Artificial Intelligence in Covid-19

Niklas Lidströmer • Yonina C. Eldar
Editors

Artificial Intelligence in Covid-19

Springer

Editors
Niklas Lidströmer
Karolinska Institute
Stockholm, Sweden

Yonina C. Eldar
Weizmann Institute of Science
Rehovot, Israel

ISBN 978-3-031-08505-5 ISBN 978-3-031-08506-2 (eBook)
https://doi.org/10.1007/978-3-031-08506-2

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

We were surprised by the COVID-19 pandemic and ill-prepared for its many challenges. Given that we live in the age of artificial intelligence, we should not have been. Artificial intelligence (AI) has emerged as the dominant technology of the past decade, and advances in machine learning, in particular, have led to remarkable innovations, ranging from automated speech recognition systems to self-driving cars. However, prior to the pandemic, the impact of AI on infectious diseases and public health had largely been anaemic, and as a result we were caught flat-footed when the first COVID cases were reported in late 2019.

But as brilliantly captured by this timely book, the AI community quickly stepped up to tackle many of the clinical and epidemiological challenges presented by the pandemic. As you will read in these pages, AI has been harnessed in creative and diverse ways to detect and treat COVID-19, as well as to develop predictive models to identify emerging threats. Tremendous advances have been achieved throughout the pandemic, but much work remains. As highlighted by this book, these outstanding and future challenges will best be tackled by multi-lingual scientific and medical teams that thoughtfully integrate human intelligence with machine intelligence.

We all need hope in the face of challenges. This book offers us scientifically grounded hope, both for the current pandemic and the next one. The tools of AI provide us with a much-needed, unfair advantage in the battle of human wits versus the genes of emerging pathogens. Thanks to AI, we will be better prepared for the next pandemic.

MIT and Harvard
Cambridge, MA, USA

James J. Collins

Preface

In 2020, earth was struck by the fifth deadliest pandemic in recorded history. It happened at the same time as modern Artificial Intelligence (AI) systems have been impacting a wide range of areas from self-driving cars to automatic speech and language recognition and healthcare.
AI has the potential to profoundly change all areas of healthcare from applications that can understand and classify clinical documentation, through automating various tasks to analysing big patient data sets, and development of algorithms for spotting tumours and other diseases, and guiding researchers in how to construct cohorts for clinical trials.

AI has pivotal importance in the way we fight the present and future pandemics. Here we present a plethora of areas where AI has been applied against COVID-19 and suggest areas in which AI can develop in order to better address future pandemics.

This book is a practical, scientific, and clinically relevant example of how clinicians and researchers in medicine, computer science, mathematics, biology, and epidemiology can fight future disasters on multiple fronts, and how scientific integration is crucial for progress.

The book aims at fellow researchers in areas ranging from medicine and biology to computer science and mathematics, medical specialists and clinicians, medical and pharmaceutical residents, and students, or interested academics in general, who wish to know about the latest usage of AI in this medical field. Equipped with this book the reader will learn about the latest AI advances against COVID-19, and how mathematics and algorithms can aid in preventing its spreading course, treatments, diagnostics, vaccines, clinical management, and future evolution.

We believe the book is needed both directly because of the seriousness of the pandemic per se and because of the spread of misinformation. It is important that the public and global organisations recognise more explicitly the *role of science,* for it is with evidence-based science we can solve this incumbent and future threats to our world and humanity.

We wish to acknowledge and express our deep appreciation to all contributing authors of this book, the Springer Nature publishing team, and our supporting families who suffered the pandemic with us.

AU2

Stockholm, Sweden
Rehovot, Israel
4 August 2022

Niklas Lidströmer
Yonina C. Eldar

Contents

About the Editors

Niklas Lidströmer Karolinska Institutet, MD, MSc, specialist physician, postgraduate researcher in AI in medicine, senior advisor in AI and medical investments, former AI entrepreneur and founder of an AI-powered medical platform, former head of Medical AI at a variety of med-tech companies, and also previous co-leader of a handful of successful medical start-ups.

His experience also encompasses widespread global clinical work spanning **20** years within numerous regions across eight countries. After graduating with a master's thesis on global medicine in **2000**, he began practicing as a medical doctor in **2002**, followed by internship, specialised residencies, and clinical work all over the world, including **1** year circumnavigating as a maritime **doctor.**

His international work experience, fluency in nearly ten languages, practical familiarisation with AI in the medical and pharmaceutical industries, and clinical specialist competence in general medicine have produced a passion for translational and educational aspects of artificial intelligence in m**edicine.**

Dr. Niklas Lidströmer edited the pivotal reference work—the new standard reference, *Artificial Intelligence in Medicine*, Springer Nature (**130** chapters, **1858** pages), which has now become the largest and most comprehensive in the scientific community. The two volumes of this work are visible in the portrai**t image.**

Homepage: https://staff.ki.se/people/niklas-li**dstromer**

Yonina C. Eldar is a Professor in the Department of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, Israel, where she heads the Center for Biomedical Engineering and Signal Processing and holds the Dorothy and Patrick Gorman Professorial Chair. She is also a Visiting Professor at MIT, a Visiting Scientist at the Broad Institute, and an Adjunct Professor at Duke University and was a Visiting Professor at Stanford. She is a member of the Israel Academy of Sciences and Humanities, an IEEE Fellow, and a EURASIP Fellow.

She received a B.Sc. degree in physics and a B.Sc. degree in electrical engineering from Tel-Aviv University, and a Ph.**D.** degree in electrical engineering and computer science from MIT, in **2002**. She has received many awards for excellence in research and teaching, including the IEEE Signal Processing Society Technical Achievement Award (**2013**), the IEEE/AESS Fred Nathanson Memorial Radar Award (**2014**). and the IEEE Kiyo Tomiyasu Award **(2016).**

She was a Horev Fellow of the Leaders in Science and Technology programme at the Technion and an Alon Fellow. She received the Michael Bruno Memorial Award from the Rothschild Foundation, the Weizmann Prize for Exact Sciences, the Wolf Foundation Krill Prize for Excellence in Scientific Research, the Henry Taub Prize for Excellence in Research (twice), the Hershel Rich Innovation Award (three times), and the Award for Women with Distinguished Contri**butions.**

Professor Yonina C. Eldar received several best paper awards and best demo awards together with her research students and colleagues, was selected as one of the 50 most influential women in Israel, and was a member of the Israel Committee for Higher Education. She is the Editor in Chief of *Foundations and Trends in Signal Processing*, a member of several IEEE Technical Committees and Award Committees, and heads the Committee for Promoting Gender Fairness in Higher Education Institutions in Israel.

More on homepage: http://www.wisdom.weizmann.ac.il/~yonina/Yoni**naEldar/**

Author Query

Chapter No.: 0005396216

Query	Details Required	Author's Response
AU1	Please check if edit to the sentence "We wish to acknowledge ..." is okay.	

Chapter 1
Introduction to Artificial Intelligence in COVID-19

Niklas Lidströmer and Yonina C. Eldar

In this introduction we also presprent basic definitions and concepts, and a general review, where we put the present pandemic into a historic and scientific context. The ongoing pandemic comes into a grander evolutionary, scientific, historic and humanistic perspective. With the use of basic scientific method, the bases for evidence, the foundations of reason and theorem—science must be front and centre of the agenda for international pandemic collaboration—we present an armamentarium of artificial intelligence in medicine to accomplish this benevolent and profound intension now and in the future.

Given the tense situation in the World, with ongoing parallel pandemics and a major war in Europe at the time of this publication, the spread of misinformation, dissemination of conspiracy theories, and recurrent violent protests and actions from, e.g., anti-vaxxers, extremists of various orientations and the rapid increase of fake news and abundancy of anti-scientific ideas and beliefs, we also take the opportunity to shed light on various scientific aspects related to COVID-19.

Pandemics

Epidemics (from Greek ἐπί epi *above* and δῆμος demos *people*) have accompanied humanity throughout all of recorded history. A usual definition of a pandemic (from Greek πᾶν pan *all* and δῆμος demos *people*) is that it is an epidemic, which has

N. Lidströmer (✉)
Karolinska Institute, Stockholm, Sweden
e-mail: niklas.lidstromer@ki.se

Y. C. Eldar
Weizmann Institute of Science, Rehovot, Israel
e-mail: yonina.eldar@weizmann.ac.il

N. Lidströmer, Y. C. Eldar (eds.), *Artificial Intelligence in Covid-19*,
https://doi.org/10.1007/978-3-031-08506-2_1

spread across a wider geographic range than an epidemic, and which has affected a significant portion of the population. Into the nineteenth century, pandemic and epidemic remained synonyms. In the 1828 edition of the Webster dictionary, Noah Webster defined pandemic as 'Incident to a whole people; epidemic; as a pandemic disease' [1]. The 1870s and 1880s saw the word pandemic used more widely and then, as David Morens, Gregory Folker and Anthony Fauci argued in 2009, the violent spread of the 1889–1892 pandemic brought the term 'into general use'. By the time of the 1918 flu, it was a 'household' word [2].

At the time of the spread of the pandemic of 1889–1892, the *Russian 'Flu*'—we shall get back to why the term the *Russian* '*Cold*' may have been a better terminology—a written evidence of the first epidemic was discovered. In 1887 a peasant woman discovered a 43 lines' clay tablet in the soil of Tel el-Amarna in Egypt (ancient Akhetaten). The tablet was written in Akkadian, giving us the first record of an epidemic; the *Plague of Megiddo* in the land of Canaan in 1350 BCE. This source is a part of the Amarna letters. Here Biridiya, the mayor of Megiddo, writes a complaint letter to the pharaoh (most likely Amenhotep III). In the lines 30–32 he writes [3]:

30	*Šum-ma-me ga-am-ra-at-me alu*^{ki}	Look, the city is consumed
31	*i-na mûti*	by death,
32	*i-na mu-ta-a-an*	by plague

However, it is unknown how many died in this epidemic. The first instance, when there is a solid written documentation, and with a qualified estimate of the victims, ca 75,000–100,000, is the *Plague of Athens* in 429–426 BCE. This epidemic can be regarded as a pandemic, the first recorded, since it is known to have affected Greece, Libya, Egypt and Ethiopia. The causing agent is unknown, but based on descriptions; typhus, typhoid fever [4] or viral haemorrhagic fever have been suggested [5]. The causing agent can be a variety of microorganisms in epidemics; the pathogenic organisms are of five main types: viruses, bacteria, fungi, protozoa, and worms.

Viruses are found wherever there is life and have probably existed since living cells first evolved [6]. In the evolution cells first emerged at least 3.8 billion years ago [7], which was approximately 750 million years after Earth was formed [6]. They existed 3.5 billion years before humans evolved on Earth. They're neither dead nor alive. Their genetic material is embedded in our own DNA, constituting close to 10% of the human genome. A key step in the viruses' evolutionary journey seems to have come about around 1.5 billion years ago—that's the age at which researchers have estimated 66 virus-specific protein folds came on the scene. These changes were affecting the proteins in the virus' outer coat—the machinery viruses use to break into host cells [8].

Bacteria have existed from very early in the history of life on Earth. Bacterial fossils discovered in rocks date from at least the Devonian Period (419.2 million to 358.9 million years ago), and there are convincing arguments that bacteria have been present since early Precambrian time, about 3.5 billion years ago [9].

Artificial intelligence (AI) has the potential to deeply improve healthcare. The past decade has witnessed a deep learning, or AI, revolution. The availability of

large-scale training data sets, which is often facilitated by Internet content; the accessibility of powerful computational resources thanks to breakthroughs in microelectronics; and advances in neural network research, such as the development of effective network architectures and efficient training algorithms, have resulted in the unprecedented success of deep learning in innumerable applications of computer vision, pattern recognition, and speech processing. More recently, this revolution has begun impacting various areas of medicine where it is used to support medical professionals in clinical settings and in ongoing research including image analysis and decision support.

Standards for AI in medicine are still being defined, but it is clear that future medicine will have many integrated AI components. For example, in oncology, AI can be used in a tailored way to analyse the complicated latticework of correlations between a large range of both genome influences and environmental factors which add up and finally constitute the cancer risk. The next-generation cancer treatment is inspired by the Evolutionary theory concepts of such as treatment vaccination, evolutionary traps and adaptive therapies [10].

In infectious disease spread in pandemics AI can provide rapid detection of the infectious agent and diagnosis of the pathogen's most updated pharmaceutical resistance profile, and also predictions of the resistance development; a dynamic forecasting of pandemic pathogen. With the use of anticipatory diagnostics, pharmaceutical regimens can be finely tuned to probabilistically channel pathogens toward less resistance-prone genotypes and act at the root cause of the problem, i.e., evading the development of resistance.

In antimicrobial drug design and to anticipate outbreaks of infectious disease, advanced computational methods can be utilized, and also to map the evolution of epidemics, e.g., as has been done in the SARS-CoV-2 pandemic. Some of the chapters highlight and detail this usage with illustrative examples of the fruitful collaboration of computer scientists, evolutionary or biological experts, public health specialists, clinicians and epidemiologists.

Evolutionary experts have suggested the dangers of an over-reliance on advanced computational tools involving "black box" algorithms and also questioned whether they undermine the synthesis of Mendelian genetics and Darwinian theory. Yet, insofar as evolutionary theory is used for hypothesis algorithm development and AI for data creation and analysis, this problem may be avoided, and the potential of both will be realised [10]. Many of the chapters in this book show how careful and purposeful use of AI can lead to important applications in various aspects of detection, analysis and treatment of COVID-19.

History of Pandemics

There have been a long range of pandemic diseases. The most fatal pandemic in recorded history was the *Black Death* (also known as *The Plague*), which killed an estimated 75–200 million people in the years around 1350, which could have been up to half of the population of Eurasia [11].

The great *influenza pandemic of 1918*, also colloquially known as the misnomer *Spanish flu* or as the *Great Influenza* epidemic, was an exceptionally deadly global influenza pandemic caused by the H1N1 influenza A virus stem. The first documentation, began with cases appearing in March 1918; firstly, in Kansas, United States, and then with further cases in France, the United Kingdom and Germany in April. Only 24 months later, almost a third of the global population, which corresponded to ca. 500 million at the time, had been infected in the four consecutive waves. The death toll estimates range from 17 to 50 million, but 100 million cannot be ruled out, making it the second deadliest pandemic in all recorded history [12].

For at least 3000 years *smallpox* had devastated humanity. *Variola major* and *Variola minor* were the two viral variants of smallpox. The agent of *variola virus* (VARV) belongs to the genus *Orthopoxvirus*. In October 1977 the last naturally occurring case was diagnosed. The World Health Organization launched an intensified plan to eradicate smallpox in 1967. Widespread immunization and surveillance were conducted around the world for several years. The last known natural case was in Somalia in 1977. In 1980 WHO declared smallpox eradicated—the only infectious disease to achieve this distinction. This remains among the most notable and profound public health successes in history [13].

The *Mycobacterium tuberculosis* bacteria causes *Tuberculosis* (TB), which has been present in humans since at least ancient times [14]. It generally affects the lungs, but it can also affect other parts of the body. Most infected people show no symptoms at all, i.e., latent tuberculosis, but ca. 10% of these cases progress to active disease which, if left without treatment, is lethal in 50% of the affected. If symptomatic, TB causes chronic cough with haemoptysis (blood-stained mucous), night sweats, fever and weight loss [15]. In older medical literature, doctors often referred to the disease as *consumption,* since it often gives an articulated weight loss [16], but a wide range of symptoms are associated with this condition [17].

Still only 2 years before the COVID-19 pandemic, in 2018, it was estimated about 25% of the global population were bearers of dormant TB [18]. De novo infections have an incidence of 1% per annum [19]. At the same time as the World was alerted by the start of the pandemic, i.e., in 2020, ca. 10 million people developed a TB activation, and 1.5 million died. In 2020 it was the second leading cause of death from an infectious disease after COVID-19 [20].

