, Maxwell Smith, and James P. Phillips. 2020. Fair Allocation of Scarce Medical Resources in the Time of Covid-19. *The New England Journal of Medicine* 382 (2020),

2049–2055.

[70] Hanie Esakandari, Mohsen Nabi-Afjadi, Javad Fakkari-Afjadi, Navid Farahmandian, Seyed-Mohsen Miresmaeili, and Elham Bahreini. 2020. A comprehensive review of COVID-19 characteristics. *Biological Procedures Online* 22 (2020), 1–10.

[71] Ali Eshragh, Saed Alizamir, Peter Howley, and Elizabeth Stojanovski. 2020. Modeling the dynamics of the COVID-19 population in Australia: A probabilistic analysis. *PLoS ONE* 15, e0240153 (2020). Issue 10.

[72] Guihong Fan, Zhichun Yang, Qianying Lin, Shi Zhao, Lin Yang, and Daihai He. 2020. Decreased case fatality rate of COVID-19 in the second wave: a study in 53 countries or regions. *Transboundary and emerging diseases* (2020), 1–7.

[73] Yaqing Fang, Yiting Nie, and Marshare Penny. 2020. Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: A data-driven analysis. *Journal of medical virology* 92 (2020), 645–659. Issue 6.

[74] Ahmed Abdullah Farid, Gamal Ibrahim Selim, and Hatem Awad A. Khater. 2020. A Novel Approach of CT Images Feature Analysis and Prediction to Screen for Corona Virus Disease (COVID-19). *International Journal of Scientific & Engineering Research* 11 (2020), 1–9. Issue 3.

[75] Dennis M Feehan and Ayesha S Mahmud. 2021. Quantifying population contact patterns in the United States during the COVID-19 pandemic. *Nature Communications* 12, 1 (2021), 1–9.

[76] Wei Feng, Ashley M. Newbigging, Connie Le, Bo Pang, Hanyong Peng, Yiren Cao, Jinjun Wu, Ghulam Abbas, Jin Song, Dian-Bing Wang, Mengmeng Cui, Jeffrey Tao, Lorne D. Tyrrell, Xian-En Zhang, Hongquan Zhang, and Chris X. Le. 2020. Molecular Diagnosis of COVID-19: Challenges and Research Needs. *Analytical Chemistry* 92 (2020), 10196–10209.

[77] Seth Flaxman, Swapnil Mishra, Axel Gandy, H. Juliette T. Unwin, Thomas A. Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W. Eaton, Mélodie Monod, Azra C. Ghani, Christl A. Donnelly, Steven Riley, Michaela A. C. Vollmer, Neil M. Ferguson, Lucy C. Okell, and Samir Bhatt. 2020. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* 584 (2020), 257–261. Imperial College COVID-19 Response Team.

[78] Simon James Fong, Gloria Li, Nilanjan Dey, Rubén González Crespo, and Enrique Herrera-Viedma. 2020. Composite Monte Carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction. *Applied Soft Computing* 93 (2020), 106282.

[79] Christophe Fraser, Christl A Donnelly, Simon Cauchemez, William P Hanage, Maria D Van Kerkhove, T Déirdre Hollingsworth, Jamie Griffin, Rebecca F Baggaley, Helen E Jenkins, Emily J Lyons, Thibaut Jombart, Wes R Hinsley, Nicholas C Grassly, Francois Balloux, Azra C Ghani, Neil M Ferguson, Andrew Rambaut, Oliver G Pybus, Hugo Lopez-Gatell, Celia M Alpuche-Aranda, Ietza Bojorquez Chapela, Ethel Palacios Zavala, Dulce Ma Espejo Guevara, Francesco Checchi, Erika Garcia, Stephane Hugonnet, and Cathy Roth. 2009. Pandemic potential of a strain of influenza A (H1N1): early findings. *science* 324, 5934 (2009), 1557–1561. WHO Rapid Pandemic Assessment Collaboration.

[80] 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. *Proceedings of the National Academy of Sciences* 117, 19 (2020), 10484–10491.

[81] Giulia Giordano, Franco Blanchini, Raffaele Bruno, Patrizio Colaneri, Alessandro Di Filippo, Angela Di Matteo, and Marta Colaneri. 2020. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature medicine* 26, 6 (2020), 855–860.

[82] Jonatan Gomez, Jeisson Prieto, Elizabeth Leon, and Arles Rodríguez. 2021. INFEKTA: An agent-based model for transmission of infectious diseases: The COVID-19 case in Bogota, Colombia. *PLoS ONE* 16, e0245787 (2021), 1–15. Issue 2.

[83] Kyra H. Grantz, Hannah R. Meredith, Derek A. T. Cummings, C. Jessica E. Metcalf, Bryan T. Grenfell, John R. Giles, Shruti Mehta, Sunil Solomon, Alain Labrique, Nishant Kishore, Caroline O. Buckee, and Amy Wesolowski. 2020. The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nat Commun.* 11, 4961 (2020), 1–8.

[84] Victor Grech and Sarah Cuschieri. 2020. COVID-19: A global and continental overview of the second wave and its (relatively) attenuated case fatality ratio. *Early Human Development* (2020), 1–4.

[85] Nathan D Grubaugh, William P Hanage, and Angela L Rasmussen. 2020. Making sense of mutation: what D614G means for the COVID-19 pandemic remains unclear. *Cell* 182, 4 (2020), 794–795.

[86] Nathan D Grubaugh, Emma B Hodcroft, Joseph R Fauver, Alexandra L Phelan, and Muge Cevik. 2021. Public health actions to control new SARS-CoV-2 variants. *Cell* 184 (2021), 1127–1132. Issue 5.

[87] Amit Kumar Gupta, Vijander Singh, Priya Mathur, and Carlos M Travieso-Gonzalez. 2020. Prediction of COVID-19 pandemic measuring criteria using support vector machine, prophet and linear regression models in Indian scenario. *Journal of Interdisciplinary Mathematics* (2020), 1–20.

[88] Deepti Gupta, Smriti Bhatt, Maanak Gupta, and Ali Saman Tosun. 2021. Future smart connected communities to fight covid-19 outbreak. *Internet of Things* 13 (2021), 100342.

[89] Rajan Guptaa, Gaurav Pandeyb, Poonam Chaudharyb, and Saibal Palc. 2020. SEIR and Regression Model based COVID-19 outbreak predictions in India. *MedRxiv* (2020).

[90] Ali M. Hasan, Mohammed M. AL-Jawad, Hamid A. Jalab, Hadil Shaiba, Rabha W. Ibrahim, and Alaa R. AL-Shamasneh. 2020. Classification of Covid-19 Coronavirus, Pneumonia and Healthy Lungs in CT Scans Using Q-Deformed Entropy and Deep Learning Features. *Entropy* 22, 5 (2020). 517.

[91] Laurent Hébert-Dufresne, Samuel V Scarpino, and Jean-Gabriel Young. 2020. Macroscopic patterns of interacting contagions are indistinguishable from social reinforcement. *Nature Physics* 16, 4 (2020), 426–431.

[92] Ezz El-Din Hemdan, Marwa A Shouman, and Mohamed Esmail Karar. 2020. Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. *arXiv:2003.11055* (2020).

[93] Inga Holmdahl and Caroline Buckee. 2020. Wrong but useful—what covid-19 epidemiologic models can and cannot tell us. *New England Journal of Medicine* 383, 4 (2020), 303–305.

[94] Hyokyoung G Hong and Yi Li. 2020. Estimation of time-varying reproduction numbers underlying epidemiological processes: A new statistical tool for the COVID-19 pandemic. *PloS ONE* 15, 7 (2020), e0236464.

[95] Zhiyuan Hou, Fanxing Du, Hao Jiang, Xinyu Zhou, and Leesa Lin. 2020. Assessment of public attention, risk perception, emotional and behavioural responses to the COVID-19 outbreak: social media surveillance in China. *medRxiv* (2020). https://doi.org/10.1101/2020.03.14.20035956 2020.03.14.20035956.

[96] Ben Hu, Hua Guo, Peng Zhou, and Zheng-Li Shi. 2021. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol* 19 (2021), 141–154. Issue 3.

[97] Fan Hu, Jiaxin Jiang, and Peng Yin. 2020. Prediction of potential commercially inhibitors against SARS-CoV-2 by multi-task deep model. *arXiv:2003.00728* (2020).

[98] Zixin Hu, Qiyang Ge, Shudi Li, Eric Boerwinkle, Li Jin, and Momiao Xiong. 2020. Forecasting and evaluating multiple interventions for COVID-19 worldwide. *Frontiers in Artificial Intelligence* 3 (2020), 41.

[99] Zixin Hu, Qiyang Ge, Shudi Li, Li Jin, and Momiao Xiong. 2020. Artificial intelligence forecasting of covid-19 in china. *arXiv:2002.07112* (2020).

[100] Jizhou Huang, Haifeng Wang, Miao Fan, An Zhuo, Yibo Sun, and Ying Li. 2020. Understanding the Impact of the COVID-19 Pandemic on Transportation-related Behaviors with Human Mobility Data. In *KDD'20*. 3443–3450.

[101] Mohamed R. Ibrahim, James Haworth, Aldo Lipani, Nilufer Aslam, Tao Cheng, and Nicola Christie. 2021. Variational-LSTM autoencoder to forecast the spread of coronavirus across the globe. *PLoS ONE* 16, e0246120 (2021), 1–23. Issue 1.

[102] Ali Imran, Iryna Posokhova, Haneya N Qureshi, Usama Masood, Muhammad Sajid Riaz, Kamran Ali, Charles N John, MD Iftikhar Hussain, and Muhammad Nabeel. 2020. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Informatics in Medicine Unlocked* 20 (2020), 100378.

[103] Md. Milon Islam, Fakhri Karray, Reda Alhajj, and Jia Zeng. 2021. A Review on Deep Learning Techniques for the Diagnosis of Novel Coronavirus (COVID-19). *IEEE Access* 9 (2021), 30551–30572.

[104] Zheng Jiang, Menghan Hu, Zhongpai Gao, Lei Fan, Ranran Dai, Yaling Pan, Wei Tang, Guangtao Zhai, and Yong Lu. 2020. Detection of Respiratory Infections Using RGB-Infrared Sensors on Portable Device. *IEEE Sensors Journal* 20, 22 (2020), 13674–13681.

[105] Wei jie Guan, Zheng yi Ni, Yu Hu, Wen hua Liang, Chun quan Ou, Jian xing He, Lei Liu, Hong Shan, Chun liang Lei, David S.C. Hui, Bin Du, Lan juan Li, Guang Zeng, Kwok-Yung Yuen, Ru chong Chen, Chun li Tang, Tao Wang, Ping yan Chen, Jie Xiang, Shi yue Li, Jin lin Wang, Zi jing Liang, Yi xiang Peng, Li Wei, Yong Liu, Ya hua Hu, Peng Peng, Jian ming Wang, Ji yang Liu, Zhong Chen, Gang Li, Zhi jian Zheng, Shao qin Qiu, Jie Luo, Chang jiang Ye, Shao yong Zhu, and Nan shan Zhong. 2020. Clinical characteristics of coronavirus disease 2019 in China. *New England journal of medicine* 382, 18 (2020), 1708–1720.

[106] Firuz Kamalov, Aswani Cherukuri, Hana Sulieman, Fadi A. Thabtah, and Akbar Hossain. 2021. Machine learning applications for COVID-19: A state-of-the-art review. *CoRR* abs/2101.07824 (2021).

[107] Hengyuan Kang, Liming Xia, Fuhua Yan, Zhibin Wan, Feng Shi, Huan Yuan, Huiting Jiang, Dijia Wu, He Sui, Changqing Zhang, and Dinggang Shen. 2020. Diagnosis of Coronavirus Disease 2019 (COVID-19) With Structured Latent Multi-View Representation Learning. *IEEE Trans. Medical Imaging* 39, 8 (2020), 2606–2614.

[108] Dharun Kasilingam, Sakthivel Puvaneswaran Sathiya Prabhakaran, Dinesh Kumar Rajendran, Varthini Rajagopal, Thangaraj Santhosh Kumar, and Ajitha Soundararaj. 2020. Exploring the growth of COVID-19 cases using exponential modelling across 42 countries and predicting signs of early containment using machine learning. *Transboundary and Emerging Diseases* (2020), 1–45.

[109] Matt J Keeling, T Deirdre Hollingsworth, and Jonathan M Read. 2020. Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19). *J Epidemiol Community Health* 74 (2020), 861–866. Issue 10.

[110] William Ogilvy Kermack and Anderson G McKendrick. 1927. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of London. Series A, Containing papers of a mathematical and physical character* 115, 772 (1927), 700–721.

[111] Arash Keshavarzi Arshadi, Julia Webb, Milad Salem, Emmanuel Cruz, Stacie Calad-Thomson, Niloofar Ghadirian, Jennifer Collins, Elena Diez-Cecilia, Brendan Kelly, Hani Goodarzi, and Jiann Shiun Yuan. 2020. Artificial Intelligence for COVID-19 Drug Discovery and Vaccine Development. *Frontiers in Artificial Intelligence* 3 (2020), 65.

[112] Nour Eldeen M Khalifa, Mohamed Hamed N Taha, Aboul Ella Hassanien, and Sally Elghamrawy. 2020. Detection of coronavirus (covid-19) associated pneumonia based on generative adversarial networks and a fine-tuned deep transfer learning model using chest x-ray dataset. *arXiv:2004.01184* (2020).

[113] Asif Iqbal Khan, Junaid Latief Shah, and Mohammad Mudasir Bhat. 2020. CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. *Computer Methods and Programs in Biomedicine* 196 (2020), 105581.

[114] Akib Mohi Ud Din Khanday, Syed Tanzeel Rabani, Qamar Rayees Khan, Nusrat Rouf, and Masarat Mohi Ud Din. 2020. Machine learning based approaches for detecting COVID-19 using clinical text data. *International Journal of Information Technology* 12 (2020), 731–739.

[115] Louis Kim, Shannon M Fast, and Natasha Markuzon. 2019. Incorporating media data into a model of infectious disease transmission. *PloS one* 14, 2 (2019), e0197646.

[116] Minseok Kim, Junhyeok Kang, Doyoung Kim, Hwanjun Song, Hyangsuk Min, Youngeun Nam, Dongmin Park, and Jae-Gil Lee. 2020. Hi-COVIDNet: Deep Learning Approach to Predict Inbound COVID-19 Patients and Case Study in South Korea. In *KDD'20*. 3466–3473.

[117] Vasilis Kontis, James E Bennett, Theo Rashid, Robbie M Parks, Jonathan Pearson-Stuttard, Michel Guillot, Perviz Asaria, Bin Zhou, Marco Battaglini, Gianni Corsetti, Martin McKee, Mariachiara Di Cesare, Colin D. Mathers, and Majid Ezzati. 2020. Magnitude, demographics and dynamics of the effect of the first wave of the COVID-19 pandemic on all-cause mortality in 21 industrialized countries. *Nature medicine* (2020), 1–10.

[118] Moritz U. G. Kraemer, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M. Pigott, Open COVID-19 Data Working Group, Louis du Plessis, Nuno R. Faria, Ruoran Li, William P. Hanage, John S. Brownstein, Maylis Layan, Alessandro Vespignani, Huaiyu Tian, Christopher Dye, Oliver G. Pybus, and Samuel V. Scarpino. 2020. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 368 (2020), 493–497. Issue 6490.

[119] SE Kreps and DL Kriner. 2020. Model uncertainty, political contestation, and public trust in science: Evidence from the COVID-19 pandemic. *Science advances* 6, 43 (2020), eabd4563.

[120] Andreas Kronbichler, Daniela Kresse, Sojung Yoon, Keum Hwa Lee, Maria Effenberger, and Jae Il Shin. 2020. Asymptomatic patients as a source of COVID-19 infections: A systematic review and meta-analysis. *International journal of infectious diseases* 98 (2020), 180–186.