In parallel to the pandemic, TB remained manifest in swathes of South-East Asia, Africa, and the Western Pacific, and with more than half of the cases diagnosed in a handful of countries: India, China, Indonesia, the Philippines, Pakistan, Nigeria, and Bangladesh [21].

Amidst the raging pandemic in 2021, with all its social distancing etc, the number of new TB cases were simultaneously decreasing by 2% per annum. In many African and Asian countries ca 80% of the population are tuberculin test positive, compared to 10% in the United States [22].

The *seventh cholera pandemic* (1961–1975), with acute diarrhoeal infections caused by the ingestion of food or water contaminated with the bacterium *Vibrio cholerae*, and which involved *El Tor* strain persists to the present. [23]. This

pandemic started in Indonesia and the following years spread to Bangladesh, India and the Soviet Union. In 1970 there was a major outbreak in Odessa, in present day Ukraine and later also in Baku, in Azerbaijan, but all information was stopped by the Soviet Union. The pandemic also spread to North Africa and the South Pacific, with a total of 155,000 cases in 1971. In 1991 in the total reported cases had reached over half a million. The rapid disease spread was facilitated by both vast migrations crises and the steep increase in modern mass transportation. However, the use of both modern curative and preventive governmental programs led to marked decrease in mortality. The original mortality, hitting every second patient, decreased to one in ten by 1980, and then to less than 3% during the 1990s [24]. When the strain re-entered South America in the early 1990s in Peru, it killed approximately 10,000 people [25]. Research has traced the origin of the strain to Africa [26].

The *Hong Kong flu* in 1968–1969, was one of the deadliest pandemics in history and killed between one and four million peoples globally, and was caused by an H3N2 strain of the influenza A virus [27]. The Hong Kong strain descended from the H2N2 strain of the *Asian flu pandemic* in 1957–1958. The process for this development went on via *antigenic shift*, i.e., the genes of several influenza A strain subtypes went through reassortment and emanated into a new virus [28].

In 1981 began the global pandemic of *HIV* (Human Immunodeficiency Virus) and *AIDS* (Acquired Immunodeficiency Syndrome). It is an ongoing global public health issue, which according to the World Health Organisation (WHO), until 2021 has killed ca. 36 million people, and ca. 38 million people are infected with HIV globally [29]. Of those infected, around 73% have access to antiretroviral treatment, and 16% are not aware that they are bearers of the infection [30].

Two years before the COVID-19 pandemic in 2018, there were ca. 770,000 deaths from HIV/AIDS [31]. During 2020, which will be remembered in history as the year when COVID-19 was declared a pandemic, there were also 680,000 deaths from HIV/AIDS [31]. The *Global Burden of Disease Study* from 2015 calculates an infection peak of HIV in 1997 with 3.3 million cases per annum; a figure that has decreased rapidly, especially due to much fewer cases in eastern and southern Africa [32]. In 2020 the global HIV incidence had dropped to ca. 1.5 million per annum [33].

The 1889–1890 pandemic, also known as the *Asiatic flu* or *Russian flu* [34], was a worldwide pandemic with a respiratory virus. It was the last major pandemic of the nineteenth century, and is counted among the deadliest pandemics in history [35]. The pandemic killed about one million people out of a world population of almost 1.5 billion (ca. 0.07%). A majority of the reported effects of the pandemic occurred from autumn 1889 until the end of 1890, with relapses in the second quarter of 1891, late autumn 1891 to early summer 1892, the winter of 1893–1894 in the Nordics, and the last wave in early 1895 [36].

Although contemporary sources described the pandemic as influenza and scholars during the last century identified several influenza strains as the possible pathological agent, some recent researchers in the 2000s have suggested, with interesting evidences as we shall see, that it may have been caused by human coronavirus OC43 [37].

Human coronavirus OC43 (HCoV-OC43) is a member of the species *Betacoronavirus 1* [38], which infects both humans and cattle [39]. The infecting coronavirus is an enveloped, positive-sense (in molecular biology and genetics, the sense of a nucleic acid molecule, particularly of a strand of DNA or RNA, refers to the nature of the roles of the strand and its complement in specifying a sequence of amino acids), single-stranded RNA virus, which enters its host cell by binding to the N-acetyl-9-O-acetylneuraminic acid receptor [40]. There are seven coronaviruses known to infect humans, and OC43 is one of these, and is one of the viruses responsible for the *common cold* [41], and may have been responsible for the 1889–1890 pandemic [42]. It has, like other coronaviruses from the genus *Betacoronavirus,* subgenus *Embecovirus*, an additional shorter spike protein called hemagglutinin esterase (HE) [43].

As mentioned above, the identification of the virus responsible for the Russian pandemic has taken over a century—a noteworthy comparison, when immediate origin tracing of COVID-19 is expected by the public. At first, an influenza virus was suspected. Researchers have tried for many years to identify the subtypes of Influenza A, deemed to be responsible for the 1889–1890, 1898–1900 and 1918 epidemics. Originally, this research was mainly focused on *seroarcheology* [44], i.e., the detection of antibodies to influenza infection in the sera of elderly people, who may have been exposed to the virus. It was assumed that the 1889–1890 pandemic was caused by Influenza A subtype *H2*, the 1898–1900 epidemic by subtype *H3*, and the 1918 pandemic by subtype *H1* [45].

Later it was confirmed that the *H1N1* was indeed the cause of the 1918 flu pandemic following identification of H1N1 antibodies in exhumed corpses. This led to a new analysis of the gathered seroarcheological data, which then suggested Influenza A subtype H3 (perhaps the H3N8 subtype) as a preliminary more plausible cause for the 1889–1890 pandemic [46].

But in the early 2000s the pathogen was suspected to have been a coronavirus. After the 2002–2004 SARS outbreak, virological researchers commenced the sequencing both human and animal coronaviruses. The two virus strains in the Betacoronavirus 1 species were compared. The evolutionary analysis of the bovine coronavirus and human coronavirus OC43, pointed towards that they had a most recent *common ancestor* in the late 1800s [47]. With several methods yielded, independently of each other, the most probable dates around 1890. When the results were published ca. 15 years ago, the authors speculated that an introduction of the bovine coronavirus strain to the human population might have caused the 1889 epidemic [48].

In the Corona year of 2020, the Danish researchers Lone Simonsen and Anders Gorm Pedersen noted that the in medical records manifestations, found in contemporary medical records, of the 1889 pandemic, i.e., runny nose, high fever, headache, severe chest inflammation, speeding up latent respiratory diseases, and

primarily killing the elder population, did indeed resemble the current COVID-19 symptomatology, a disease caused by a coronavirus, much more than an influenza [49].

The Danish researchers were able to calculate that the human coronavirus OC43 had split from bovine coronavirus ca. 130 years earlier, i.e., hence coinciding with the Russian pandemic in 1889–1890. Before the Danes had published it at the end of 2020, also genetic comparisons between bovine coronavirus and different strains of OC43, constituted the mathematical basis of the calculations by a Dutch team [48]. In November 2020 they had not yet published their results, and another team at University of Leuven in Belgium, did simultaneously execute a similar calculation of OC43. This team independently identified the crossover date to the last decade of the 1800s [50].

In 2021, again the contemporary medical records were systematically analysed, and it was again concluded that the clinical manifestations resembled those of COVID-19 rather than influenza. The similarities with a typical loss of taste and smell perception, long term sequelae similar to post-Covid, and the big picture of a multisystem disease. Moreover, the central nervous system symptoms described in medical records back then and today showed a congruent latticework [37, 50].

Also epidemiologically, it has been noticed that the mortality curve of the Russian Flu was *J-shaped* as found in COVID-19 (little mortality in the very young, high mortality in the old), while the influenza infections have *U-shaped* mortality curves (high mortality in the very young and very old). Although, a scientific consensus that the Russian pandemic was caused by a coronavirus, has not been reached when we write this introduction. A systematic review of the literature suggests that the evidence is to be considered hypothetical [51].

Regarding other pandemics, e.g., the *1957–1958 Asian flu*, it is a clear it was caused by the influenza A virus subtype H2N2, and that it originated in Guizhou in southern China [52]. The number of excess deaths caused by the pandemic, has been estimated to ca. one to four million, hence one of the deadliest pandemics in human history. A reassorted strain of the same virus, the H3N2 has been identified as the causative agent of the *Hong Kong flu* in 1968–1969 [28].

The COVID-19 Pandemic

The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) caused the ongoing COVID-19 pandemic. This novel virus was first identified from an outbreak in the Chinese city of *Wuhan* in December 2019. The *World Health Organization* (WHO) declared a *Public Health Emergency of International Concern*

on 30 January 2020 and a *pandemic* on 11 March 2020. Until March 2022 more than 466 million cases and 6.06 million deaths *officially*, making it one of the deadliest in history. Recent calculations have suggested higher figures, see below.

An attempt to list the ca. 20 deadliest pandemics with all over one million deaths throughout recorded history

#	Common epidemy name	Disease/Causative agent	Deaths in millions	Occurrence	Ref.
1	Black Death	Bubonic plague/Yersinia Pestis	75–200	1346–1353	[53]
2	Spanish Flu	Influenza A/H1N1	17–100	1918–1920	[54]
3	Plague of Justinian	Bubonic plague/Yersinia Pestis	15–100	541–549	[55]
4	HIV Pandemic	AIDS/HIV	27.2–47.8 (2020)	1981–now	[29]
5	COVID-19 Pandemic	COVID-19/SARS-CoV-2	6.09-18.6 (March 2022)	2019–now	[56]
6	Third Plague Pandemic	Bubonic plague/Yersinia Pestis	12–15	1855–1960	[57]
7	Cocoliztli Epidemic	Cocoliztli/Agent discussed	5–15	1545–1548	[58]
8	Antonine Plague	Smallpox or measles	5–10	165–180	[59]
9	Mexican Epidemic	Smallpox	5–8	1519–1520	[60]
10	Russian Epidemic	Typhus	2–3	1918–1922	[61]
11	Influenza Pandemic	Influenza A/H2N2	1–4	1957–1958	[62]
12	Hong Kong Flu	Influenza A/H3N2	1–4	1968–1969	[63]
13	Cocoliztli Epidemic	Cocoliztli/Agent discussed	2–2.5	1576–1580	[64]
14	Japanese Epidemic	Smallpox	2	735–737	[65]
15	Persian Plague	Bubonic plague/Yersinia Pestis	2	1772–1773	[66]
16	Naples Plague	Bubonic plague/Yersinia Pestis	1.25	1656–1658	[67]
17	Cholera Pandemic	Cholera	1+	1846–1860	[68]
18	Italian Plague	Bubonic plague/Yersinia Pestis	1	1629–1631	[23]
19	Russian “Flu”	Corona recently suggested	1	1889–1895	[69]

A recent article in *Nature*, presents a modelling, which suggests three times more deaths than the official figures, i.e., as much as 18 million deaths in the pandemic until the end 2021 [56].

Official death counts have been claimed to underreport the actual death toll, because excess mortality (the number of deaths in a period compared to a long-term average) data show an increase in deaths that is not explained by COVID-19 deaths alone. Using such data, estimates of the true number of deaths from COVID-19 worldwide have included a range from 9.5 to 18.6 million by *The Economist* [70],

as well as over 10.3 million by the Institute for Health Metrics and Evaluation. Such deaths include deaths due to healthcare capacity constraints and priorities, as well as reluctance to seek care (to avoid possible infection) [71].

The evolutionary origins of some recent SARS viruses during the recent 20 years, have been suggested to arise from a "common ancestor" in the form of the Russian Pandemic, which started in 1888. Also, the causative agent of the Plague, Yersinia Pestis, had been suggested for many years, when it was finally proved to be correct only in the last decade. Likewise, the Spanish Flu was classified quite recently. These are three important examples of the recent rise of *viral archaeology* [72]. The scientific community has proved the causative agent of all major pandemics in modern times and are gaining an increasingly solid foundation for pandemics in earlier history [73].

The evolution of human-virus associations is usually reconstructed from contemporary patterns of genomic diversity. An intriguing, though still rarely implemented, alternative is to search for the genetic material of viruses in archaeological and medical archive specimens to document evolution as it happened. Ancient DNA research has incorporated insights from virology to explore the potential range of applications and likely limitations of *archeovirological* approaches. Hopefully numerous questions may be answered or tackled by archeovirology in the near future, and also the main expected roadblocks to these avenues of research [74].

In these times of the COVID-19 pandemic and sociological consequences such as antivaxxer demonstrations etc., it is important to bring up the *Renaissance*, and the *Enlightenment*, the development of the scientific method, logic and reason. To yet again bring forward the great thinkers of the Enlightenment, e.g., *Voltaire*, who believed above all in the *efficacy of reason*. He believed social progress could be achieved through reason and that no authority—religious or political or otherwise—should be immune to challenge by reason [75].

In his 1727 *Essay on Epic Poetry*, he popularised the story of Newton being struck by a stray apple. He preached the importance of trusting the knowledge one gains from their own senses, a philosophy better known as empiricism. Voltaire shared what knowledge and good beliefs he found, regardless of how such sharing would personally affect him [76]. Voltaire's failure to produce an original philosophy was, in a sense, counterbalanced by his deliberate cultivation of a philosophy of action; his 'common sense' crusade against superstition and prejudice and in favour of religious toleration was his single greatest contribution to the progress of Enlightenment [75].

As early as 1742, Voltaire trumpeted the benefits of smallpox inoculation, a practice that preceded Edward Jenner's invention of modern vaccination. Inoculation used small doses of the smallpox virus itself to create future immunity. Voltaire wrote [77]:

> *Out of a hundred people in the world at least sixty have smallpox, and of these sixty, twenty die of it in the flower of their youth and twenty keep the unpleasant marks for ever. That*

> *makes one fifth of all human beings that this disease kills or permanently disfigures. Of all those inoculated in Turkey or England not one dies unless he is infirm and predisposed to die anyway, nobody is disfigured, nobody has smallpox a second time, assuming that the inoculation was properly done* [78].

George Washington put Voltaire's view into practice. In 1777, he ordered a doctor to inoculate all soldiers in the Continental Army who came through Philadelphia. *Necessity not only authorises but seems to require the measure*, he wrote, *for should the disorder infect the Army … we should have more to dread from it, than from the Sword of the Enemy.* The following month, he wrote to the Continental Congress about his plan [79]:

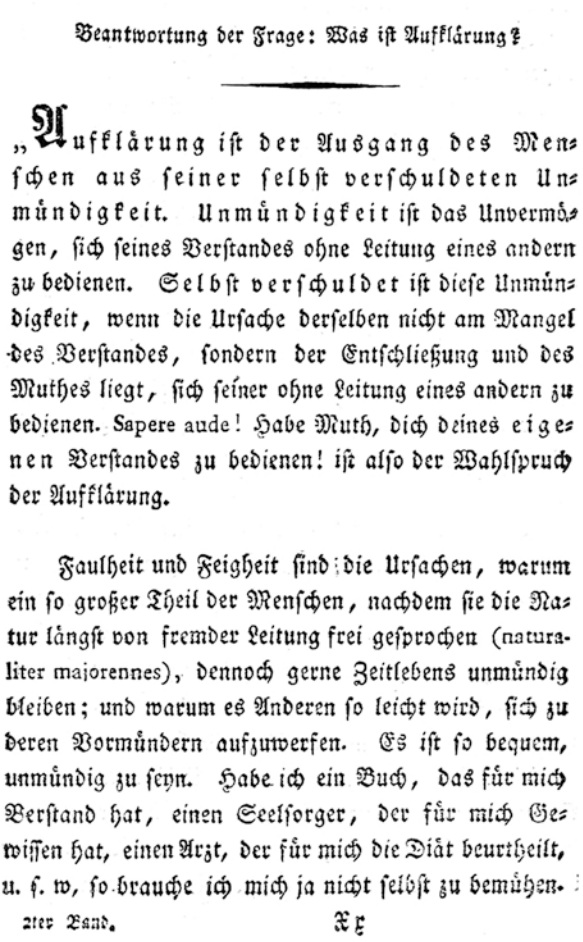
> *The small pox has made such Head in every Quarter that I find it impossible to keep it from spreading thro' the whole Army in the natural way. I have therefore determined, not only to inoculate all the Troops now here, that have not had it, but shall order Dr Shippen to inoculate the Recruits as fast as they come in to Philadelphia* [80].