[121] Shengjie Lai, Nick W Ruktanonchai, Liangcai Zhou, Olivia Prosper, Wei Luo, Jessica R Floyd, Amy Wesolowski, Mauricio Santillana, Chi Zhang, Xiangjun Du, Hongjie Yu, and Andrew J. Tatem. 2020. Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* 585, 7825 (2020), 410–413.

[122] Siddique Latif, Muhammad Usman, Sanaullah Manzoor, Waleed Iqbal, Junaid Qadir, Gareth Tyson, Ignacio Castro, Adeel Razi, Maged N. Kamel Boulos, Adrian Weller, and Jon Crowcroft. 2020. Leveraging Data Science to Combat COVID-19: A Comprehensive Review. *IEEE Transactions on Artificial Intelligence* 1, 1 (2020), 85–103.

[123] Ardashel Latsuzbaia, Malte Herold, Jean-Paul Bertemes, and Joël Mossong. 2020. Evolving social contact patterns during the COVID-19 crisis in Luxembourg. *PloS one* 15, 8 (2020), e0237128.

[124] Stephen A Lauer, Kyra H Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin Lessler. 2020. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine* 172, 9 (2020), 577–582.

[125] Cheryl Yi-Pin Lee, Raymond T. P. Lin, Laurent Renia, and Lisa F. P. Ng. 2020. Serological Approaches for COVID-19: Epidemiologic Perspective on Surveillance and Control. *Frontiers in Immunology* 11 (2020), 879.

[126] Yan Leng, Yujia Zhai, Shaojing Sun, Yifei Wu, Jordan Selzer, Sharon Strover, Hezhao Zhang, Anfan Chen, and Ying Ding. 2021. Misinformation During the COVID-19 Outbreak in China: Cultural, Social and Political Entanglements. *IEEE Trans. Big Data* 7, 1 (2021), 69–80.

[127] Kathy Leung, Joseph T Wu, Di Liu, and Gabriel M Leung. 2020. First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. *The Lancet* 395, 10233 (2020), 1382–1393.

[128] Guanjian Li, Weiran Li, Xiaojin He, and Yunxia Cao. 2020. Asymptomatic and Presymptomatic Infectors: Hidden Sources of Coronavirus Disease 2019 (COVID-19). *Clinical Infectious Diseases* 71, 8 (2020), 2018–2018.

[129] Jinjing Li, Yogi Vidyattama, Hai Anh La, Riyana Miranti, and Denisa M. Sologon. 2020. The Impact of COVID-19 and Policy Responses on Australian Income Distribution and Poverty. *CoRR* (2020), 1–21. arXiv:2009.04037

[130] Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman2. 2020. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* 368, 6490 (2020), 489–493.

[131] Xiaoya Li, Mingxin Zhou, Jiawei Wu, Arianna Yuan, Fei Wu, and Jiwei Li. 2020. Analyzing Covid-19 on online social media: Trends, sentiments and emotions. *arXiv:2005.14464* (2020).

[132] Mayara Lisboa Bastos, Gamuchirai Tavaziva, Syed Kunal Abidi, Jonathon R Campbell, Louis-Patrick Haraoui, James C Johnston, Zhiyi Lan, Stephanie Law, Emily MacLean, Anete Trajman, Dick Menzies, Andrea Benedetti, and Faiz Ahmad Khan. 2020. Diagnostic accuracy of serological tests for covid-19: systematic review and meta-analysis. *BMJ* 370 (2020), 1–13.

[133] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. 2017. A survey on deep learning in medical image analysis. *Medical image analysis* 42 (2017), 60–88.

[134] Jiangtao Liu, Ji Zhou, Jinxi Yao, Xiuxia Zhang, Lanyu Li, Xiaocheng Xu, Xiaotao He, Bo Wang, Shihua Fu, Tingting Niu, Jun Yan, Yanjun Shi, Xiaowei Ren, Jingping Niu, Weihao Zhu, Sheng Li, Bin Luo, and Kai Zhang. 2020. Impact of meteorological factors on the COVID-19 transmission: A multi-city study in China. *Science of the total environment* 726 (2020), 138513.

[135] Quan-Hui Liu, Marco Ajelli, Alberto Aleta, Stefano Merler, Yamir Moreno, and Alessandro Vespignani. 2018. Measurability of the epidemic reproduction number in data-driven contact networks. *PNAS* 115, 50 (2018), 12680–12685.

[136] Mohamed Loey, Florentin Smarandache, and Nour Eldeen M. Khalifa. 2020. Within the Lack of Chest COVID-19 X-ray Dataset: A Novel Detection Model Based on GAN and Deep Transfer Learning. *Symmetry* 12, 4 (2020). 651.

[137] Leonardo López and Xavier Rodó. 2020. The end of social confinement and COVID-19 re-emergence risk. *Nature Human Behaviour* 4, 7 (2020), 746–755.

[138] Roujian Lu, Xiang Zhao, Juan Li, Peihua Niu, Bo Yang, Honglong Wu, Wenling Wang, Hao Song, Baoying Huang, Na Zhu, Yuhai Bi, Xuejun Ma, Faxian Zhan, Liang Wang, Tao Hu, Hong Zhou, Zhenhong Hu, Weimin Zhou, Li Zhao, Jing Chen, Yao Meng, Ji Wang, Yang Lin, Jianying Yuan, Zhihao Xie, Jinmin Ma, William J Liu, Dayan Wang, Wenbo Xu, Edward C Holmes, George F Gao, Guizhen Wu, Weijun Chen, Weifeng Shi, and Wenjie Tan. 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* 395, 10224 (2020), 565–574.

[139] Rishikesh Magar, Prakarsh Yadav, and Amir Barati Farimani. 2021. Potential neutralizing antibodies discovered for novel corona virus using machine learning. *Scientific Reports* 11, 5261 (2021), 1–11.

[140] Gianluca Maguolo and Loris Nanni. 2020. A Critic Evaluation of Methods for COVID-19 Automatic Detection from X-Ray Images. (2020). arXiv:2004.12823

[141] Benjamin F Maier and Dirk Brockmann. 2020. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science* 368, 6492 (2020), 742–746.

[142] Zohair Malki, El-Sayed Atlam, Aboul Ella Hassanien, Guesh Dagnew, Mostafa A. Elhosseini, and Ibrahim Gadb. 2020. Association between weather data and COVID-19 pandemic predicting mortality rate: Machine learning approaches. *Chaos Solitons Fractals* 138, 110137 (2020), 1–11.

[143] Ying Mao, Susiyan Jiang, and Daniel Nametz. 2020. Data-driven Analytical Models of COVID-2019 for Epidemic Prediction, Clinical Diagnosis, Policy Effectiveness and Contact Tracing: A Survey. (2020).

[144] David-A. Mendels, Laurent Dortet, Cécile Emeraud, Saoussen Oueslati, Delphine Girlich, Jean-Baptiste Ronat, Sandrine Bernabeu, Silvestre Bahi, Gary J. H. Atkinson, and Thierry Naas. 2021. Using artificial intelligence to improve COVID-19 rapid diagnostic test result interpretation. *Proceedings of the National Academy of Sciences* 118, 12 (2021). e2019893118.

[145] Hayden C Metsky, Catherine A Freije, Tinna-Solveig F Kosoko-Thoroddsen, Pardis C Sabeti, and Cameron Myhrvold. 2020. CRISPR-based surveillance for COVID-19 using genomically-comprehensive machine learning design. *BioRxiv* (2020).

[146] Hayden C Metsky, Nicole L Welch, Nicholas J Haradhvala, Laurie Rumker, Yibin B Zhang, Priya P Pillai, David K Yang, Cheri M Ackerman, Juliane Weller, Paul C Blainey, Cameron Myhrvold, Michael Mitzenmacher, and Pardis C Sabeti. 2020. Designing viral diagnostics with model-based optimization. *BioRxiv* (2020), 1–83.

[147] Nicholas Micallef, Bing He, Srijan Kumar, Mustaque Ahamad, and Nasir D. Memon. 2020. The Role of the Crowd in Countering Misinformation: A Case Study of the COVID-19 Infodemic. In *IEEE Big Data'2020*. 748–757.