Contemporary generals are maybe more fainthearted than General Washington, perhaps avoiding to demand mandatory COVID vaccination for the military. The irony is though that smallpox inoculation was, incomparably, far more dangerous than any of the modern vaccinations, including those against the SARS-CoV-2 virus [81].

Several misinformation campaigns, e.g., from anti-vaxxers, have gained traction online amidst rising COVID-19 infection rates worldwide [82], using the same debunked *standard set of strategies*—repetitively since the very start of vaccinations over 200 years ago [83]. The same arguments can be found in pamphlets, which then circulated against smallpox vaccination despite the evidence in favour of its effectiveness [84]. These arguments are not new and have changed little over time. Learning to recognize their repackaging in modern form can help with effectively combating their power [83]. Despite mortality rates between 30% and 40%, and the extreme contagiousness of the disease, it was common for anti-vaccinationists to claim that smallpox was only a minor threat to the population [83]. The *minimization of threat* is a common tactic in contemporary debates as well. Many who promote the anti-vaccination agenda claim vaccines to be more dangerous than the disease [85].

As Voltaire, Immanuel Kant valued the essential ideals of the Enlightenment. In an essay entitled *What Is Enlightenment?* (1784), he contended that the Enlightenment marked a new way of thinking and eloquently affirmed the Enlightenment's confidence in and commitment to reason [86]. Kant answers the question in the first sentence of the essay: *Enlightenment is man's emergence from his self-incurred immaturity* (German: *Unmündigkeit*) [86]. He argues that: *immaturity is self-inflicted not from a lack of understanding, but from the lack of courage to use one's reason, intellect, and wisdom without the guidance of another* [86]. He exclaims that the motto of the Enlightenment is *Sapere aude*! (Latin: Dare to be wise!) [86] (Fig. 1.1).

AU1



Beantwortung der Frage: Was iſt Aufklärung?

„Aufklärung iſt der Ausgang des Menſchen aus ſeiner ſelbſt verſchuldeten Unmündigkeit. Unmündigkeit iſt das Unvermögen, ſich ſeines Verſtandes ohne Leitung eines andern zu bedienen. Selbſt verſchuldet iſt dieſe Unmündigkeit, wenn die Urſache derſelben nicht am Mangel des Verſtandes, ſondern der Entſchließung und des Muthes liegt, ſich ſeiner ohne Leitung eines andern zu bedienen. Sapere aude! Habe Muth, dich deines eigenen Verſtandes zu bedienen! iſt alſo der Wahlſpruch der Aufklärung.

Faulheit und Feigheit ſind die Urſachen, warum ein ſo großer Theil der Menſchen, nachdem ſie die Natur längſt von fremder Leitung frei geſprochen (naturaliter majorennes), dennoch gerne Zeitlebens unmündig bleiben; und warum es Anderen ſo leicht wird, ſich zu deren Vormündern aufzuwerfen. Es iſt ſo bequem, unmündig zu ſeyn. Habe ich ein Buch, das für mich Verſtand hat, einen Seelſorger, der für mich Gewiſſen hat, einen Arzt, der für mich die Diät beurtheilt, u. ſ. w, ſo brauche ich mich ja nicht ſelbſt zu bemühen.

4ter Band. Xx

Fig. 1.1 The first page of the 1799 edition of Immanuel Kant's, *Answering the Question: What Is Enlightenment?* Original in German: *Beantwortung der Frage: Was ist Aufklärung*. Beginn des Traktats. (EA 1784). (Immanuel Kants vermischte Schriften; Hrsg.: Johann Heinrich Tieftrunk; Bd. 2. Halle 1799). In the December 1784 publication of the Berlinische Monatsschrift (Berlin Monthly), edited by Friedrich Gedike and Johann Erich Biester, Kant replied to the question posed a year earlier by the Reverend Johann Friedrich Zöllner, who was also an official in the Prussian government. (Photo of original edition: Dr. Niklas Lidströmer)

Origins of the COVID-19 Pandemic

Today there are also widespread delusions about the spread of HIV, Ebola, influenza, and now also of *COVID-19*. The pandemic has, at times, sadly been caught in a superpower struggle, with the accusation of one country *to have caused this plague* or that another country *controls the WHO*—the political debacle may have impacted WHO's effective response to coronavirus [87].

Pandemics can arise from the very pure evolution, and from zoonotic jumps. It has occurred before, it has been a constant ingredient of the evolution in nature, including of ourselves—ancient viruses are buried in our DNA. The human genome contains billions of pieces of information and around 22,000 genes, but not all of it

is, strictly speaking, human. Eight percent of our DNA consists of remnants of ancient viruses, and another 40% is made up of repetitive strings of genetic letters that is also thought to have a viral origin. Those extensive viral regions are much more than evolutionary relics: They may be deeply involved with a wide range of diseases [88].

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a *novel* severe acute respiratory syndrome coronavirus. It was first isolated from three people with pneumonia connected to the cluster of acute respiratory illness cases in Wuhan, but *all structural features* of the *novel* SARS-CoV-2 virus particle occur in related coronaviruses *in nature* [89].

Over 2 years since the first cases started appearing in Wuhan, China, there is much we don't know about the *exact* origins of SARS-CoV-2, the virus causing COVID-19. Though, recent history offers little promise for exactness to come quickly. For example, about 14 years elapsed between the identification of HIV as the virus that caused AIDS and a demonstration of its modern transition to humans from a specific group of chimpanzees, although this had been suspected some years earlier [90]. About a decade passed from the time of the 2003–2004 epidemic of SARS and definitive delineation of the origin of its causal coronavirus, and 7 years passed before the 2009–2010 influenza pandemic was shown to have originated in Mexican swine [91]. Working with a hypothesis and drawing conclusions based on observation is something which requires a high-standing patience.

The alternative possibility to a natural origin—a laboratory leak—was originally deemed to be difficult to definitively prove or disprove—when the pandemic started, but now we have more research supporting natural processes. First of all, SARS-CoV-2 is closely related to the original SARS-CoV [92]. Secondly, the coronavirus clusters genetically with the genus Betacoronavirus, previously mentioned in this introduction, and clusters in the subgenus *Sarbecovirus* (lineage B) together with two other bat-derived strains. Thirdly, several studies, have independently from each other, proved it is *96% identical* at the complete genome level with other bat coronavirus specimens (*BatCov RaTG13*) [93, 94]. Even the individual protein parts have been analysed, showing even greater resemblance, e.g., the M-protein of SARS-CoV and SARS-CoV2 shows a 98% homology [95].

The P4 epidemiological laboratory at the Wuhan Institute of Virology in Wuhan in China's central Hubei was built in cooperation with French bio-industrial firm Institut Merieux and the Chinese Academy of Sciences. The facility is among a handful of labs around the world cleared to handle Class 4 pathogens (P4)—dangerous viruses that pose a high risk of person-to-person transmission [96].

Several studies have now concluded the *Wuhan market* was the epicentre of the pandemic's start—the coronavirus SARS-CoV-2 jumped to people from animals sold at the market on two occasions in late 2019, but some scientists want more definitive evidence [97].

In an article in *Nature* a year ago, virologists such as Michael Worobey, University of Arizona in Tucson, said that his thinking on the origins of COVID-19 had shifted. In May 2021, he led a letter published in *Science* in which he and others pressed the

scientific community to keep an open mind about whether the pandemic stemmed from a laboratory, a *controversial hypothesis*, suggesting that SARS-CoV-2 was either created in a lab or was accidentally or intentionally released by researchers at the Wuhan Institute of Virology. Clearly the research community took such a hypothesis seriously, and investigated it in depth, and as we shall see the controversial lab hypothesis has been completely debunked [98, 99].

Hence, much more evidence has come to light that supports a *zoonotic* origin, similar to that of HIV, Zika virus, Ebola virus and multiple influenza viruses. With all the evidence we presently have it is clear to most researchers that the pandemic started at the market in Wuhan. *Separate lines of analysis point to it, and it's extremely improbable that two distinct lineages of SARS-CoV-2 could have been derived from a laboratory and then coincidentally ended up at the market,* Michael Worobey concludes [97].

Nonetheless, others said they were not completely convinced of two spillover events because the virus might have evolved from one lineage into the other in a person who was immunocompromised. More data collected from people and animals is needed to answer this question, and to show that the first spillover occurred at the Huanan market. David Relman, a microbiologist at Stanford University in California, said that the preprints of the studies were not definitive, and that they excluded the possibility that people were infected prior to the outbreak at the market, but went undiagnosed [100].

During the twenty-first century the WHO has alerted the World community about potential pandemics at a handful of occasions before COVID-19, with relatively (compared to COVID-19) minor reactions. It may be as if there are more epidemic threats in the modern global World. The reasons are likely a mix of the evolutionary mechanism per se, overpopulation, climate change, *anthropomorphisation* (in a wider definition) of nature, *homo sapiens sapiens* living in new areas closer to wild animals, etc. Parallels can be drawn to the emergence of HIV, retroviruses, zoonotic mechanisms and viral evolution, as mentioned.

In October 2021 WHO named the researchers to reboot outbreak origin investigations. The group, called SAGO, would create a permanent framework for probing epidemics, and initiate a next phase of the COVID origins hunt [100]. In February 2022 two massive twin-studies were published, which concluded that *SARS-CoV-2 emergence very likely resulted from at least two zoonotic events* [99] and that the *Huanan market was the epicenter of SARS-CoV-2 emergence* [101].

The live animal markets is an omnipresent theme in virus spillover events [102, 103], with markets such as the Huanan market selling live mammals being in the highest risk category [104]. The events leading up to the COVID-19 pandemic mirror the SARS-CoV-1 outbreaks from 2002 to 2004 [105].

At the end of July 2022 an analysis in *Science* indicated that the emergence of SARS-CoV-2 occurred via the live wildlife trade in China and showed that the Huanan market was the epicenter of the COVID-19 pandemic—the study could pinpoint the precise animal stalls, with environmental samples obtained from known live animal vendors, in the southwest corner of the westmost wing of the market complex, presented with detailed maps in the article [106].

The same day and in the same scientific journal another study showed that SARS-CoV-2 genomic diversity before February 2020 likely comprised only two distinct viral lineages, denoted A and B. Phylodynamic rooting methods, coupled with epidemic simulations, revealed that these lineages were the result of at least two separate cross-species transmission events into humans. As with other coronaviruses, SARS-CoV-2 emergence likely resulted from multiple zoonotic events [107].

Continuous Fight for Science and Reason

Until this present day, there is a constantly ongoing battle for Enlightenment, including on the highest international level. Speaking at the opening plenary of the Science Summit of the UN General Assembly, *Magdalena Skipper*, Editor in Chief of the world leading scientific journal *Nature*, called on the United Nations to recognise more explicitly the *role of science* and the scientific community in solving the most pressing global challenges facing the world [108].

Opening against the backdrop of COVID-19 and the accelerating climate emergency, the UN's 76th General Assembly in New York marks a pivotal point to call together government and civil society to collaborate and take action. This call for collaboration must be explicitly inclusive of the scientific community. Too often policy and decision makers are only exposed to narrow sections of research evidence or constrained by political factors, and, as Skipper asserts, to truly achieve global goals, science need to be at its heart. At the Science Summit, Magdalena Skipper, Editor in Chief of *Nature*, said:

> *The UN has recognised the role of science in the path towards achieving SDGs; the second UN Global Sustainable Development Report has been commissioned by the Secretary General. But it is time for the UN to take the next step and place the science agenda front and centre on the UN General Assembly agenda itself.*
>
> *The UN sees its General Assembly as the means to bring together world leaders, civil society champions, young people and global businesses to reinforce and reinvigorate our collective determination to solve our shared problems. The grave omission in this vision is the absence of an explicit mention of science. It's an omission that is likely to be costly. The world cannot afford it; it's time to formally and explicitly extend the invitation* [108].

Hence, all the more important that AI in healthcare rests in the hands of benevolent, well-regulated and enlightened societies and people, given the incredible sharpness of this weapon.

Modern Tools for Pandemic Control

Previous pandemics lacked not only vaccinations, but also means to warn the public, no ways to track or trace the disease spread, no effective treatments (often neither curative nor palliative) when people were hospitalised, non-existent disease

monitoring, and of course before Nobel laureate Wilhelm Röntgen no proper imaging at all. Nowadays we have better tools for all of the these, but we live in a globalised World, with more international mobility than ever before, making disease spread to great areas easier than ever. To this we can add climate change and overpopulation.

Several of the chapters in this book will dive deeper into vaccine development, tracing, testing, tracking and monitoring of the present pandemic. It is also worth noticing the simultaneous rise of *visual and data journalism*, which has shown the potential to inform the public more in-depth than ever before.

Among the most important medical innovations we find Edward Jenner's *vaccination* of 1796, Horace Wells' *anaesthesia* of 1844 and Alexander Fleming's *penicillin* of 1928. The impact of vaccination cannot be praised and elaborated enough. Overall, smallpox infected about 300 million people in the twentieth century. Through vaccination, this has become the only human disease successfully eradicated in global health history, with the WHO making it official in 1980. And since 1977, no cases of smallpox have been reported again.

But there is a long range of diseases, which the general public have almost forgotten, or have vague ideas about, all thanks to vaccinations, for example; Polio, Tetanus, Hepatitis A and B, Rubella, Haemophilus Influenzae, Measles, Whooping Cough (Pertussis), Pneumococcal Disease, Rotavirus, Mumps, Chickenpox, Diphtheria and to some extent even Influenza. The tremendous and laudable work by international organisations such as the WHO, show the very best sides of human cooperation and progress.

A Brief Chronology of the Chapters of This Book

Below we present and explain the scope and content of the book, with a chronology of the 12 following chapters and their contents, and embedded links to these 12 (Chaps. 2–13). AU2

- The first chapter, *AI and Pooling Tests for COVID-19* (Chap. 2), presents AI tools for pooled testing and recovery, and delivers a broad description of the use of algorithmic tools for pooled testing in the context of COVID-19 has emerged as a key tool in improving the prediction accuracy as well as efficiency of pooled testing. Well-organized testing of individuals for COVID-19, with reverse transcription polymerase chain reaction (RT-PCR) is a pivotal ingredient in the fight against the pandemic.

  Testing resources, i.e., testing kits, key reagents, trained personnel, useful clinic spaces and time slots, are indeed widely limited resources globally. Hence, pooled testing is a method increase the velocity—pooling implicates *mixing* small portions of *samples* from different individuals. The pools are then tested, instead of the *individual* samples. A limited number of pools, in comparison to the total pool samples, suffice, deliver precise *prediction* of the health status of

the entire pool, assuming only a small number of the samples were positive using appropriate AI tools to analyze the pools.