[148] Michael J. Mina, Roy Parker, and Daniel B. Larremore. 2020. Rethinking Covid-19 Test Sensitivity — A Strategy for Containment. *New England Journal of Medicine* 383, 22 (2020), e120.

[149] Shervin Minaee, Rahele Kafieh, Milan Sonka, Shakib Yazdani, and Ghazaleh Jamalipour Soufi. 2020. Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning. *Medical image analysis* 65 (2020), 101794.

[150] Adam S Miner, Liliana Laranjo, and A Baki Kocaballi. 2020. Chatbots in the fight against the COVID-19 pandemic. *NPJ digital medicine* 3, 1 (2020), 1–4.

[151] Youssoufa Mohamadou, Aminou Halidou, and Pascalin Tiam Kapen. 2020. A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19. *Applied Intelligence* 50, 11 (2020), 3913–3925.

[152] Himadri Mukherjee, Subhankar Ghosh, Ankita Dhar, Sk Obaidullah, KC Santosh, and Kaushik Roy. 2021. Shallow convolutional neural network for COVID-19 outbreak screening using chest X-rays. (2021), 1–14.

[153] R Muthusami and K Saritha. 2020. Statistical analysis and visualization of the potential cases of pandemic coronavirus. *VirusDisease* 31, 2 (2020), 204–208.

[154] Kyle R. Myers, Wei Yang Tham, Yian Yin, Nina Cohodes, Jerry G. Thursby, Marie C. Thursby, Peter E. Schiffer, Joseph T. Walsh, Karim R. Lakhani, and Dashun Wang. 2020. Quantifying the Immediate Effects of the COVID-19 Pandemic on Scientists. *CoRR* (2020), 1–39. arXiv:2005.11358

[155] Ali Narin, Ceren Kaya, and Ziynet Pamuk. 2020. Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks. (2020). arXiv:2003.10849

[156] M. Saqib Nawaz, Philippe Fournier-Viger, Abbas Shojaee, and Hamido Fujita. 2021. Using artificial intelligence techniques for COVID-19 genome analysis. *Applied Intelligence* (2021), 1–18.

[157] Muhammad Nazrul Islam, Toki Tahmid Inan, Suzzana Rafi, Syeda Sabrina Akter, Iqbal H Sarker, and AKM Najmul Islam. 2020. A Survey on the Use of AI and ML for Fighting the COVID-19 Pandemic. *arXiv e-prints* (2020), arXiv–2008.

[158] László Nemes and Attila Kiss. 2021. Social media sentiment analysis based on COVID-19. *Journal of Information and Telecommunication* 5 (2021), 1–15. Issue 1.

[159] Yixiang Ng, Zongbin Li, Yi Xian Chua, Wei Liang Chaw, Zheng Zhao, Benjamin Er, Rachael Pung, Calvin J Chiew, David C Lye, Derrick Heng, and Vernon J. Lee. 2020. Evaluation of the effectiveness of surveillance and containment measures for the first 100 patients with COVID-19 in Singapore—January 2–February 29, 2020. *Morbidity and mortality weekly report* 69, 11 (2020), 307–311.

[160] Thanh Thi Nguyen. 2020. Artificial intelligence in the battle against coronavirus (COVID-19): a survey and future research directions. *arXiv:2008.07343* (2020).

[161] Majid Niazkar and Hamid Reza Niazkar. 2020. Covid-19 outbreak: Application of multi-gene genetic programming to country-based prediction models. *Electronic Journal of General Medicine* 17, 5 (2020).

[162] Rene Niehus, Pablo M De Salazar, Aimee R Taylor, and Marc Lipsitch. 2020. Using observational data to quantify bias of traveller-derived COVID-19 prevalence estimates in Wuhan, China. *The Lancet*

Infectious Diseases (2020).

[163] Mizuho Nishio, Shunjiro Noguchi, Hidetoshi Matsuo, and Takamichi Murakami. 2020. Automatic classification between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy on chest X-ray image: combination of data augmentation methods. *Scientific reports* 10, 1 (2020), 1–6.

[164] Barbara Oliveiros, Liliana Caramelo, Nuno C Ferreira, and Francisco Caramelo. 2020. Role of temperature and humidity in the modulation of the doubling time of COVID-19 cases. *MedRxiv* (2020).

[165] Ryosuke Omori, Ryota Matsuyama, and Yukihiko Nakata. 2020. The age distribution of mortality from novel coronavirus disease (COVID-19) suggests no large difference of susceptibility by age. *Scientific Reports* 10, 16642 (2020), 1–9.

[166] Philip D O'Neill and Gareth O Roberts. 1999. Bayesian inference for partially observed stochastic epidemics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 162, 1 (1999), 121–129.

[167] World Health Organization et al. 2020. *Transmission of SARS-CoV-2: implications for infection prevention precautions: Scientific brief, 09 July 2020*. Technical Report. World Health Organization.

[168] Dave Osthus, James Gattiker, Reid Priedhorsky, and Sara Y Del Valle. 2019. Dynamic Bayesian influenza forecasting in the United States with hierarchical discrepancy (with discussion). *Bayesian Analysis* 14, 1 (2019), 261–312.

[169] Xiuyuan Ou, Yan Liu, Xiaobo Lei, Pei Li, Dan Mi, Lili Ren, Li Guo, Ruixuan Guo, Ting Chen, Jiaxin Hu, Zichun Xiang, Zhixia Mu, Xing Chen, Jieyong Chen, Keping Hu, Qi Jin, Jianwei Wang, and Zhaohui Qian. 2020. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nature communications* 11, 1 (2020), 1–12.

[170] Saban Öztürk, Umut Özkaya, and Mücahid Barstugan. 2021. Classification of Coronavirus (COVID-19) from X-ray and CT images using shrunken features. *Int. J. Imaging Syst. Technol.* 31, 1 (2021), 5–15.

[171] Tulin Ozturk, Muhammed Talo, Eylul Azra Yildirim, Ulas Baran Baloglu, Ozal Yildirim, and U Rajendra Acharya. 2020. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in biology and medicine* 121 (2020), 103792.

[172] Ratnabali Pal, Arif Ahmed Sekh, Samarjit Kar, and Dilip K Prasad. 2020. Neural network based country wise risk prediction of COVID-19. *Applied Sciences* 10, 18 (2020), 6448.

[173] Minah Park, Alex R Cook, Jue Tao Lim, Yinxiaohe Sun, and Borame L Dickens. 2020. A systematic review of COVID-19 epidemiology based on current evidence. *Journal of Clinical Medicine* 9, 4 (2020), 967.

[174] Sansao A Pedro, Frank T Ndjomatchoua, Peter Jentsch, Jean M Tcheunche, Madhur Anand, and Chris T Bauch. 2020. Conditions for a second wave of COVID-19 due to interactions between disease dynamics and social processes. *Frontiers in Physics* (2020), 1–9.

[175] Ana Luisa Pedrosa, Letícia Bitencourt, Ana Cláudia Fontoura Fróes, Maria Luíza Barreto Cazumbá, Ramon Gustavo Bernardino Campos, Stephanie Bruna Camilo Soares de Brito, and Ana Cristina Simões e Silva. 2020. Emotional, behavioral, and psychological Impact of the COVID-19 Pandemic. *Frontiers in psychology* 11 (2020).

[176] Rosanna W Peeling, Catherine J Wedderburn, Patricia J Garcia, Debrah Boeras, Noah Fongwen, John Nkengasong, Amadou Sall, Amilcar Tanuri, and David L Heymann. 2020. Serology testing in the COVID-19 pandemic response. *The Lancet Infectious Diseases* 20, 9 (2020), e245–e249.

[177] Liangrong Peng, Wuyue Yang, Dongyan Zhang, Changjing Zhuge, and Liu Hong. 2020. Epidemic analysis of COVID-19 in China by dynamical modeling. *arXiv:2002.06563* (2020).

[178] Igor Gadelha Pereira, Joris Michel Guerin, Andouglas Gonçalves Silva Júnior, Gabriel Santos Garcia, Prisco Piscitelli, Alessandro Miani, Cosimo Distante, and Luiz Marcos Garcia Goncalves. 2020. Forecasting Covid-19 Dynamics in Brazil: A Data Driven Approach. *International Journal of Environmental Research and Public Health* 17, 14 (2020). 5115.

[179] Eskild Petersen, Marion Koopmans, Unyeong Go, Davidson H Hamer, Nicola Petrosillo, Francesco Castelli, Merete Storgaard, Sulien Al Khalili, and Lone Simonsen. 2020. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *The Lancet infectious diseases* (2020).

[180] Nicola Petrosillo, Giulio Viceconte, Onder Ergonul, Giuseppe Ippolito, and Eskild Petersen. 2020. COVID-19, SARS and MERS: are they closely related? *Clinical Microbiology and Infection* 26, 6 (2020), 729–734.

[181] Anton Pichler, Marco Pangallo, R. Maria del Rio-Chanona, François Lafond, and J. Doyne Farmer. 2020. Production networks and epidemic spreading: How to restart the UK economy? *CoRR* arXiv:2005.10585 (2020).

[182] Nejc Plohl and Bojan Musil. 2021. Modeling compliance with COVID-19 prevention guidelines: The critical role of trust in science. *Psychology, Health & Medicine* 26, 1 (2021), 1–12.

[183] Canelle Poirier, Wei Luo, Maimuna S. Majumder, Dianbo Liu, Kenneth D. Mandl, Todd A. Mooring, and Mauricio Santillana. 2020. The role of environmental factors on transmission rates of the COVID-19 outbreak: An initial assessment in two spatial scales. *Scientific Reports* 10, 17002 (2020), 1–11. Issue 1.

[184] Hasan Polat, Mehmet Siraç Özerdem, Faysal Ekici, and Veysi Akpolat. 2021. Automatic detection and localization of COVID-19 pneumonia using axial computed tomography images and deep convolutional neural networks. *Int J Imaging Syst Technol* (2021), 1–16.

[185] Kiesha Prem, Yang Liu, Timothy W Russell, Adam J Kucharski, Rosalind M Eggo, Nicholas Davies, Stefan Flasche, Samuel Clifford, Carl AB Pearson, James D Munday, Sam Abbott, Hamish Gibbs, Alicia Rosello, Billy J Quilty, Thibaut Jombart, Fiona Sun, Charlie Diamond, Amy Gimma, Kevin van Zandvoort, Sebastian Funk, Christopher I Jarvis, W John Edmunds, Nikos I Bosse, Joel Hellewell, Mark Jit, and Petra Klepac. 2020. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study. *The Lancet Public Health* 5, 5 (2020), e261–e270.

[186] Viola Priesemann, Rudi Balling, Melanie M Brinkmann, Sandra Ciesek, Thomas Czypionka, Isabella Eckerle, Giulia Giordano, Claudia Hanson, Zdenek Hel, Pirta Hotulainen, Peter Klimek, Armin Nassehi, Andreas Peichl, Matjaz Perc, Elena Petelos, Barbara Prainsack, and Ewa Szczurek. 2021. An action plan for pan-European defence against new SARS-CoV-2 variants. *The Lancet* 397, 10273 (2021), 469–470.

[187] Heather Prime, Mark Wade, and Dillon T Browne. 2020. Risk and resilience in family well-being during the COVID-19 pandemic. *American Psychologist* (2020).

[188] Narinder Singh Punn and Sonali Agarwal. 2020. Automated diagnosis of COVID-19 with limited posteroanterior chest X-ray images using fine-tuned deep neural networks. *Applied Intelligence* (2020).

[189] Xiaolong Qi, Zicheng Jiang, Qian Yu, Chuxiao Shao, Hongguang Zhang, Hongmei Yue, Baoyi Ma, Yuancheng Wang, Chuan Liu, Xiangpan Meng, Shan Huang, Dan Xu, Junqiang Lei, Huihong Huang, Jie Yang, Jiansong Ji, Hongqiu Pan, Shengqiang Zou, Shenghong Ju, and Xiaolong Qi. 2020. Machine learning-based CT radiomics model for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: A multicenter study. *Medrxiv* (2020).

[190] Cox D. R. 1972. Regression Models and Life Tables. *Journal of the Royal Statistic Society* B, 34 (1972), 187–202.

[191] Benjamin Rader, Samuel V Scarpino, Anjalika Nande, Alison L Hill, Ben Adlam, Robert C Reiner, David M Pigott, Bernardo Gutierrez, Alexander E Zarebski, Munik Shrestha, John S Brownstein, Marcia C Castro, Christopher Dye, Huaiyu Tian, Oliver G Pybus, and Moritz U G Kraemer. 2020. Crowding and the shape of COVID-19 epidemics. *Nature medicine* (2020), 1–6.

[192] Iman Rahimi, Fang Chen, and Amir H Gandomi. 2021. A review on COVID-19 forecasting models. *Neural Computing and Applications* (2021), 1–11.

[193] Gurjit S Randhawa, Maximillian PM Soltysiak, Hadi El Roz, Camila PE de Souza, Kathleen A Hill, and Lila Kari. 2020. Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. *PloS ONE* 15, 4 (2020), e0232391.

[194] Jawad Rasheed, Akhtar Jamil, Alaa Ali Hameed, Fadi Al-Turjman, and Ahmad Rasheed. 2021. COVID-19 in the Age of Artificial Intelligence: A Comprehensive Review. *Interdisciplinary Sciences: Computational Life Sciences* (2021), 1–23.

[195] Khaiwal Ravindra, Vivek Singh Malik, Bijaya K Padhi, Sonu Goel, and Madhu Gupta. 2020. Consideration for the asymptomatic transmission of COVID-19: Systematic Review and Meta-Analysis. *medRxiv* (2020).

[196] Matheus Henrique Dal Molin Ribeiro, Ramon Gomes da Silva, Viviana Cocco Mariani, and Leandro dos Santos Coelho. 2020. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. *Chaos, Solitons & Fractals* 135 (2020), 109853.

[197] 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, AIX-COVNET, James H. F. Rudd, Evis Sala, and Carola-Bibiane Schönlieb. 2021. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nature Machine Intelligence* 3 (2021), 199–217.

[198] Weston C Roda, Marie B Varughese, Donglin Han, and Michael Y Li. 2020. Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling* (2020).

[199] Jon Roozenbeek, Claudia R Schneider, Sarah Dryhurst, John Kerr, Alexandra LJ Freeman, Gabriel Recchia, Anne Marthe Van Der Bles, and Sander Van Der Linden. 2020. Susceptibility to misinformation about COVID-19 around the world. *Royal Society open science* 7, 10 (2020), 201199.

[200] Rohit Salgotra, Mostafa Gandomi, and Amir H Gandomi. 2020. Time series analysis and forecast of the COVID-19 pandemic in India using genetic programming. *Chaos, Solitons & Fractals* 138 (2020), 109945.

[201] Steven Sanche, Yen Ting Lin, Chonggang Xu, Ethan Romero-Severson, Nick Hengartner, and Ruian Ke. 2020. High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerging Infectious Disease journal* 26, 7 (2020), 1470.

[202] Björn W Schuller, Dagmar M Schuller, Kun Qian, Juan Liu, Huaiyuan Zheng, and Xiao Li. 2020. Covid-19 and computer audition: An overview on what speech & sound analysis could contribute in the Sars-cov-2 corona crisis. *arXiv:2003.11117* (2020).

[203] Patrick Schwab, Arash Mehrjou, Sonali Parbhoo, Leo Anthony Celi, Jürgen Hetzel, Markus Hofer, Bernhard Schölkopf, and Stefan Bauer. 2021. Real-time prediction of COVID-19 related mortality using electronic health records. *Nature communications* 12, 1 (2021), 1–16.

[204] Priya Seetharaman. 2020. Business models shifts: Impact of Covid-19. *International Journal of Information Management* 54 (2020), 102173.