- The second chapter *AI for Drug Repurposing in the Pandemic Response* (Chap. 3), deals with the discovery of treatments for the present and future pandemics. It shows how drugs can be repurposed, which is an effective strategy in the first line of action, when the applicable *de novo* pharmaceuticals, both vaccines and anti-viral therapies, are still in the development phase. In a wider perspective, AI has generally boosted pharmaceutical development. The chapter exhibits and discusses a variety of AI solutions, platforms and applications in use to target COVID-19. It also presents various strategies and global collaborations for the optimal use of AI in the drug development against the present pandemic, and in order stay alert for future ones.
- The third chapter *AI and Point Of Care Image Analysis For COVID-19* (Chap. 4) teaches about the application of AI techniques in medical imaging and discusses a literature review on diagnostic and prognostic models using AI tools. In the clinical management of the COVID-19 pandemic, the point-of-care imaging, i.e., chest x-ray, computed tomography and ultrasound, has played a pivotal role. This chapter explores the role of AI in point-of-care imaging analysis throughout the pandemic.
- The fourth chapter *Machine Learning and Laboratory Values in the Diagnosis, Prognosis and Vaccination Strategy of COVID-1* (Chap. 5) unpins the fact that an early diagnosis is pivotal for the correct treatment of infectious diseases. Both prognosis and the control of disease spread of an epidemic are affected by this—an additional layer of complexity if the patients are frequently asymptomatic.
  As in Chap. 1, the gold-standard for diagnosing SARS-CoV-2 infection is the identification, with the use of molecular techniques, of viral genomic material (RNA), and it is demonstrated how AI can help augment testing power, make testing more available, quick and precise. It is also shown how AI can define the diagnoses per se and also prognosticate patients. AI models based on laboratory tests and biometric facts lead to improved outcome. The chapter also analyses the vaccination campaign against SARS-CoV-2, including heterogeneous patient selection, laboratory parameters and AI tools for, e.g., validation, and for strategy optimisation and planning.
- The fifth chapter *AI and the Infectious Medicine of COVID-19* (Chap. 6) concludes that the rise of AI for infectious medicines coincides with the global need for the same during the pandemic. There has been a long gone digitalisation of infectious medicine, and the pandemic boosted the adoption of AI into the field. The chapter reviews COVID-19 seen from a clinical infectious medicine specialists' perspective, with the patient management, the use of AI in clinical research, decision support system in infectious medicine, as well as diagnostics, care and preventative measures.
- The sixth chapter *AI and ICU Monitoring in COVID-19* (Chap. 7) shows that a new paradigm rises on the intensive care unit horizon, with the help of AI and data dense care. Modern ICUs depend profoundly on data from patient monitoring devices, which opens up for complex data from a plethora of electronic

sources, to boost evidence-based medicine (EBM), based on the use of AI approaches. The chapter presents how AI could help in the monitoring of patients in the intensive care units during the COVID-19 pandemic, which flooded the ICUs, and tested the limits of hospital performance.

- The seventh chapter *Symptom Based Detection Models of COVID-19 Infection Using AI* (Chap. 8) demonstrates the pivotal importance of early diagnosis, and how this can be accomplished with the use of AI, based on characteristic symptoms. It shows how to avoid problems with confusion of these with other infectious diseases, i.e., clinical differentiation of COVID-19 patients, with the help of AI, underpinning evidence based medicine and the most accurate decision making. Further it is presented how patient demographics and health data build up a useful clinical AI based detection model.
- The eighth chapter *AI Techniques for Forecasting Epidemic Dynamics: Theory and Practice* (Chap. 9) deals with a research area, which has been flourishing since at least the 1990s. It greatly impacts decision makers, the general public, builds up the foundation of new policies, and other researchers, all appreciating accurate forecasts on time. A large pandemic brings the epidemic dynamics into a stress test, which encompasses many complex variables, including human behaviour, viral factors, transmission patterns, legal regulations, and more. The COVID-19 pandemic was no exception, and to predict its evolution has been a complex exertion, with average to limited outcome. This chapter presents the *state-of-the-art of epidemic forecasting*, and focuses the task of forecasting during an ongoing pandemic, and gives a review of a handful of methods and their challenges.
- The ninth chapter *Regulatory Aspects on AI and Pharmacovigilance for COVID-19* (Chap. 10) investigates the present and upcoming legislative and regulatory situations, and presents the rules, which govern AI in the health sector at European Union level. The chapter focuses on pharmacovigilance, and the use of AI in the pharmaceutical industry in drug and medical product development, controlling accurate quality, efficacy and safety. It also describes the European strategy concentrating on the use of AI and how a better health policy can be created. Ethical and personal data protection aspects are also elaborated.
- The tenth chapter *AI and the Clinical Immunology/Immunoinformatics for COVID-19* (Chap. 11) presents applications of AI in protein analysis, structure predictions, biochemical simulations, molecular functions predictions in vaccine design, biomedical image processing, and genome analysis in SARS-CoV-2 genomic datasets. The authors shows how AI in immunology contribute to the fight against COVID-19, with an ever increasing number of AI-applications, with increased scalability and wide-ranging complex pattern recognition in vast amounts of data. AI can be particularly helpful in the complex immunological fields, with the help of accelerated processing power, and strategies to outperform traditional methods in multiple areas.
- The eleventh chapter *AI and Dynamic Prediction of Deterioration in Covid-19* (Chap. 12) shows how changes in manual measurements of vital parameters are associated with predictability of sepsis and mortality in COVID-19 patients. The

chapter focuses on the dynamic changes in vital signs associated with sepsis and mortality, i.e., deterioration, in COVID-19, covering both existing, and elaborates other applicable, approaches in the future.
- The twelfth chapter *AI, Epidemiology and Public Health in the Covid Pandemic* (Chap. 13) fuses two major and tightly related research areas. Both epidemiology and public health depend on quantifiable and measurable entities. These quanta describe and measure a health issue in society, and regardless of whether they depict disease, seek to explain the aetiology or underpin the policy making, they form the pivotal scientific bases in both twin disciplines.

New facts and figures can now be collected in epidemiology, with the rise of digitalisation, encompassing hitherto aspects beyond scientific reach. AI now helps processing these data, opening up new public health perspectives. COVID-19 has provided a grand scale testing field for these new explorations with AI for epidemiology and public health. This chapter examines how which types of AI in these two fields could practically be used in a pandemic, and in relation to the modern concept *infodemics*, and also evaluates the algorithm performances and maturities.

Power of Science

On the architrave above the entrance of the grand aula in Uppsala University's, alma mater for one of the editors, main building from 1887 there is a gilded quotation by *Thomas Thorild* (1759–1808) [109]:

Tänka fritt är stort,	Thinking freely is great,
men tänka rätt är större.	but thinking rightly is greater [110].

Thomas Thorild, a Swedish poet, critic, feminist and philosopher, noted for his early support of women's rights, did not refer to what would be morally correct, but rather logically [111]. Though, at the time of the university main building's construction, the leading professors intended to use the quotation in the moral sense, to curb the culturally radical tendencies they feared, and judged to be a sign of moral decay [112]. We use it to underline the all-important common education on science and reason, to facilitate human progress (Fig. 1.2).

Gustav Mahler composed his fifth symphony in 1901 and 1902, mostly during the summer months at Mahler's holiday cottage at Maiernigg. Among its most distinctive features is the trumpet solo that opens the work with a rhythmic motif similar to the opening of Ludwig van Beethoven's Symphony No. 5. The symphony starts with the trumpet's soloistic *Trauermarsch* (funeral march) theme, which lonely and sadly rises and "collapses" several times. The Boston Philharmonic Orchestra's chief conductor *Benjamin Zander* described it, in one of his trumpet master classes, referring to Mahler's instructions to the conductor, which are not part of the scores; *it is a military fanfare…a solitary bugler left on the battle field with four thousand dead soldiers…* [113].

Fig. 1.2 The entrance of the aula in the main building of Uppsala University. Above the entrance is the quotation by Thomas Thorild. (Photo: Dr Niklas Lidströmer)

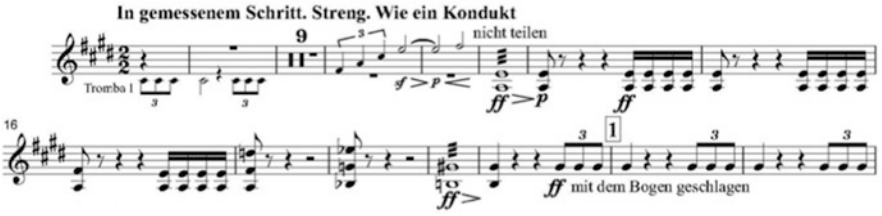


Fig. 1.3 The initial bars of Gustav Mahler's 5th symphony, part 1: Trauermarsch (Funeral march). *In gemessenem Schritt. Streng. Wie ein Kondukt* (At a measured pace. Strict. Like a funeral procession.) In C♯ minor. (Here, rendered and adapted from Mahler's original partiture by Dr. Niklas Lidströmer)

After several years of social distancing, isolation for many already lonely individuals, numerous deaths from this disease, the fifth deadliest pandemic in history, it is an atmosphere like the one in Mahler's fifth starting theme, that fits the musical sound carpet at the end of this introduction (Fig. 1.3).

Fig. 1.4 Detail from the book frontispiece of the *Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers* (1772), drawn by Charles-Nicolas Cochin (1715–1790) and engraved by Benoît-Louis Prévost (1735 or 1747–1804). The work is heavily laden with symbolism: The figure in the centre represents *truth*, which is surrounded by a bright light—the central symbol of the *Enlightenment*. The two other figures, on the right, *reason* and *philosophy*, are tearing the veil from *truth*. *Copyright*: This work is in the public domain in its country of origin, France, and other countries and areas where the copyright term is the author's life plus 70 years or fewer

It is though with an almost passionate *nebenbei* [114] sentiment, as expressed in a Mendelssohnian violin cadenza, that we wished to compile this book on the essential use of artificial intelligence in the fight against the present and future pandemics.

The use of AI in the fight against COVID-19 and future pandemics is presented here, with the intention of benevolently wishing to repair the world—to quote a Jewish concept: *tikkun olam* (תיקון עולם, Hebrew: *repairing of the world*).

It is with this background and under these circumstances we wish this book to be a beacon of light in the fight against this and future pandemics. To show how science can be our main power in fighting future disasters on many possible fronts, and how within science, the integration of different domains, from the computational to the biological, is essential for progress and results (Fig. 1.4).

References

1. An American dictionary of the English language. 1st ed. Noah Webster; 1828.
2. Morens DM, Folkers GK, Fauci AS. What is a pandemic? J Infect Dis. 2009;200(7):1018–21. https://doi.org/10.1086/644537.

3. Amarna letters EA 244, (Amarna Tablet 244). Inventory Number: C 4768 (12200) Tablet Number: EA 244 Knudtzon 1907–1915).
4. Papagrigorakis MJ, Yapijakis C, Synodinos PN, Baziotopoulou-Valavani E. DNA examination of ancient dental pulp incriminates typhoid fever as a probable cause of the Plague of Athens. Int J Infect Dis. 2007;10(3):206–14. https://doi.org/10.1016/j.ijid.2005.09.001.
5. Olson PE, Hames CS, Benenson AS, Genovese EN. The Thucydides syndrome: Ebola déjà vu? (or Ebola reemergent?). Emerg Infect Dis. 1996;2(2):155–6. https://doi.org/10.3201/eid0202.960220.
6. Cooper GM. The origin and evolution of cells. The cell: a molecular approach. 2nd ed; 2000. AU3
7. Schopf JW, Kudryavtsev AB, Czaja AD, Tripathi AB. Evidence of Archean life: stromatolites and microfossils. Precambrian Res. 2007;158:141–55.
8. Iyer LM, Balaji S, Koonin EV, Aravind L. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res. 2006;117(1):156–84. https://doi.org/10.1016/j.virusres.2006.01.009.
9. Schopf JW. Fossil evidence of Archaean life. Philos Trans R Soc Lond B Biol Sci. 2006;361(1470):869–85.
10. Goodman JR, Wohns N. AIM and evolutionary theory. In: Lidströmer N, et al., editors. Artificial intelligence in medicine. Cham: Springer; 2021. https://doi.org/10.1007/978-3-030-58080-3_41-1. AU4
11. Gould GM, Pyle WL. Historic epidemics. In: Anomalies and curiosities of medicine. Blacksleet River; 1896. ISBN 978-1-4499-7722-1.
12. Spreeuwenberg P, Kroneman M, Paget J. Reassessing the global mortality burden of the 1918 influenza pandemic. Am J Epidemiol. 2018;187(12):2561–7. https://doi.org/10.1093/aje/kwy191.
13. WHO Fact Sheet, Smallpox, who.int/health-topics/smallpox.
14. Lawn SD, Zumla AI (July 2011). "Tuberculosis". Lancet 378 (9785): 57–72. doi:https://doi.org/10.1016/S0140-6736(10)62173-3
15. WHO Fact Sheet. https://www.who.int/news-room/fact-sheets/detail/tuberculosis.
16. The Chambers Dictionary. New Delhi: Allied Chambers India Ltd. 1998. p. 352. ISBN 978-81-86062-25-8. Archived from the original on 6 Sept 2015.
17. Adkinson NF, Bennett JE, Douglas RG, Mandell GL. Mandell, Douglas, and Bennett's principles and practice of infectious diseases. 7th ed. Philadelphia, PA: Churchill Livingstone/Elsevier; 2010. Chapter 250. ISBN 978-0-443-06839-3.
18. Tuberculosis (TB). World Health Organization (WHO); 2018.
19. Tuberculosis. World Health Organization (WHO); 2002.
20. Tuberculosis (TB). WHO. Archived from the original on 30 Jul 2020.
21. Global tuberculosis report. World Health Organization (WHO). Archived from the original on 30 Dec 2013.
22. Kumar V, Robbins SL. Robbins basic pathology. 8th ed. Philadelphia: Elsevier; 2007. ISBN 978-1-4160-2973-1. OCLC 69672074.
23. Hays JN. Epidemics and pandemics: their impacts on human history. ABC-CLIO; 2005. p. 421. ISBN 9781851096589. AU5
24. Ahiarah L. Cholera. 2008. www.austincc.edu.
25. Cholera's seven pandemics. www.cbc.ca.
26. Lam C, Octavia S, Reeves P, Wang L, Lan R. Evolution of seventh cholera pandemic and origin of 1991 epidemic, Latin America. Emerg Infect Dis. 2010;16(7):1130–2. https://doi.org/10.3201/eid1607.100131. PMC 3321917.
27. Rogers K. 1986 flu pandemic. Encyclopaedia Britannica. Open Publishing; 2010.
28. Jester BJ, Uyeki TM, Jernigan DB. Fifty years of influenza A(H3N2) following the pandemic of 1968. Am J Public Health. 2020;110(5):669–76. https://doi.org/10.2105/AJPH.2019.305557. ISSN 0090-0036. PMC 7144439. AU6
29. HIV/AIDS Factsheet. World Health Organization. AU7
30. HIV/AIDS overview, symptoms and treatment. World Health Organization.
31. UN AIDS DATA2019. UNAIDS.org. AU8
32. Foreword—AIDS 2020. UNAIDS.