[205] Andrew W. Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Zídek, Alexander W. R. Nelson, Alex Bridgland, Hugo Penedones, Stig Petersen, Karen Simonyan, Steve Crossan, Pushmeet Kohli, David T. Jones, David Silver, Koray Kavukcuoglu, and Demis Hassabis. 2020. Improved protein structure prediction using potentials from deep learning. *Nature* 577, 7792 (2020), 706–710.

[206] Farah Shahid, Aneela Zameer, and Muhammad Muneeb. 2020. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos, Solitons & Fractals* 140 (2020), 110212.

[207] Osama Shahid, Mohammad Nasajpour, Seyedamin Pouriyeh, Reza M Parizi, Meng Han, Maria Valero, Fangyu Li, Mohammed Aledhari, and Quan Z Sheng. 2020. Machine Learning Research Towards Combating COVID-19: Virus Detection, Spread Prevention, and Medical Assistance. *arXiv:2010.07036* (2020).

[208] Karishma Sharma, Sungyong Seo, Chuizheng Meng, Sirisha Rambhatla, and Yan Liu. 2020. Covid-19 on social media: Analyzing misinformation in twitter conversations. *arXiv e-prints* (2020), arXiv–2003.

[209] Feng Shi, Jun Wang, Jun Shi, Ziyan Wu, Qian Wang, Zhenyu Tang, Kelei He, Yinghuan Shi, and Dinggang Shen. 2021. Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19. *IEEE Reviews in Biomedical Engineering* 14 (2021), 4–15.

[210] Nistha Shrestha, Muhammad Yousaf Shad, Osman Ulvi, Modasser Hossain Khan, Ajlina Karamehic-Muratovic, Uyen-Sa D.T. Nguyen, Mahdi Baghbanzadeh, Robert Wardrup, Nasrin Aghamohammadi, Diana Cervantes, Kh. Md Nahiduzzaman, Rafdzah Ahmad Zaki, and Ubydul Haque. 2020. The impact of COVID-19 on globalization. *One Health* 11, 100180 (2020), 1–9.

[211] Harkaran Singh and Joydip Dhar. 2018. *Mathematical population dynamics and epidemiology in temporal and spatio-temporal domains.* CRC Press.

[212] Rajesh Singh and Ronojoy Adhikari. 2020. Age-structured impact of social distancing on the COVID-19 epidemic in India. *arXiv:2003.12055* (2020).

[213] Ram Kumar Singh, Meenu Rani, Akshaya Srikanth Bhagavathula, Ranjit Sah, Alfonso J Rodriguez-Morales, Himangshu Kalita, Chintan Nanda, Shashi Sharma, Yagya Datt Sharma, Ali A Rabaan, Jamal Rahmani, and Pavan Kumar. 2020. Prediction of the COVID-19 pandemic for the top 15 affected countries: Advanced autoregressive integrated moving average (ARIMA) model. *JMIR public health and surveillance* 6, 2 (2020), e19115.

[214] Ying Song, Shuangjia Zheng, Liang Li, Xiang Zhang, Xiaodong Zhang, Ziwang Huang, Jianwen Chen, Huiying Zhao, Yusheng Jie, Ruixuan Wang, Yutian Chong, Jun Shen, Yunfei Zha, and Yuedong Yang. 2021. Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (2021), 1–10.

[215] Pedro Soto-Acosta. 2020. COVID-19 pandemic: Shifting digital transformation to a high-speed gear. *Information Systems Management* 37, 4 (2020), 260–266.

[216] Nicholas Soures, David Chambers, Zachariah Carmichael, Anurag Daram, Dimpy P Shah, Kal Clark, Lloyd Potter, and Dhireesha Kudithipudi. 2020. SIRNet: understanding social distancing measures with hybrid neural network model for COVID-19 infectious spread. *arXiv:2004.10376* (2020).

[217] Hiteshi Tandon, Prabhat Ranjan, Tanmoy Chakraborty, and Vandana Suhag. 2020. Coronavirus (COVID-19): ARIMA based time-series analysis to forecast near future. *arXiv:2004.07859* (2020).

[218] Yi-Wei Tang, Jonathan E. Schmitz, David H. Persing, and Charles W. Stratton. 2020. Laboratory Diagnosis of COVID-19: Current Issues and Challenges. *Journal of Clinical Microbiology* 58 (2020), e00512–20. Issue 6.

[219] Mahdi Tavakoli, Jay Carriere, and Ali Torabi. 2020. Robotics, smart wearable technologies, and autonomous intelligent systems for healthcare during the COVID-19 pandemic: An analysis of the state of the art and future vision. *Advanced Intelligent Systems* 2, 7 (2020), 2000071.

[220] Irani Thevarajan, Thi HO Nguyen, Marios Koutsakos, Julian Druce, Leon Caly, Carolien E van de Sandt, Xiaoxiao Jia, Suellen Nicholson, Mike Catton, Benjamin Cowie, Steven Y. C. Tong, Sharon R. Lewin, and Katherine Kedzierska. 2020. Breadth of concomitant immune responses prior to patient recovery: a case report of non-severe COVID-19. *Nature medicine* 26, 4 (2020), 453–455.

[221] Huaiyu Tian, Yonghong Liu, Yidan Li, Chieh-Hsi Wu, Bin Chen, Moritz UG Kraemer, Bingying Li, Jun Cai, Bo Xu, Qiqi Yang, Ben Wang, Peng Yang, Yujun Cui, Yimeng Song, Pai Zheng, Quanyi Wang, Ottar N. Bjornstad, Ruifu Yang, Bryan T. Grenfell, Oliver G. Pybus, and Christopher Dye. 2020. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* 368, 6491 (2020), 638–642.

[222] Mesut Toğaçar, Burhan Ergen, and Zafer Cömert. 2020. COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches. *Computers in Biology and Medicine* 121 (2020), 103805.

[223] Nicole Tovstiga and George Tovstiga. 2020. COVID-19: a knowledge and learning perspective. *Knowledge Management Research & Practice* (2020), 1–6.

[224] Vincent S Tseng, Josh Jia-Ching Ying, Stephen TC Wong, Diane J Cook, and Jiming Liu. 2020. Computational Intelligence Techniques for Combating COVID-19: A Survey. *IEEE Computational Intelligence Magazine* 15, 4 (2020), 10–22.

[225] Turker Tuncer, Sengul Dogan, and Fatih Ozyurtb. 2020. An automated Residual Exemplar Local Binary Pattern and iterative ReliefF based COVID-19 detection method using chest X-ray image. *Chemometrics and Intelligent Laboratory Systems* 203, 104054 (2020), 1–11.

[226] Buddhisha Udugama, Pranav Kadhiresan, Hannah N Kozlowski, Ayden Malekjahani, Matthew Osborne, Vanessa YC Li, Hongmin Chen, Samira Mubareka, Jonathan B Gubbay, and Warren CW Chan. 2020. Diagnosing COVID-19: the disease and tools for detection. *ACS Nano* 14, 4 (2020), 3822–3835.

[227] Anwaar Ulhaq, Jannis Born, Asim Khan, Douglas Pinto Sampaio Gomes, Subrata Chakraborty, and Manoranjan Paul. 2020. Covid-19 control by computer vision approaches: A survey. *IEEE Access* 8 (2020), 179437–179456.

[228] Henning Ulrich, Micheli M Pillat, and Attila Tárnok. 2020. Dengue Fever, COVID-19 (SARS-CoV-2), and Antibody-Dependent Enhancement (ADE): A Perspective. *Cytometry Part A* (2020).

[229] US CDC. 2021. SARS-CoV-2 Variant Classifications and Definitions. https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html. (2021).

[230] Isabelle van der Vegt and Bennett Kleinberg. 2020. Women Worry About Family, Men About the Economy: Gender Differences in Emotional Responses to COVID-19. In *Social Informatics'2020 (Lecture Notes in Computer Science)*, Vol. 12467. 397–409.

[231] Olivier Vandenberg, Delphine Martiny, Olivier Rochas, Alex van Belkum, and Zisis Kozlakidis. 2021. Considerations for diagnostic COVID-19 tests. *Nature Reviews Microbiology* 19 (2021), 171–183.