33. Fact Sheet—latest global and regional data on the status of the HIV/AIDS global epidemic (PDF). UNAIDS.
34. Ryan JR, editor. Chapter 1—past pandemics and their outcome. In: Pandemic influenza: emergency planning and community preparedness. CRC Press; 2008. p. 16. ISBN 978-1-42006088-1. The Asiatic Flu of 1889-1890 was first reported in Bukhara, Russia.
35. Garmaroudi, Farshid S. (2007). The last great uncontrolled plague of mankind. Sci Creat Quart. Accessed 25 Mar 2020. The Asiatic flu, 1889-1890: It was the last great pandemic of the nineteenth century.
36. Shally-Jensen M, editor. Influenza. Encyclopedia of contemporary American social issues, vol. 2. ABC-CLIO; 2010. p. 1510. ISBN 978-0-31339205-4.
37. King A. An uncommon cold. New Sci. 2020;246(3280):32–5. Bibcode: 2020NewSc.246...32K. https://doi.org/10.1016/S0262-4079(20)30862-9.
38. Lee P. Molecular epidemiology of human coronavirus OC43 in Hong Kong (Thesis). The University of Hong Kong Libraries. https://doi.org/10.5353/th_b4501128. hdl:10722/131538.
39. Lim YX, Ng YL, Tam JP, Liu DX. Human coronaviruses: a review of virus–host interactions. Diseases. 2016;4(3):26. https://doi.org/10.3390/diseases4030026.
40. Li F. Structure, function, and evolution of coronavirus spike proteins. Annu Rev Virol. 2016;3(1):237–61. https://doi.org/10.1146/annurev-virology-110615-042301. PMC 5457962. PMID 27578435. BCoV S1-NTD does not recognize galactose as galectins do. Instead, it recognizes 5-N-acetyl-9-O-acetylneuraminic acid (Neu5,9Ac2) (30, 43). The same sugar receptor is also recognized by human coronavirus OC43 (43, 99). OC43 and BCoV are closely related genetically, and OC43 might have resulted from zoonotic spillover of BCoV (100, 101).
41. Lau SKP, Lee P, Tsang AKL, Yip CCY, Tse H, Lee RA, So L-Y, Lau Y-L, Chan K-H, Woo PCY, Yuen K-Y. Molecular epidemiology of human coronavirus OC43 reveals evolution of different genotypes over time and recent emergence of a novel genotype due to natural recombination. J Virol. 2011;85(21):11325–37. https://doi.org/10.1128/JVI.05512-11.
42. Brüssow H, Brüssow L. Clinical evidence that the pandemic from 1889 to 1891 commonly called the Russian flu might have been an earlier coronavirus pandemic. Microb Biotechnol. 2021;14(5):1860–70. https://doi.org/10.1111/1751-7915.
43. Taxonomy browser (Betacoronavirus 1). www.ncbi.nlm.nih.gov.
44. Dowdle WR. Influenza A virus recycling revisited. Bull World Health Organ. 1999;77(10):820–8.
45. Anhlan D, Grundmann N, Makalowski W, Ludwig S, Scholtissek C. Origin of the 1918 pandemic H1N1 influenza A virus as studied by codon usage patterns and phylogenetic analysis. RNA. 2011;17(1):64–73. https://doi.org/10.1261/rna.2395211.
46. Odun-Ayo F, Odaibo G, Olaleye D. Influenza virus A (H1 and H3) and B co-circulation among patient presenting with acute respiratory tract infection in Ibadan, Nigeria. Afr Health Sci. 2018;18(4):1134–43. https://doi.org/10.4314/ahs.v18i4.34.
47. Forni D, Cagliani R, Clerici M, Sironi M. Molecular evolution of human coronavirus genomes. Trends Microbiol. 2017;25(1):35–48. https://doi.org/10.1016/j.tim.2016.09.001.
48. Vijgen L, Keyaerts E, Moës E, Thoelen I, Wollants E, Lemey P, Vandamme AM, Van Ranst M. Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. J Virol. 2005;79(3):1595–604. https://doi.org/10.1128/JVI.79.3.1595-1604.2005.
49. Knudsen JK. Overraskende opdagelse: Coronavirus har tidligere lagt verden ned [Surprising discovery: Coronavirus has previously brought down the world]. DR (in Danish). 2020. Accessed 13 Aug 2020. A presumed influenza pandemic in 1889 was actually caused by coronavirus, Danish research shows.
50. Van Ranst M. Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. J Virol. 2020;79(3):1595–604. https://doi.org/10.1128/JVI.79.3.1595-1604.2005. PMC 54410.

51. Brüssow H. What we can learn from the dynamics of the 1889 'Russian flu' pandemic for the future trajectory of COVID-19. Microb Biotechnol. 2021;14(6):2244–53.
52. Hilleman, Maurice R. (2002). Realities and enigmas of human viral influenza: pathogenesis, epidemiology and control. Vaccine. 20 (25–26): 3068–3087. CiteSeerX 10.1.1.523.7697. doi: https://doi.org/10.1016/S0264-410X(02)00254-2.
53. Howard J. Plague was one of history's deadliest diseases—then we found a cure. National Geographic; 2020. Archived from the original on 2 Dec 2020. Accessed 3 Dec 2020.
54. Compare: 1918 Spanish influenza pandemic versus COVID-19. BioSpace.
55. Stathakopoulos D. Plague, Justinianic (early medieval pandemic). The Oxford Dictionary of Late Antiquity, Oxford University Press; 2018. https://doi.org/10.1093/acref/9780198662778.001.0001. ISBN 978-0-19-866277-8
56. Adam D. COVID's true death toll: much higher than official records. Modelling suggests that by the end of 2021, some 18 million people had died because of the pandemic. Nature. 2022;603:562. https://doi.org/10.1038/d41586-022-00708-0. AU12
57. Stenseth NC. Plague through history. Science. 2008;321(5890):773–4. https://doi.org/10.1126/science.1161496. ISSN 0036-8075.
58. Prem H. Disease outbreaks in Central Mexico during the sixteenth century. In: Cook ND, Lovell WG, editors. "Secret judgments of god": old world disease in colonial Spanish America. Norman: University of Oklahoma Press; 1991. p. 20–48. ISBN 0806123729.
59. Antonine Plague. World History Encyclopedia.
60. Acuna-Soto R, Stahle DW, Cleaveland MK, Therrell MD. Megadrought and megadeath in 16th century Mexico. Emerg Infect Dis. 2002;8(4):360–2. https://doi.org/10.3201/eid0804.010175.
61. Goodall EW. Typhus fever in Poland, 1916 to 1919. Sect Epidemiol State Med. 1920;13:261–76. https://doi.org/10.1177/003591572001301507.
62. Viboud C, Simonsen L, Fuentes R, Flores J, Miller MA, Chowell G. Global mortality impact of the 1957–1959 influenza pandemic. J Infect Dis. 2016;213(5):738–45. https://doi.org/10.1093/infdis/jiv534.
63. Paul WE. Fundamental immunology; 2008. p. 1273. ISBN 9780781765190.
64. Skaarup BO. Anatomy and anatomists in early modern Spain. 1st ed. Routledge; 2015. p. 205. ISBN 978-1472448262.
65. Suzuki A. Smallpox and the epidemiological heritage of modern Japan: towards a total history. Med Hist. 2011;55(3):313–8. https://doi.org/10.1017/S0025727300005329.
66. Hashemi Shahraki A, Carniel E, Mostafavi E. Plague in Iran: its history and current status. Epidemiol Health. 2016;38:e2016033. https://doi.org/10.4178/epih.e2016033.
67. Scasciamacchia S, Serrecchia L, Giangrossi L, Garofolo G, Balestrucci A, Sammartino G, Fasanella A. Plague epidemic in the Kingdom of Naples, 1656–1658. Emerg Infect Dis. 2012;18(1):186–8. https://doi.org/10.3201/eid1801.110597.
68. Frerichs RR. Asiatic cholera pandemics during the life of John Snow: Asiatic cholera pandemic of 1846-63. John Snow—a historical giant in epidemiology. UCLA Department of Epidemiology—Fielding School of Public Health.
69. Rosenwald MS. History's deadliest pandemics, from ancient Rome to modern America. Washington Post; 2020.
70. Tracking covid-19 excess deaths across countries. The Economist.
71. What 'excess deaths' do and don't tell us about COVID-19. Reason. 29 Apr 2020.
72. Calvignac-Spencer S, Düx A, Gogarten JF, Patrono LV. Molecular archeology of human viruses. Adv Virus Res. 2021;111:31–61. https://doi.org/10.1016/bs.aivir.2021.07.002.
73. Piret J, Boivin G. Pandemics throughout history. Front Microbiol. 2021;11:631736.
74. Calvignac-Spencer S, Düx A, Gogarten JF, Patrono LV. Chapter two—molecular archeology of human viruses. In: Kielian M, Mettenleiter TC, Roossinck MJ, editors. Advances in virus research, vol. 111. Academic Press; 2021. p. 31–61., ISSN 0065-3527, ISBN 9780128246054. https://doi.org/10.1016/bs.aivir.2021.07.002.

75. Pomeau RH. Voltaire. Encyclopedia Britannica. 17 Nov 2021. https://www.britannica.com/biography/Voltaire. Accessed 28 Mar 2022.
76. Roberts RM. Serendipity: accidental discoveries in science. New York: Wiley; 1989. Chapter 4.
77. Letters concerning the English nation. London; 1733.
78. Voltaire, Letter on Inoculation with Smallpox, Letter XI, *Letters on England* (also known as *Philosophical Letters*), as translated by Leonard Tancock for the Penguin edition.
79. Transcript of a letter from George Washington to Doctor William Shippen, Jr. (a draft in the handwriting of Washington's aid, Robert Hanson) In George Washington Papers at the Library of Congress, 1741–1799: Series 4. General Correspondence. 1697–1777.
80. From George Washington to William Shippen, Jr., 6 February 1777. Founders Online, National Archives. https://founders.archives.gov/documents/Washington/03-08-02-0281. [Original source: *The Papers of George Washington*, Revolutionary War Series, vol. 8, *6 January 1777–27 March 1777*, ed. Frank E. Grizzard, Jr. Charlottesville: University Press of Virginia; 1998, p. 264.
81. Edsall G. Smallpox vaccination vs inoculation. JAMA. 1964;190(7):689–90.
82. Romer D, Jamieson KH. Conspiracy theories as barriers to controlling the spread of COVID-19 in the U.S. Soc Sci Med. 2020;263:113356.
83. Larsson P. COVID-19 anti-vaxxers use the same arguments from 135 years ago. The Conversation. 2020;
84. Marsh J. The 1885 Montreal smallpox epidemic. In: The Canadian encyclopedia. 2021. https://www.thecanadianencyclopedia.ca/en/article/plague-the-red-death-strikes-montreal-feature.
85. Jennings W. Lack of trust, conspiracy beliefs, and social media use predict COVID-19 vaccine hesitancy. Vaccines. 2021;9(6):593. https://doi.org/10.3390/vaccines9060593.
86. Kant I. Answering the question: what is enlightenment? Original in German: *Beantwortung der Frage: Was ist Aufklärung*. Beginn des Traktats. (EA 1784). (Immanuel Kants vermischte Schriften; Hrsg.: Johann Heinrich Tieftrunk; Bd. 2. Halle 1799). In the December 1784 publication of the Berlinische Monatsschrift (Berlin Monthly), edited by Friedrich Gedike and Johann Erich Biester, Kant replied to the question posed a year earlier by the Reverend Johann Friedrich Zöllner, who was also an official in the Prussian government.
87. Borger J. Caught in a superpower struggle: the inside story of the WHO's response to coronavirus. The Guardian. 2020;
88. Arnold C. The non-human living inside of you. Nautilus Biol + Beyond. 2020;
89. Hu B, Guo H, Zhou P, et al. Characteristics of SARS-CoV-2 and COVID-19. Nat Rev Microbiol. 2021;19:141–54. https://doi.org/10.1038/s41579-020-00459-7.
90. Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. Cold Spring Harb Perspect Med. 2011;1(1):a006841. https://doi.org/10.1101/cshperspect.a006841. PMID: 22229120; PMCID: PMC3234451.
91. Mena I, Nelson MI, Quezada-Monroy F, Dutta J, Cortes-Fernández R, Lara-Puente JH, Castro-Peralta F, Cunha LF, Trovão NS, Lozano-Dubernard B, Rambaut A, van Bakel H, García-Sastre A. Origins of the 2009 H1N1 influenza pandemic in swine in Mexico. eLife. 2016;5:e16777. https://doi.org/10.7554/eLife.16777.
92. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020;382(8):727–33. https://doi.org/10.1056/NEJMoa2001017.
93. Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19) (PDF) (Report). World Health Organization (WHO).
94. Rathore JS, Ghosh C. Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), a newly emerged pathogen: an overview. Pathog Dis. 2020;78(6):ftaa042. https://doi.org/10.1093/femspd/ftaa042.
95. Thomas S. The structure of the membrane protein of SARS-CoV-2 resembles the sugar transporter SemiSWEET. Pathog Immun. 2020;5(1):342–63. https://doi.org/10.20411/pai.v5i1.377.

96. Fact check: The Wuhan Institute of Virology is not owned by GlaxoSmithKline. Reuters. 2020.
97. Maxmen A. Wuhan market was epicentre of pandemic's start, studies suggest. Nature. 2022;603:15–6. https://doi.org/10.1038/d41586-022-00584-8. AU19
98. Worobey M, Levy JI, Malpica LM, Serrano AC-C, Pekar JE, Goldstein SA, Rasmussen AL, Kraemer MUG, Newman C, Koopmans MPG, Suchard MA, Wertheim JO, Lemey P, Robertson DL, Garry RF, Holmes EC, Rambaut A, Andersen KG. The Huanan market was the epicenter of SARS-CoV-2 emergence. Zenodo. 2022; https://doi.org/10.5281/zenodo.6299600. AU20
99. Pekar JE, Magee A, Parker E, Moshiri N, Izhikevich K, Havens JL, Gangavarapu K, Serrano M, Lorena M, Crits-Christoph A, Matteson NL, Zeller M, Levy JI, Wang JC, Hughes S, Lee J, Park H, Park M-S, Yan CZ, Katherine TP, Lin R, et al. SARS-CoV-2 emergence very likely resulted from at least two zoonotic events. Zenodo. 2022; https://doi.org/10.5281/zenodo.6291628. AU21
100. Maxmen A. WHO names researchers to reboot outbreak origin investigations. Nature. 2021; https://doi.org/10.1038/d41586-021-02813-y. AU22
101. Worobey M, Levy JI, Malpica LM, Serrano AC-C, Pekar JE, Goldstein SA, Rasmussen AL, Kraemer MUG, Newman C, Koopmans MPG, Suchard MA, Wertheim JO, Lemey P, Robertson DL, Garry RF, Holmes EC, Rambaut A, Andersen KG. The Huanan market was the epicenter of SARS-CoV-2 emergence. Zenodo. 2022; https://doi.org/10.5281/zenodo.6299116.
102. Karesh WB, Cook RA, Bennett EL, Newcomb J. Wildlife trade and global disease emergence. Emerg Infect Dis. 2005;11:1000–2.
103. Johnson CK, Hitchens PL, Pandit PS, Rushmore J, Evans TS, Young CCW, Doyle MM. Global shifts in mammalian population trends reveal key predictors of virus spillover risk. Proc Biol Sci. 2020;287:20192736.
104. Lin B, Dietrich ML, Senior RA, Wilcove DS. A better classification of wet markets is key to safeguarding human health and biodiversity. Lancet Planet Health. 2021;5:e386–94.
105. Shi Z, Hu Z. A review of studies on animal reservoirs of the SARS coronavirus. Virus Res. 2008;133:74–87.
106. Worobey M, et al. The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. Science. 2022; https://doi.org/10.1126/science.abp8715.
107. Pekar JE, et al. The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2. Science. 2022; https://doi.org/10.1126/science.abp8337.
108. UN must explicitly recognise role of science in combating global challenges, according to Nature Editor in Chief, *Press Release*, Springer Nature Group. 14 Sept 2021. https://group.springernature.com/gp/group/media/press-releases/un-must-explicitly-recognise-role-of-science/19660842. AU23
109. Lindqvist H. Historien om Sverige. Gustavs dagar (The history of Sweden. The days of Gustav III).
110. Free translation by the editor Dr. Niklas Lidströmer.
111. Mortensen BME. Thorild, Thomas. In: Steinberg SH, editor. Cassell's encyclopedia of world literature. London: Cassell; 1953.
112. Frängsmyr T. Svensk idehistoria: bildning och vetenskap under tusen år. Natur och kultur. 1938–(2000). sid. 143–144. ISBN 91-27-07793-4.
113. Boston Philharmonic, Benjamin Zander, Master Class. (Mahler: Trumpet Solo from Symphony No. 5 (Benjamin Zander)—Interpretation Class—Elmer Churampi). Quotation taken from YouTube video recording at 8'16-8'24, also available at benjaminzander.org.
114. Boston Philharmonic, Benjamin Zander. Mendelssohn: Violin Concerto—1st movement—Interpretation Class, Term taken from YouTube video recording at 15'39-16'48, also available at benjaminzander.org. The full quotation with the term; "…there is a wonderful German phrase *nebenbei*…to the side…". Benjamin Zander goes on and explains the term; "there are four German (violin) concertos – Beethoven, Brahms, Bruch…and Mendelssohn"…He points out that Mendelssohn should be played with a *nebenbei* sentiment, while the other three exemplified can be played with a more "strict" or "upright" posture.

Author Queries

Chapter No.: 1 0005460491

Queries	Details Required	Author's Response
AU1	Missing citation for "Ïmages 1.1–1.4" have been inserted in the appropriate places. Kindly check and confirm.	
AU2	Please check and confirm if the chapter numbers (first chapter, second chapter, ...) are presented correctly. Kindly correct if necessary.	
AU3	References 6 and 8 were identical and Reference 8 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	
AU4	References 11 and 12 were identical and Reference 12 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	
AU5	References 25 and 77 were identical and Reference 77 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	
AU6	References 30 and 59 were identical and Reference 59 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	
AU7	References 31 and 63 were identical and Reference 63 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	
AU8	References 33 and 34 were identical and Reference 34 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	
AU9	References 40 and 55 were identical and Reference 55 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	
AU10	References 51 and 53 were identical and Reference 53 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	
AU11	References 54 and 56 were identical and Reference 56 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	
AU12	References 64 and 79 were identical and Reference 79 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	
AU13	References 85 and 87 were identical and Reference 87 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	
AU14	References 94 and 96 were identical and Reference 96 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	
AU15	References 94 and 97 were identical and Reference 97 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	

AU16	References 99 and 100 were identical and Reference 100 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	
AU17	References 99 and 101 were identical and Reference 101 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	
AU18	References 99 and 102 were identical and Reference 102 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	
AU19	References 113 and 116 were identical and Reference 116 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	
AU20	Please clarify that Ref. [98] can be duplicate of ref. [101]?	
AU21	References 115 and 119 were identical and Reference 119 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	
AU22	References 117 and 118 were identical and Reference 118 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	
AU23	References 127 and 128 were identical and Reference 128 has been deleted. The subsequent references have been renumbered. Please check and confirm if appropriate.	

Chapter 2
AI for Pooled Testing of COVID-19 Samples

Ajit Rajwade, Nir Shlezinger, and Yonina C. Eldar

Introduction

The Coronavirus disease 2019 (COVID-19) pandemic has already forced lockdowns all over the globe, and has claimed more than five million lives worldwide. In order to handle and contain pandemics, and particularly COVID-19, large portions of the population should be frequently tested [1]. One of the main difficulties in doing so stems from the limited testing resources and the lengthy duration required to identify the presence of an infection [2].

The main diagnosis tool for COVID-19 tests is based on ribonucleic acid (RNA) extraction via qualitative reverse transcription polymerase chain reaction (RT-PCR). Although various technological alternatives have been proposed [3, 4], the RT-PCR process remains the leading and most widely spread method for COVID-19 testing. The output of this procedure represents an estimate of the viral load in the tested sample [5]. A major bottleneck associated with this form of COVID-19 testing follows from the fact that each RT-PCR machine can simultaneously process a fixed number of samples, and its procedure tends to be quite lengthy, on the order of a few hours for each test.

A leading method to tackle the time-consuming nature of RT-PCR, whose duration often lasts between 3 and 4 h, and the limited capacity associated with the detection scheme, is to increase the efficiency of COVID-19 tests using *pooling*

A. Rajwade (✉)
IIT Bombay, Mumbai, India

N. Shlezinger
Ben-Gurion University, Be'er Sheva, Israel
e-mail: nirshl@bgu.ac.il

Y. C. Eldar
Weizmann Institute of Science, Rehovot, Israel
e-mail: yonina.eldar@weizmann.ac.il

N. Lidströmer, Y. C. Eldar (eds.), *Artificial Intelligence in Covid-19*,
https://doi.org/10.1007/978-3-031-08506-2_2

techniques [6, 7]. Pooling builds on the fact that empirical observations suggest a low prevalence rate, namely, most tests are wasteful as most individual samples will return as negative. Consequently, for large populations, testing grouped samples can facilitate the aim of providing quick yet accurate results economically at an increased throughput. In pooling, each sample processed, i.e., each input to the RT-PCR machine, is comprised of a mixture of small portions of several samples taken from different patients (one sample per patient). When the infected patients constitute a small subset of the overall tested individuals, pooling-based schemes allow accurate recovery using a number of tests which is notably smaller than the number of patients [8]. While pooling can significantly increase the throughput of testing and thus boost rapid and accessible diagnosis for tracking of the pandemic, the fact that it relies on the examination of mixed samples implies that dedicated artificial intelligence (AI) methods are required for the automated and rapid identification of the infected subjects and the recovery of their viral loads.

In this chapter we review recovery methods for pooled COVID-19 tests. These AI methods can be divided according to two main mathematical frameworks for such recovery procedures: group testing (GT) and compressed sensing (CS). To describe methods belonging to these families, we begin by presenting a mathematical model which formulates the problem of pooled testing recovery in section "System Model". We elaborate on the subtleties and the specific assumptions and requirements which characterize COVID-19 testing. Based on the modelling of the pooled recovery setting, we review the application of GT, which was originally designed for testing infections in large populations, for COVID-19 pooled testing in section "Group Testing Methods for COVID-19". Our review examines different GT schemes along with their relevance in the context of COVID-19. We proceed to discuss recovery methods which arise from CS theory in section "Compressed Sensing for Pooled Testing for COVID-19". The framework of CS deals with the recovery of sparse signals [9, 10], and such methods were proposed for recovery from pooled COVID-19 tests in [11–16]. We conclude the chapter with a qualitative comparison between these families of AI methods in section "Comparative Discussion and Summary", and identify the setting in which one is likely to be preferable over the other for pooled COVID-19 tests.

System Model

In this section we present the system model for which we review pooled recovery algorithms in sections "Group Testing Methods for COVID-19" and "Compressed Sensing for Pooled Testing for COVID-19". As we focus on COVID-19 testing, we begin by identifying the specific characteristics that arise from this application of pooling in section "Pooled COVID-19 Tests", based on which we formulate the recovery problem in section "Recovery from Pooled Tests".

The PCR Process

We begin with an overview of the RT-PCR process following [12]. We refer the reader to [17, 18] for a more detailed description.

In the RT-PCR method for COVID-19 testing, a sample in the form of naso- or oro-pharyngeal swabs is collected from a patient, following which it is mixed into a liquid medium. The RNA molecules of the virus (if present) in the original sample and thus also in the liquid medium, get transformed into *complementary DNA* (cDNA). This conversion from RNA to cDNA occurs by a process termed *reverse transcription*. To this mixture, DNA fragments called *primers* are added. These primers behave in a manner that is 'complementary' to the existing cDNA and get attached to certain specific sections of the cDNA (if the virus is present in the sample). This cDNA then undergoes a process of exponential amplification inside the RT-PCR machine during several cycles of alternate heating and cooling. Roughly per cycle, the cDNA is doubled in number. Fluorescent markers added to this mixture produce a visible fluorescence effect, observed almost immediately on a computer screen, in response to, and directly proportional to, the total amount of amplified cDNA. The time when the amount of fluorescence is observed to exceed a machine-specific threshold is known as the *cycle threshold* (CT). A smaller CT value indicates greater number of copies of the virus, and a larger CT value indicates a lower number (potentially even zero). Usually CT values lie anywhere between 16–35 cycles in real experiments. Statistics of typically observed CT values are reported in [19]. In particular, we note that RT-PCR can detect even single molecules. A single molecule typically would have a CT value of around 40 cycles. An RT-PCR machine can typically test 96 samples in parallel, and the testing time is between 3 and 4 h.

Mathematical Model

The readings provided in RT-PCR testing provide an indication on the presence and level and infection. In some cases, one can approximate the relationship between the cycle time of a sample and its viral load using a statistical model [12, 18, 20]. Due to the fact that the growth of cDNA molecules is exponential [18] (as also explained in section "The PCR Process"), the number of molecules of viral cDNA in a sample i at cycle time t is given by: $z_i(1 + q_a)^t$, where z_i is the initial viral load in the sample (before the RT-PCR process began), and $q_a \in [0, 1]$ is an amplification factor. Here t is a real number, and $\lfloor t \rfloor$ denotes the number of RT-PCR cycles that have passed. The fluorescence $\xi_i(t)$ of the ith pool at time t is directly proportional to the number of virus molecules. Hence we have:

$$\xi_i(t) = K \cdot z_i \left(1 + q_a\right)^t, \tag{2.1}$$

where K is a constant of proportionality.

Let F be the threshold value of the fluorescence, and assume that the fluorescence of pool i reaches the value F at some cycle time τ_i, according to (2.1). Due to factors such as the stochasticity of the RT-PCR process reaction and the measurement error/quantization in the RT-PCR machine, the CT value that is output by the machine will not reflect this true cycle time. In other words, the true cycle time τ_i and the recorded cycle time t_i are statistically related, and this relationship can be modelled as $\tau_i = t_i + e_i$, where e_i is interpreted as the error in recording the value of τ_i. Here e_i is modeled as a zero-mean Gaussian random variable with variance σ_e^2. We assume $0 < \sigma_e^2 \ll 1$, as we expect this error to not be too high. Plugging in $\xi_i(\tau_i) = F$, we have from (2.1) that

$$F = K \cdot z_i \left(1+q_a\right)^{\tau_i} = K \cdot y_i \left(1+q_a\right)^{t_i}, \tag{2.2}$$

where y_i is defined to be the noisy initial viral load of sample i corresponding to a noisy observed CT value t_i. Hence,

$$y_i = z_i \left(1+q_a\right)^{\tau_i - t_i} = z_i \left(1+q_a\right)^{e_i}. \tag{2.3}$$

The above stochastic relationship constitutes an approximate model of the underlying statistics of RT-PCR measurements, which one may use for designing and analyzing pooling methods, as discussed in the sequel.

Pooled COVID-19 Tests

The common approach in testing for the presence of COVID-19 is based on the RT-PCR method. Here, a sample is collected, most commonly based on nasopharyngeal or oropharyngeal swabs or saliva [21]. The presence of infection is then examined by RNA extraction via RT-PCR measurements, quantifying the viral load in the sample. The RT-PCR process is quite time consuming (on the typical order of several hours), and can simultaneously examine up to a given number of m inputs (commonly on the order of several tens of samples). This results in a major bottleneck, particularly when the number of patients, denoted by n, is much larger than m.

A candidate approach to reduce the test duration utilizes pooling [6]. Pooling considers a small mixture of pooled samples rather than the total number of samples. To model this procedure in a generic manner, let $\boldsymbol{f} \in \mathcal{R}_+^n$ represent the samples vector, i.e., the viral loads of each of the subjects, whose entries $f_1, \ldots, f_n$ denote the viral loads of the n individuals. If $f_i = 0$, then the ith individual is not infected. Pooled testing is comprised of two stages: pooling and measurement.

Pooling refers to the mapping of the n samples $\boldsymbol{f}$ into a lower-dimensional vector $\boldsymbol{z} \in \mathcal{R}_+^m$, with $m < n$. Here, each entry of $\boldsymbol{z}$ is obtained by mixing a subset of the samples in $\boldsymbol{f}$ determined by the binary valued pooling matrix $\boldsymbol{\Phi} \in \{0,1\}^{m \times n}$, such that

$$z_i = g_{\text{pool}}\left(\left\{f_j | \Phi_{i,j} = 1\right\}\right), \quad i \in \{1,\ldots,m\}. \tag{2.4}$$

In (2.4), $g_{\text{pool}}(\cdot)$ represents how the viral loads of the mixed samples are mapped into the viral load of the pool. In direct (non-pooled) testing, one can write $z_i = f_i$ and $\mathbf{\Phi}$ as the $n \times n$ diagonal matrix. While pooling is commonly modelled as averaging, i.e., $g_{\text{pool}}(\mathcal{X}) = \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}} x$, in many applications, including COVID-19 testing, the relationship between the pool and the mixed samples is more complex and includes stochasticity and dilution [6].

Measurement refers to reading of the pooled vector $\boldsymbol{z}$ into the observed vector $\boldsymbol{y} \in \mathcal{R}_+^m$. This is carried out via the element-wise mapping $g_{\text{meas}}(\cdot)$, such that $y_i = g_{\text{meas}}(z_i)$ for each $i \in \{1,\ldots,m\}$. For RT-PCR, the mapping $g_{\text{meas}}(\cdot)$ represents a procedure which involves the conversion of the sample into complementary DNA through reverse transcription, followed by examining its reaction in response to the heating and cooling cycle through a chemical process to quantify the number of virus particles (as earlier described in section "The PCR Process"). The measurement procedure typically involves some level of uncertainty and distortion, and thus $g_{\text{meas}}(\cdot)$ is typically stochastic. Under some approximations detailed in section "Mathematical Model", one can approximate $g_{\text{meas}}(\cdot)$ using the noise model in (2.3).

An illustration of the overall flow of pooled RT-PCR-based COVID-19 testing is depicted in Fig. 2.1. On the left side of the figure, we show the true viral loads of all n items. In particular, we see that the first item is infected in a medium level, the fourth item is infected in a low level, and all other items are not infected. Next,

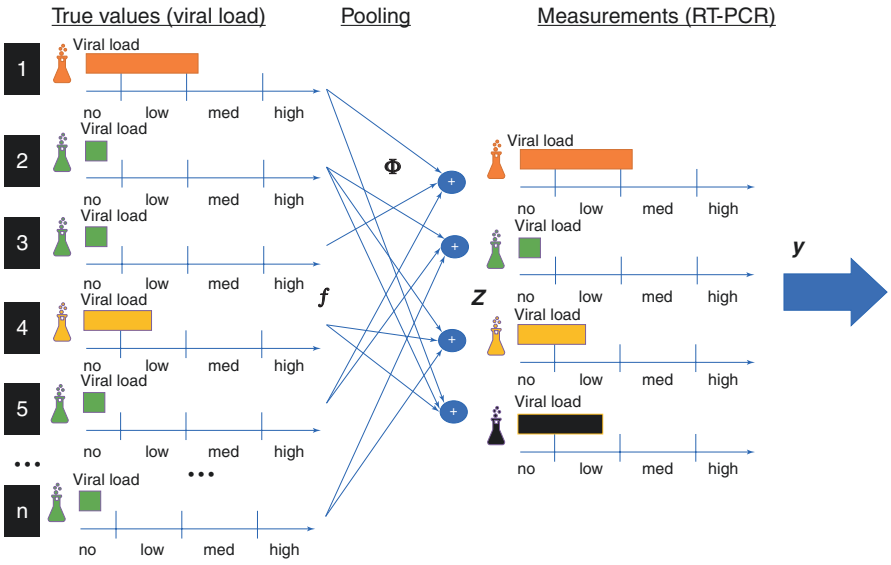


Fig. 2.1 Pooled RT-PCR testing illustration

pooling is done based on a testing matrix, which is generated prior to obtaining the samples. For example, the first pooled test involves samples from the first, third, and fifth items. This results in a measurement vector, denoted by $\boldsymbol{z}$. This vector is fed to the recovery algorithm, which should be able to tell that the first item is infected (with some medium level viral load), the fourth item is infected (with some low level viral load), and all other items are not infected.

While the above description of the pooling and measurement procedure are generic, its application to RT-PCR-based COVID-19 tests gives rise to the following characteristics:

A1 The number of infected measurements, denoted by k, is much smaller than the number of tested individuals n. Typically $k \leq 0.1n$, i.e., up to 10% of the tested population is infected. This number often varies between non-symptomatic patients (where k is typically much smaller than 10%) and symptomatic ones.