[232] Alessandro Vespignani, Huaiyu Tian, Christopher Dye, James O Lloyd-Smith, Rosalind M Eggo, Munik Shrestha, Samuel V Scarpino, Bernardo Gutierrez, Moritz UG Kraemer, Joseph Wu, Kathy Leung, and Gabriel M. Leung. 2020. Modelling COVID-19. *Nature Reviews Physics* 2 (2020), 279–281.

[233] Nishant Vishwamitra, Ruijia Roger Hu, Feng Luo, Long Cheng, Matthew Costello, and Yin Yang. 2020. On Analyzing COVID-19-related Hate Speech Using BERT Attention. In *ICMLA'2020*. 669–676.

[234] Thuy Dung Vo and Manh Dung Tran. 2020. The Impact of Covid-19 Pandemic on the Global Trade. *International Journal of Social Science and Economics Invention* 7, 1 (2020), 1–7.

[235] Erik Volz, Swapnil Mishra, Meera Chand, Jeffrey C Barrett, Robert Johnson, Lily Geidelberg, Wes R Hinsley, Daniel J Laydon, Gavin Dabrera, Áine O'Toole, Roberto Amato, Manon Ragonnet-Cronin, Ian Harrison, Ben Jackson, Cristina V. Ariani, Olivia Boyd, Nicholas J Loman, John T McCrone, Sónia Gonçalves, David Jorgensen, Richard Myers, Verity Hill, David K. Jackson, Katy Gaythorpe, Natalie Groves, John Sillitoe, Dominic P. Kwiatkowski, Seth Flaxman, Oliver Ratmann, Samir Bhatt, Susan Hopkins, Axel Gandy, Andrew Rambaut, and Neil M Ferguson. 2021. Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. *medRxiv* (2021),

2020–12. The COVID-19 Genomics UK (COG-UK) consortium.

[236] Patrick GT Walker, Charles Whittaker, Oliver J Watson, Marc Baguelin, Peter Winskill, Arran Hamlet, Bimandra A Djafaara, Zulma Cucunubá, Daniela Olivera Mesa, Will Green, Hayley Thompson, Shevanthi Nayagam, Kylie E C Ainslie, Sangeeta Bhatia, Samir Bhatt, Adhiratha Boonyasiri, Olivia Boyd, Nicholas F Brazeau, Lorenzo Cattarino, Gina Cuomo-Dannenburg, Amy Dighe, Christl A Donnelly, Ilaria Dorigatti, Sabine L van Elsland, Rich FitzJohn, Han Fu, Katy A M Gaythorpe, Lily Geidelberg, Nicholas Grassly, David Haw, Sarah Hayes, Wes Hinsley, Natsuko Imai, David Jorgensen, Edward Knock, Daniel Laydon, Swapnil Mishra, Gemma Nedjati-Gilani, Lucy C Okell, H Juliette Unwin, Robert Verity, Michaela Vollmer, Caroline E Walters, Haowei Wang, Yuanrong Wang, Xiaoyue Xi, David G Lalloo, Neil M Ferguson, and Azra C Ghani. 2020. The impact of COVID-19 and strategies for mitigation and suppression in low-and middle-income countries. *Science* 369, 6502 (2020), 413–422.

[237] Alexandra C Walls, Young-Jun Park, M Alejandra Tortorici, Abigail Wall, Andrew T McGuire, and David Veesler. 2020. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181, 2 (2020), 281–292.

[238] Alexandra C Walls, Xiaoli Xiong, Young-Jun Park, M Alejandra Tortorici, Joost Snijder, Joel Quispe, Elisabetta Cameroni, Robin Gopal, Mian Dai, Antonio Lanzavecchia, Maria Zambon, Félix A Rey, Davide Corti, and David Veesler. 2019. Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. *Cell* 176, 5 (2019), 1026–1039.

[239] Liang Wang, Xavier Didelot, Jing Yang, Gary Wong, Yi Shi, Wenjun Liu, George F Gao, and Yuhai Bi. 2020. Inference of person-to-person transmission of COVID-19 reveals hidden super-spreading events during the early outbreak phase. *Nature Communications* 11, 5006 (2020), 1–6. Issue 1.

[240] 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. *Scientific Reports* 10, 1 (2020), 1–12.

[241] Lili Wang, Yiwang Zhou, Jie He, Bin Zhu, Fei Wang, Lu Tang, Michael Kleinsasser, Daniel Barker, Marisa C Eisenberg, and Peter XK Song. 2020. An epidemiological forecast model and software assessing interventions on the COVID-19 epidemic in China. *Journal of Data Science* 18, 3 (2020), 409–432.

[242] Mao Wang, Aili Jiang, Lijuan Gong, Lina Luo, Wenbin Guo, Chuyi Li, Jing Zheng, Chaoyong Li, Bixing Yang, Jietong Zeng, Youping Chen, Ke Zheng, and Hongyan Li. 2020. Temperature significant change COVID-19 Transmission in 429 cities. *medrxiv* (2020).

[243] Peipei Wang, Xinqi Zheng, Jiayang Li, and Bangren Zhu. 2020. Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. *Chaos, Solitons & Fractals* 139 (2020), 110058.

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

[245] Tianyi Wang, Ke Lu, Kam-Pui Chow, and Qing Zhu. 2020. COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model. *IEEE Access* 8 (2020), 138162–138169.

[246] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. 2017. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *CVPR*. 2097–2106.

[247] Xia Wang, Dorcas Washington, and Georg F Weber. 2021. Complex Systems Analysis Informs on the Spread of COVID-19. *bioRxiv* (2021).

[248] Yan Wang, Shangde Gao, and Wenyu Gao. 2021. Can Predominant Credible Information Suppress Misinformation in Crises? Empirical Studies of Tweets Related to Prevention Measures during COVID-19. *CoRR* abs/2102.00976 (2021).

[249] Yubo Wang, Yong He, Jin Tong, Yalan Qin, Ting Xie, Jianghua Li, Jianrong Li, Jianhua Xiang, Yong Cui, Elizabeth S Higgs, and Jianglin Xiang. 2020. Characterization of an asymptomatic cohort of SARS-COV-2 infected individuals outside of Wuhan, China. *Clinical Infectious Diseases* 71 (2020), 2132–2138. Issue 16.

[250] Joshua S Weitz, Stephen J Beckett, Ashley R Coenen, David Demory, Marian Dominguez-Mirazo, Jonathan Dushoff, Chung-Yin Leung, Guanlin Li, Andreea Măgălie, Sang Woo Park, Rogelio Rodriguez-Gonzalez, Shashwat Shivam, and Conan Y. Zhao. 2020. Modeling shield immunity to reduce COVID-19 epidemic spread. *Nature medicine* 26, 6 (2020), 849–854.

[251] WHO. 2021. Tracking SARS-CoV-2 variants. https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/. (2021).

[252] World Bank. 2020. The global economic outlook during the COVID-19 pandemic: A changed world. *World Bank Feature Story* (2020).

[253] Jiangpeng Wu, Pengyi Zhang, Liting Zhang, Wenbo Meng, Junfeng Li, Chongxiang Tong, Yonghong Li, Jing Cai, Zengwei Yang, Jinhong Zhu, Meie Zhao, Huirong Huang, Xiaodong Xie, and Shuyan Li. 2020. Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results. *MedRxiv* (2020).

[254] Joseph T Wu, Kathy Leung, Mary Bushman, Nishant Kishore, Rene Niehus, Pablo M de Salazar, Benjamin J Cowling, Marc Lipsitch, and Gabriel M Leung. 2020. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nature medicine* 26, 4 (2020), 506–510.

[255] Joseph T Wu, Kathy Leung, and Gabriel M Leung. 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet* 395, 10225 (2020), 689–697.

[256] Shuai Xia, Meiqin Liu, Chao Wang, Wei Xu, Qiaoshuai Lan, Siliang Feng, Feifei Qi, Linlin Bao, Lanying Du, Shuwen Liu, Chuan Qin, Fei Sun, Zhengli Shi, Yun Zhu, Shibo Jiang, and Lu Lu. 2020. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell research* 30, 4 (2020), 343–355.