A2 The RT-PCR measurements are noisy, i.e., some level of random distortion is induced in the overall process, encapsulating the randomness in the acquisition of the samples, their mixing, and the RT-PCR reading. As a result, the pooling and measurement procedures $g_{\text{pool}}(\cdot)$ and $g_{\text{meas}}(\cdot)$ may be noisy and unknown, with one possible model for $g_{\text{meas}}(\cdot)$ being the relationship formulated in section "Mathematical Model".

A3 There is a limit, denoted by $L > 1$, on the number of subjects which can be mixed together in a single measurement. A typical limit on the number of subjects mixed together is $L = 32$ [6]. Furthermore, the portion taken from each sample for the pooled measurements is identical, e.g., one cannot mix 30% from one patient with 70% from another patient into a single pool.

While the characteristics are identified for swab-based RT-PCR tests, they also hold for other forms of contagious disease testing, including, e.g., serological tests [22].

Recovery from Pooled Tests

Pooling is based on mixing the samples of groups of patients together, forming m mixed samples out of the overall n patients. Then, the presence of COVID-19 for each of the tested individuals must be recovered from the mixed RT-PCR measurements, either directly, i.e., in a *one-shot* fashion, or in an adaptive manner, which involves additional tests [23]. When testing for COVID-19, the following requirements must also be accounted when designing the recovery procedure: [label = *R*].

R1 One-shot tests, where we fully identify all subjects from a single RT-PCR operation, are often preferable over adaptive techniques, which involve having to carry out additional tests based on the results. In particular, one-shot methods are simpler to automate, being less dependent on human intervention compared

to adaptive schemes, and thus typically reduce the overall duration of the testing procedure (though not necessarily the number of pooled tests m).

R2 In some cases, one is interested not in only identifying whether a subject is infected or not, but also in some score on the viral load. This can be either the viral load itself, or a discrete score.

Per 2 the objective of recovery from COVID-19 pooled tests is to recover a score on the infection of each subject. To formulate this, we define a mapping, $Q : \mathcal{R}_+ \mapsto \mathcal{Q}$, which translates a viral load into the corresponding score, while $\mathcal{Q}$ is a set describing all possible infection status. Some examples of such mappings of interest are:

Example 2.1 *When one is only interested in detecting whether a subject is infected or not, then* $\mathcal{Q} = \{0,1\}$*, with* $Q(x) = 1$ *when* x *is larger than some infection threshold, and* $Q(x) = 0$ *otherwise.*

Example 2.2 *Often in practice, one is interested in a discrete score of the viral load, For example, possible discrete score outputs are no (no virus), low (borderline), mid (infected), and high (highly infected); this requirement implies that* $\mathcal{Q} = \{no, low, med, high\}$*. The decision regions are defined by the pre-specified thresholds* $\tau_1, \ldots, \tau_{|\mathcal{Q}|-1}$*, such that* $Q(x)$ *maps a value of* x *not larger than* τ_1 *is treated as not infected, while, e.g., a value in the range* $(\tau_1, \tau_2]$ *is treated as a low level of infection.*

Example 2.3 *When the desired output is the actual viral load, then* $\mathcal{Q} = \mathcal{R}_+$*, with* $Q(x) = x$*.*

Based on the characteristics of pooled COVID-19 tests detailed above, the recovery problem can be formulated as the estimation of $Q(\boldsymbol{f})$ (where $Q(\cdot)$ is applied element-wise) from $\boldsymbol{z}$, denoted by $\hat{\boldsymbol{f}} \in \mathcal{Q}^n$. In particular, for the subset of k infected items of a total of n inspected patients, the goal in pooled recovery is to formulate an algorithm for reconstructing $\hat{\boldsymbol{f}}$ from $\boldsymbol{y}$, possibly along with the design of the $m \times n$ measurement matrix $\boldsymbol{\Phi}$. The measurement matrix should guarantee that at most L subjects are mixed in each pool-test, such that the recovery algorithm can identify the subset of infected items and their representation of the viral load using the recovery algorithm, from the observed vector $\boldsymbol{y}$. The recovery procedure is visualized in Fig. 2.2.

AI-based recovery algorithms typically fall into either of the following mathematical frameworks: GT, which is concerned with detecting the presence of infection (e.g., Example 2.1), and cs, which originates from the recovery of sparse continuous-amplitude vectors, and is thus more naturally suitable for recovering the actual viral loads (e.g., Example 2.3). We discuss these approaches in sections "Group Testing Methods for COVID-19" and "Compressed Sensing for Pooled Testing for COVID-19", respectively. The recovery of a discrete yet non-binary score, as in Example 2.2, requires an extension of GT techniques, which we also present in section "Group Testing Methods for COVID-19".

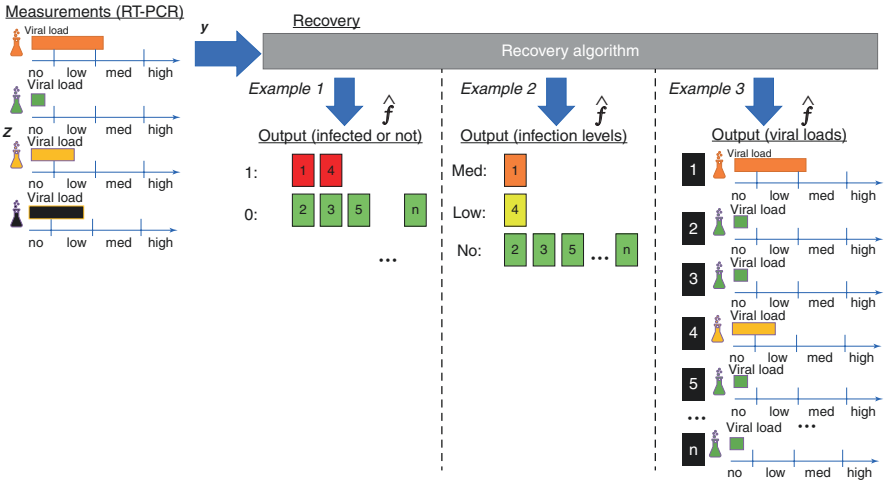


Fig. 2.2 Pooled testing recovery illustration, divided into desired outputs following Examples 2.1–2.3

Group Testing Methods for COVID-19

GT is a mathematical theory derived for detecting infections in large populations. It was introduced by Robert Dorfman in [24], who suggested GT for identifying syphilis-infected draftees during World War II. Since then, GT has been long studied and utilized in many fields, such as biology and chemistry [25, 26], communications [27–29], sensor networks [30], pattern matching [31], web services [32], and cyber security [33–35].

Broadly speaking GT is a framework for group detection problems of binary variables, i.e., each subject can either be infected or not infected [8]. GT is traditionally adaptive, requiring multiple sequential tests [24] in order to achieve a minimal number of overall tests from which the presence of infection can be inferred. Nonetheless, GT can also be applied in a one-shot (non-adaptive) manner [36], avoiding the need to mix new samples during the testing procedure, as studied for COVID-19 pooled testing in [37, 38]. Therefore, in the following we divide our description of GT methods into adaptive and non-adaptive. Unless stated otherwise, our description henceforth focuses on linear measurements, that is we consider $g_{\text{pool}}(\boldsymbol{f}) = \boldsymbol{\Phi f}$ to be the common setting for GT studies.

Adaptive GT Methods

GT was originally proposed as an adaptive procedure. The simple scheme proposed by Dorfman partitions the patients into distinct sets of size B that are pooled together. Then, the subjects of each pool which is tested as positive are re-tested separately one-by-one, where the measurements are assumed to be noiseless, i.e., $g_{\text{meas}}(z) = z$. The resulting formulation is stated below as Algorithm 2.1.

Algorithm 2.1: Dorfman Adaptive GT [24]

Init: Subjects to be inspected $\boldsymbol{f} \in \mathcal{R}_+^n$, set size B
Pooling:
1. Set number of pools to $m = \lceil n/B \rceil$;
2. Set measurement matrix such that $(\boldsymbol{\Phi})_{i,j} = 1$ if $i \in [(j-1) \cdot B, j \cdot B)$ and $(\boldsymbol{\Phi})_{i,j} = 0$ otherwise

Measurement:
3. Measure $\boldsymbol{y}$

Recovery:
4. Set $\hat{\boldsymbol{f}} = 0$
5. **for** each j such that $y_j \neq 0$ **do**
6. Set each $\hat{f}_i$ for which $(\boldsymbol{\Phi})_{i,j} = 1$ by testing f_i individually
7. end

Dorfman's method, which is simple to implement and highly intuitive, has been used for COVID-19 pooled testing in [6, 7, 39]. When one has prior knowledge of the number of defective items k, the setting of B which minimizes the maximal number of tests is $B = \sqrt{n/k}$ [40]. The adaptive nature of Algorithm 2.1 stems from the fact that it operates in two stages. Yet, adaptive GT can be more efficient in terms of the maximal number of tests when operating with multiple stages in a recursive manner. A representative implementation of recursive adaptive GT is the methods proposed in [41] which is based on binary splitting. Broadly speaking, binary splitting recursively partitions the subjects as long as they contain at least one infected item. The resulting procedure is formulated as Algorithm 2.2. Recursive adaptive GT based on binary splitting is known to be most beneficial when the k is much smaller compared with n, Furthermore, it was conjectured in [42] that gt, and pooling in general, should only be applied when the subjects are indeed sparse (as per *A1*), and particularly when $k < n/3$.

Algorithm 2.2: Binary Splitting Adaptive GT [41]

Init: Subjects to be inspected $\boldsymbol{f} \in \mathcal{R}_+^n$, index of first subject i
Pooling:
Set measurement matrix $\boldsymbol{\Phi}$ to be the $1 \times n$ all-ones vector;
Measurement:
Measure (the scalar) y;
Recovery:
if $y \neq 0$ **then**
  if $n = 1$ **then**
    Set $\hat{f}_i$ as infected;
  end
  else
    Invoke Algorithm 2 for $f_1, \ldots, f_{\lceil n/2 \rceil}$ with first index i;
    Invoke Algorithm 2 for $f_{\lceil n/2 \rceil + 1}, \ldots, f_n$ with first index $i + \lceil n/2 \rceil$;
  end
end
else
  Set $\hat{f}_i, \ldots, \hat{f}_{i+n-1}$ as not infected;
end

The major drawback of adaptive GT for COVID-19 pooled testing stems from the fact that it requires additional procedures based on the results. This implies that laboratory technicians must save swabs taken from patients in addition to those mixed into the pools, and then possibly mix them again based on the outcome of the lengthy RT-PCR procedure. This can constitute a major bottleneck and notably burden laboratory technicians for both two-stage GT (as in Algorithm 2.1), and even more so for recursive GT (as in Algorithm 2.2).

Non-Adaptive GT Methods

Non-adaptive GT operates in a one-shot manner, detecting the presence of infections among all patients without requiring additional tests to be acquired based on previous results. Unlike the adaptive techniques detailed in the previous section, where the recovery algorithm typically dictates the measurement matrix $\boldsymbol{\Phi}$, non-adaptive methods can typically be carried out with different polling patterns. Therefore, in the following we first discuss how to set the pooling matrix $\boldsymbol{\Phi}$, after which we review some representative recovery methods designed for both noiseless and noisy settings.

Pooling Matrix

To determine the testing matrix $\boldsymbol{\Phi}$, one must first fix the number of the pool-tests m. GT theory dictates that m should satisfy $m = (1 + \varepsilon)k\log_2 n$, for some $\varepsilon > 0$, as this is the sufficient number of pool-test for which reliable recovery in noiseless GT can be

guaranteed [43]. The parameter ε controls the probability of error of the procedure [43]. In practice, the number of pools-tests is often dictated by the measurement setup, e.g., it may be constrained to be an integer multiple of the number of inputs accepted by an RT-PCR machine.

The traditional GT method of generating $\mathbf{\Phi}$ for non-adaptive recovery draws its elements in an i.i.d. fashion according to a Bernoulli distribution with parameter p. A common choice for p is $p = 1 - 2^{-1/k}$, for which the probability of each element in $\boldsymbol{z}$ to be zero is 1/2. When k is unknown, p is chosen using a rough approximation of k. A major drawback of this approach is that *A3* is not necessarily satisfied, and there is some chance that too many patients will be mixed into the same pool causing dilution. This can be relieved by forcing the columns of $\mathbf{\Phi}$, as well as the rows of $\mathbf{\Phi}$, to be typical, such that every column/row to have exactly $\lceil p \cdot m \rceil \leq L$ and $\lceil p \cdot n \rceil$ ones, respectively [37]. In practical testing setups, one is interested in using a fixed deterministic matrix, rather than having to work with random matrices. Thus, $\mathbf{\Phi}$ should be generated once before the pooling starts, after which it can be used for multiple pooling experiments.

Noiseless Linear Non-Adaptive Recovery

Noiseless recovery methods are designed for linear noiseless measurements, such that $\boldsymbol{y} = \boldsymbol{z} = \mathbf{\Phi} \boldsymbol{f}$. For such settings, common strategies attempt to classify the subjects into DND and DD based on observing which are mixed into a non-infected pool and which are mixed into an infected one. The identification of the DND is carried out using COMP. The purpose of COMP is to characterize DND subjects as those pooled into a non-infected pool, while classifying the remaining patients as possible defective (PD). The resulting procedure is summarized below as Algorithm 2.3.

Algorithm 2.3: Combinatorial Orthogonal Matching Pursuit [44, 45]

Init: Pooled measurements $\boldsymbol{y} \in \mathcal{R}_+^m$, pooling matrix $\mathbf{\Phi}$, infection threshold τ
Mark each $\{\hat{f}_j\}_{j=1}^n$ as PD;
for $i = 1, \ldots, m$ **do**
  if $y_i \leq \tau$ **then**
    Set $\hat{f}_j$ as DND ($Q(0)$) for each j such that $(\mathbf{\Phi})_{i,j} \neq 0$;
  end
end

While Algorithm 2.3 only divides the patients into DND and PD, it can also be extended to identify some of the subjects as DD. The rationale here is that whenever a pool is tested as infected, one of the subjects mixed into that specific pool must be infected. Therefore if only one PD subject is mixed into an infected pool then this subject is DD. The resulting steps are summarized as Algorithm 2.4.

Algorithm 2.4: Definitely Defective Non-Adaptive Recovery [46]

Init: Pooled measurements $\boldsymbol{y} \in \mathcal{R}_+^m$, pooling matrix $\boldsymbol{\Phi}$, infection threshold τ
Apply COMP (Algorithm 3) to obtain DND and PD subjects;
for *each PD subject* $\hat{f}_j$ **do**
  if $\exists y_i > \tau$ *such that* $(\boldsymbol{\Phi})_{i,j} \neq 0$ *while* $(\boldsymbol{\Phi})_{i,j'} = 0$ *for each* $j' \neq j$ *with PD* $\hat{f}_{j'}$ **then**
    Set $\hat{f}_j$ as DD;
  end
end

The DD procedure in Algorithm 2.4 can be repeated sequentially, i.e., iteratively using previously identified DD subjects to repeat the detection procedure. This iterative operation is referred to as Sequential COMP [46]. It is noted though that these algorithms may result in some of the patients being classified as neither DD nor DND implying that a policy for dealing with such cases is required (though it is likely to be infrequently applied).

Algorithms 2.3 and 2.4 can be used as preliminary processing, after which the remaining PD patients are detected using alternative recovery methods. In particular, for settings where one is interested in recovering the viral loads and not just which patients are infected, COMP can be used at pre-processing to identify the DND subjects and reduce the number of patients whose viral load must be estimated; the viral loads of these remaining PD subjects can then be recovered using methods that are not limited to detection, e.g., loopy belief propagation (LBP) [38], CS [12], and even least-squares estimation [37].