[257] Jingui Xie and Yongjian Zhu. 2020. Association between ambient temperature and COVID-19 infection in 122 cities from China. *Science of the Total Environment* 724 (2020), 138201.

[258] Jiaqi Xiong, Orly Lipsitz, Flora Nasri, Leanna MW Lui, Hartej Gill, Lee Phan, David Chen-Li, Michelle Iacobucci, Roger Ho, Amna Majeed, and Roger S. McIntyre. 2020. Impact of COVID-19 pandemic on mental health in the general population: A systematic review. *Journal of affective disorders* 277 (2020), 55–64.

[259] Li Yan, Hai-Tao Zhang, Jorge Goncalves, Yang Xiao, Maolin Wang, Yuqi Guo, Chuan Sun, Xiuchuan Tang, Liang Jin, Mingyang Zhang, Xiang Huang, Ying Xiao, Haosen Cao, Yanyan Chen, Tongxin Ren, Fang Wang, Yaru Xiao, Sufang Huang, Xi Tan, Niannian Huang, Bo Jiao, Yong Zhang, Ailin Luo, Laurent Mombaerts, Junyang Jin, Zhiguo Cao, Shusheng Li, Hui Xu, and Ye Yuan. 2020. A machine learning-based model for survival prediction in patients with severe COVID-19 infection. *MedRxiv* (2020).

[260] Li Yan, Hai-Tao Zhang, Yang Xiao, Maolin Wang, Yuqi Guo, Chuan Sun, Xiuchuan Tang, Liang Jing, Shusheng Li, Mingyang Zhang, Ying Xiao, Haosen Cao, Yanyan Chen, Tongxin Ren, Junyang Jin, Fang Wang, Yanru Xiao, Sufang Huang, Xi Tan, Niannian Huang, Bo Jiao, Yong Zhang, Ailin Luo, Zhiguo Cao, Hui Xu, and Ye Yuan. 2020. Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan. *MedRxiv* (2020).

[261] Zifeng Yang, Zhiqi Zeng, Ke Wang, Sook-San Wong, Wenhua Liang, Mark Zanin, Peng Liu, Xudong Cao, Zhongqiang Gao, Zhitong Mai, Jingyi Liang, Xiaoqing Liu, Shiyue Li, Yimin Li, Feng Ye, Weijie Guan, Yifan Yang, Fei Li, Shengmei Luo, Yuqi Xie, Bin Liu, Zhoulang Wang, Shaobo Zhang, Yaonan Wang, Nanshan Zhong, and Jianxing He. 2020. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of thoracic disease* 12 (2020), 165–174. Issue 3.

[262] Yanfang Ye, Shifu Hou, Yujie Fan, Yiyue Qian, Yiming Zhang, Shiyu Sun, Qian Peng, and Kenneth Laparo. 2020. α-Satellite: An AI-driven System and Benchmark Datasets for Hierarchical Community-level Risk Assessment to Help Combat COVID-19. *arXiv:2003.12232* (2020).

[263] CY Yu, KG Chan, CY Yean, and GY Ang. 2021. Nucleic Acid-Based Diagnostic Tests for the Detection SARS-CoV-2: An Update. *Diagnostics (Basel)* 11 (2021), 53:1–37. Issue 1.

[264] Ping Yu, Jiang Zhu, Zhengdong Zhang, and Yingjun Han. 2020. A familial cluster of infection associated with the 2019 novel coronavirus indicating possible person-to-person transmission during the incubation period. *The Journal of infectious diseases* 221, 11 (2020), 1757–1761.

[265] Abdelhafid Zeroual, Fouzi Harrou, Abdelkader Dairi, and Ying Sun. 2020. Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study. *Chaos, Solitons & Fractals* 140 (2020), 110–121.

[266] Gengpei Zhang and Xiongding Liu. 2021. Prediction and control of COVID-19 spreading based on a hybrid intelligent model. *PLoS ONE* 16, e0246360 (2021), 1–13. Issue 2.

[267] Juanjuan Zhang, Maria Litvinova, Yuxia Liang, Yan Wang, Wei Wang, Shanlu Zhao, Qianhui Wu, Stefano Merler, Cécile Viboud, Alessandro Vespignani, Marco Ajelli, and Hongjie Yu. 2020. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* 368, 6498 (2020),

1481–1486.

[268] Kang Zhang, Xiaohong Liu, Jun Shen, Zhihuan Li, Ye Sang, Xingwang Wu, Yunfei Zha, Wenhua Liang, Chengdi Wang, Ke Wang, Linsen Ye, Ming Gao, Zhongguo Zhou, Liang Li, Jin Wang, Zehong Yang, Huimin Cai, Jie Xu, Lei Yang, Wenjia Cai, Wenqin Xu, Shaoxu Wu, Wei Zhang, Shanping Jiang, Lianghong Zheng, Xuan Zhang, Li Wang, Liu Lu, Jiaming Li, Haiping Yin, Winston Wang, Oulan Li, Charlotte Zhang, Liang Liang, Tao Wu, Ruiyun Deng, Kang Wei, Yong Zhou, Ting Chen, Johnson Yiu-NamLau, Manson Fok, Jianxing He, Tianxin Lin, Weimin Li, and Guangyu Wang. 2020. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell* 181, 6 (2020), 1423–1433.

[269] Yi Zhang, Xiaojing Cai, Caroline V Fry, Mengjia Wu, and Caroline Wagner. 2020. Topic Evolution, Disruption and Resilience in Early COVID-19 Research. *Scientometrics* (2020), 1–29.

[270] Alex Zhavoronkov, Vladimir Aladinskiy, Alexander Zhebrak, Bogdan Zagribelnyy, Victor Terentiev, Dmitry S Bezrukov, Daniil Polykovskiy, Rim Shayakhmetov, Andrey Filimonov, Philipp Orekhov, Yilin Yan, Olga Popova, Quentin Vanhaelen, Alex Aliper, and Yan Ivanenkov. 2020. Potential COVID-2019 3C-like protease inhibitors designed using generative deep learning approaches. *chemrxiv* (2020). 10.26434/chemrxiv.11829102.

[271] Alex Zhavoronkov, Bogdan Zagribelnyy, Alexander Zhebrak, Vladimir Aladinskiy, Victor Terentiev, Quentin Vanhaelen, Dmitry S. Bezrukov, Daniil Polykovskiy, Rim Shayakhmetov, Andrey Filimonov, Michael Bishop, Steve McCloskey, Edgardo Leija, Deborah Bright, Keita Funakawa, Yen-Chu Lin, Shih-Hsien Huang, Hsuan-Jen Liao, Alex Aliper, and Yan Ivanenkov. 2020. Potential Non-Covalent SARS-CoV-2 3C-like Protease Inhibitors Designed Using Generative Deep Learning Approaches and Reviewed by Human Medicinal Chemist in Virtual Reality. *ChemRxiv* (2020), 1–18.

[272] Chuansheng Zheng, Xianbo Deng, Qing Fu, Qiang Zhou, Jiapei Feng, Hui Ma, Wenyu Liu, and Xinggang Wang. 2020. Deep learning-based detection for COVID-19 from chest CT using weak label. *MedRxiv* (2020).

[273] Nanning Zheng, Shaoyi Du, Jianji Wang, He Zhang, Wenting Cui, Zijian Kang, Tao Yang, Bin Lou, Yuting Chi, Hong Long, Mei Ma, Qi Yuan, Shupei Zhang, Dong Zhang, Feng Ye, and Jingmin Xin. 2020. Predicting COVID-19 in China using hybrid AI model. *IEEE transactions on cybernetics* 50, 7 (2020), 2891–2904.

[274] Si-qian Zheng, Li Yang, Peng-xiang Zhou, Hui-bo Li, Fang Liu, and Rong-sheng Zhao. 2020. Recommendations and guidance for providing pharmaceutical care services during COVID-19 pandemic: a China perspective. *Research in social and administrative pharmacy* 17 (2020), 1819–1824.

[275] Tianjian Zhou and Yuan Ji. 2020. Semiparametric Bayesian inference for the transmission dynamics of COVID-19 with a state-space model. *Contemporary Clinical Trials* 97 (2020), 106146.