Noisy Non-Linear Non-Adaptive Recovery

The non-adaptive GT recovery algorithms detailed in the previous section are designed for settings where the measurements are linear and noiseless, such that $\boldsymbol{y} = \boldsymbol{z} = \boldsymbol{\Phi f}$. Some of these techniques can be extended to the presence of additive noise, i.e., when $\boldsymbol{y} = \boldsymbol{z} + \boldsymbol{w}$ where $\boldsymbol{w}$ is a noise signal, by increasing the number of pools, as in e.g., noisy COMP [36]. However, these methods do not naturally extend to setting where the relationship between the measurements and the pooled viral loads is not only noisy, but also non-linear (namely $\boldsymbol{z}$ is not equal to $\boldsymbol{\Phi f}$) as in (2.4), which is often the case in RT-PCR-based COVID tests.

When the pooling and measurement mapping are stochastic yet their distribution is known, the recovery rule which minimizes the error probability, and holds for both linear and non-linear measurements, is the maximum a-posteriori proability (MAP) rule

$$\hat{\boldsymbol{f}} = \arg\max_{\boldsymbol{s} \in \mathcal{Q}^n} P\left(\boldsymbol{s}|\boldsymbol{y}\right). \tag{2.5}$$

However, even when the underlying distribution is fully known, computing MAP estimates is a computationally prohibitive task whose burden grows exponentially with n. Nonetheless, its decision rule can be approached with complexity that only grows linearly with n via the LBP algorithm [40, Chap. 3.3, 47].

Consider our pooled testing setup, where the task is to estimate the discrete-valued vector $\boldsymbol{s} \in \mathcal{Q}^n$ with entries $s_i = Q(f_i)$ from the observed $\boldsymbol{y} \in \mathcal{R}_+^m$. Such settings correspond to Example 2.1 (where $|\mathcal{Q}| = 2$) and Example 2.2. Given a pooling pattern matrix $\boldsymbol{\Phi}$, the joint probability distribution of $\boldsymbol{y}$, $\boldsymbol{s}$ can be expressed as a partition

$$P(\boldsymbol{y},\boldsymbol{q}) = P(\boldsymbol{y}|\boldsymbol{s})P(\boldsymbol{s}) = P(\boldsymbol{s})\prod_{i=1}^{m} P(y_i|\mathcal{S}_i), \tag{2.6}$$

where $\mathcal{S}_i = \left\{ s_j | (\boldsymbol{\Phi})_{i,j} = 1 \right\}$. Note that the factorization in (2.6) is dictated by the pooling pattern $\boldsymbol{\Phi}$, while the combined stochastic effect of the pooling and measurement procedures is encapsulated in the conditional distribution $P(y_i|\mathcal{S}_i)$. By defining $g_i(\mathcal{S}_i) := P(y_i|\mathcal{S}_i)$, (2.6) can be represented as a factor graph as illustrated in Fig. 2.3, with *m function nodes* $\{g_i\}$ and *n variable nodes* $\{s_i\}$.

LBP is an iterative algorithm that approximates the marginal a-posteriori probabilities by propagating messages in the factor graph domain according to a flooding schedule [48]. Messages are initialized $\left\{ \mu_{g_i \to s_j}^{(0)}(s), \mu_{s_j \to g_i}^{(0)}(s) \right\}$ for each $s \in \mathcal{Q}$ and for every function node g_i and variable node s_j which share an edge in the factor graph, i.e., $\boldsymbol{\Phi}_{i,j} = 1$. Then, for each vertex, the messages are shared downstream with their neighbors in an iterative fashion. For iteration index t, the variable-to-function messages are computed via

$$\mu_{s_j \to g_i}^{(t+1)}(s) = \prod_{f \in \mathcal{N}(s_j)/g_i} \mu_{f \to s_j}^{(t)}(s), \tag{2.7}$$

and the function-to-variable messages are obtained as

$$\mu_{g_i \to s_j}^{(t+1)}(s) = \sum_{\mathcal{S}_i / s_j} g_i(\mathcal{S}_i) \prod_{z \in \mathcal{N}(g_i)/s_j} \mu_{z \to g_i}^{(t)}(z), \tag{2.8}$$

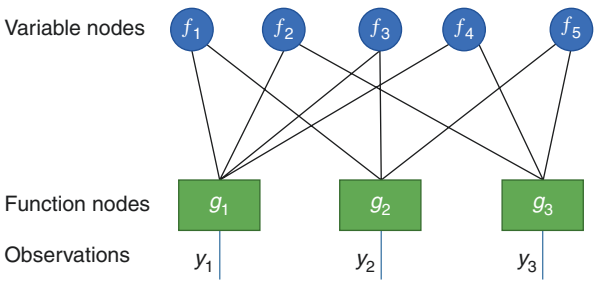


Fig. 2.3 Representation of pooling procedure as a loopy factor graph

where $\mathcal{N}(v)$ denotes the set of neighbors of a given node v. Note that the neighbours of variable nodes are all function nodes, while the neighbors of function nodes are all variable nodes.

Assuming convergence after T iterations, the a-posteriori probability can be approximated as the product of all incoming messages

$$\hat{P}(s_j = s|\boldsymbol{y}) \propto \prod_{f \in \mathcal{N}(s_j)} \mu^{(T)}_{f \to s_j}(s). \tag{2.9}$$

The resulting lbp algorithm is summarized below as Algorithm 2.5.

Algorithm 2.5: Loopy Belief Propagation ([49], Ch. 26])

Init: Pooled measurements $\boldsymbol{y} \in \mathcal{R}_+^m$, pooling matrix $\boldsymbol{\Phi}$

Initialize $\{\mu^{(0)}_{f_i \to s_j}(s), \mu^{(0)}_{s_j \to f_i}(s)\}$ uniformly, i.e., setting $(\mu)_i = \frac{1}{|\mathcal{Q}|}$ for each $i \in \mathcal{Q}$;

for $t = 0, \cdots T-1$ **do**

  Update variable-to-function messages via (7);

  Update function-to-variable messages via (8) to compute $f(\mathcal{S}_i)$;

end

for $j = 1 \ldots, n$ **do**

  Recover infection levels of possibly defective samples by

$$\hat{f}_j = \arg\max_{s \in \mathcal{Q}} \prod_{f \in \mathcal{N}(s_j)} \mu^{(T)}_{f \to s_j}(s).$$

end

Algorithm 2.5 is suitable for non-linear and noisy settings, and is not constrained to detecting infection, being allowed to produce estimates within any predefined finite dictionary. However, it can be computationally demanding, particularly when the number of patients n is relatively large. For this reason, it is often preferable to carry out Comp as a preliminary stage for reducing the dimensionality of the problem. Furthermore, it relies on knowledge of the stochastic model relating the viral loads and the measurements. In traditional GT, simplified models are often adopted ([40], Chap. 3). In COVID-19 testing, when the measurements correspond to the RT-PCR readings of the pool tests, one may either resort to approximated models, as suggested in section "Mathematical Model". Alternatively, as proposed in [38], deep learning tools can be exploited to learn to compute the function nodes as a form of learned factor graphs [50], by training a neural classifier to predict $\mathcal{S}_i = \{s_j | \boldsymbol{A}_{i,j} = 1\} \in \mathcal{Q}^{|\mathcal{S}_i|}$ from y_i. This form of AI-aided recovery allows to carry out LBP recovery with either binary levels (Example 2.1) or multi-level classification (Example 2.2) over the learned factor graph as illustrated in Fig. 2.4. This form of recovery preserves the suitability of LBP to exploit the statistical structures induced by known pooling pattern $\boldsymbol{\Phi}$ while leveraging to model-agnostic nature of neural networks to bypass the need to impose an approximated model on the pooling and RT-PCR measurement processes.

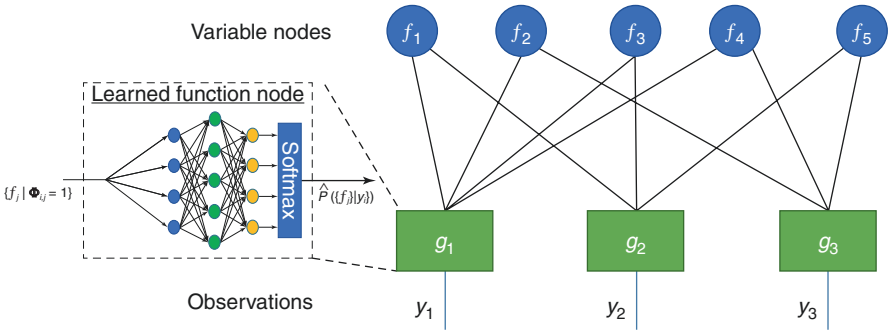


Fig. 2.4 Pooling procedure with loopy learned factor graph

Summary

GT provides powerful AI methods for pooling and recovery from pooled measurements. These techniques are most suitable when one is interested in merely identifying the presence of infection, as in Example 2.1, and when the pooling and measurement procedures can be reliably modeled as linear (and preferably noiseless) operations. GT tools can be applied in both an adaptive and a non-adaptive manner, where the former is quite intuitive yet is less attractive for COVID-19 tests due to its associated overhead in terms of additional technician labour. Furthermore, GT methods, and particularly COMP, can be used as pre-processing steps combined with alternative recovery algorithms, including CS-based discussed in the following section, as means for reducing the dimensionality and complexity of automated recovery.

Compressed Sensing for Pooled Testing for COVID-19

CS is a popular sub-field in the areas of signal and image processing [9, 10]. It involves the acquisition of signals or images directly in compressed format, with a focus on saving acquisition or measurement time. This is different from conventional sensing where a signal or image is acquired in its entirety followed by possibly lossy compression using techniques such as JPEG, JPEG2000 or MPEG.

Consider a signal/image $\boldsymbol{f}$ in vectorized format with n elements. Instead of directly measuring every element of $\boldsymbol{f}$, it is sensed indirectly via a 'sensing matrix' $\boldsymbol{\Phi}$ of size $m \times n$, $m \ll n$ leading to the acquisition of m measurements in a vector $\boldsymbol{y}$. This is represented mathematically as:

$$\boldsymbol{y} = \boldsymbol{\Phi f} + \eta, \tag{2.11}$$

where η is a noise vector with m elements. We note that the matrix-vector operation $\boldsymbol{\Phi f}$ is accomplished via hardware, in a specific manner dictated by the architecture of the compressive device. The aim is to recover $\boldsymbol{f}$ given $\boldsymbol{y}$ and $\boldsymbol{\Phi}$. This problem is

ill-posed in general since $m \ll n$ and this forms an under-determined linear system, even if η were equal to the zero vector. However CS theory states that in the absence of measurement noise, this system of linear equations has a unique solution provided two sufficient conditions are met: ($Q1$) the underlying signal $\boldsymbol{f}$ is sparse, i.e. most of its entries are zero, and ($Q2$) the sensing matrix $\boldsymbol{\Phi}$ has a nullspace which contains no sparse vector other than the zero vector. These conditions are sufficient to ensure unique recovery of $\boldsymbol{f}$ from $\boldsymbol{y}$, $\boldsymbol{\Phi}$ [51]. Moreover, the recovery of $\boldsymbol{f}$ can be accomplished by efficient techniques such as linear programming [52] and various greedy methods. Furthermore, the error in the recovery of $\boldsymbol{f}$ is robust against noise and increases gradually as the noise variance increases.

In many imaging applications, $\boldsymbol{f}$ is not sparse in itself, but is expressible as a sparse or weakly sparse linear combination of column vectors from an orthonormal basis $\boldsymbol{\Psi}$ such as the discrete wavelet transform (DWT) or discrete cosine transform (DCT) in the following manner:

$$\boldsymbol{f} = \boldsymbol{\Psi}\boldsymbol{\theta} = \sum_{l=0}^{n} \boldsymbol{\Psi}_l \theta_l. \tag{2.12}$$

Here a weakly sparse linear combination refers to a linear combination in which most of the coefficients are equal to or very close to zero. That is, the coefficient vector $\boldsymbol{\theta}$ is sparse or weakly-sparse. In this case, the measurement model can be written as:

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\theta} + \eta, \tag{2.13}$$

and we seek to recover $\boldsymbol{\theta}$ from $\boldsymbol{y}$, $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$. Upon recovering $\boldsymbol{\theta}$, it is easy to recover $\boldsymbol{f}$ since $\boldsymbol{f} = \boldsymbol{\Psi}\boldsymbol{\theta}$. Recovery can be efficiently performed using estimators such as the LASSO:

$$\hat{\boldsymbol{\theta}}_\lambda = \text{argmin}_{\boldsymbol{\theta}} \, || \, \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\theta} \, ||_2^2 + \lambda \, || \, \boldsymbol{\theta} \, ||_1 \tag{2.14}$$

or basis pursuit denoising:

$$\hat{\boldsymbol{\theta}}_\lambda = \text{argmin}_{\boldsymbol{\theta}} \, || \, \boldsymbol{\theta} \, ||_1 \text{ s.t.} \, || \, \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\theta} \, ||_2 \leq \varepsilon \tag{2.15}$$

where λ and ε depend on the variance of the noise in η. In both cases, there are theoretical guarantees for the recovery of $\boldsymbol{\theta}$ [52, 53]. These guarantees extend to a variety of different noise models. Matrices generated randomly from the zero mean Gaussian distribution or the ±1 Bernoulli distribution have been proved to obey a property called the Restricted Isometry Property (RIP) [54] with very high probability provided m is lower bounded by a quantity proportional to $s \log n$ where s is the number of non-zero entries in the underlying signal. The RIP implies the property ($Q2$) mentioned earlier involving the null-space.

Compressed Sensing Forward Model for Pooled RT-PCR

A model similar to that in (2.11) can be presented for pooled inference in the case of RT-PCR. Using the notations introduced in section "System Model", this can be expressed by writing $g_{\text{pool}}(\boldsymbol{f}) = \boldsymbol{\Phi f}$, such that the measurement output $\boldsymbol{y} = g_{\text{meas}}(\boldsymbol{\Phi f})$ is given by:

$$\boldsymbol{y} = \boldsymbol{\Phi f} \circ \left(1 + q_a\right)^{\boldsymbol{e}} \Rightarrow \forall i \in [m], y_i = \boldsymbol{\Phi}^i \boldsymbol{f} \left(1 + q_a\right)^{e_i}, \tag{2.16}$$

where $\circ$ denotes the element-wise multiplication of two vectors. Here we note the following:

1. $\boldsymbol{f}$ is a real-valued but *sparse* vector of n elements where each entry f_i represents the viral load in the sample of the i^{th} individual. The sparsity assumption on $\boldsymbol{f}$ is essential for good quality recovery using CS algorithms and it is well justified due to the low prevalance rates in COVID-19 testing [55] as explained earlier.
2. $\boldsymbol{\Phi}$ is a $m \times n$ binary pooling matrix. A pool is created by mixing together small but equal portions of the samples of a subset of the n individuals. For the ith pool, this information is contained in the non-zero entries of the ith row of $\boldsymbol{\Phi}$, i.e. in $\boldsymbol{\Phi}^i$.
3. $\boldsymbol{y}$ is a vector of viral loads in the m different pools.
4. The noise model here is not additive unlike that in (2.11) which is typically used in CS, but multiplicative as explained in section "Mathematical Model" with $q_a \in [0, 1]$ representing an amplification factor and each entry of $\boldsymbol{e}$ being distributed as per $\mathcal{N}\left(0, \sigma_e^2\right)$ where $0 < \sigma_e \ll 1$. Comparing with (2.3), the role of f_i is played by the appropriate viral load value in the ith pool, given by $\boldsymbol{\Phi}^i \boldsymbol{f}$.

The reader may refer to Fig. 2.5 for greater clarity in the pooling process.

CS Algorithms for Recovery

Referring to (2.16), the aim is to recover the sparse vector $\boldsymbol{f}$ from $\boldsymbol{y}$, $\boldsymbol{\Phi}$. Since $\boldsymbol{f}$ is real-valued (and also non-negative), such a procedure gives us viral-load values as opposed to just a binary indicator health status indicator ('infected' versus 'non-infected') as obtained from traditional group testing techniques surveyed in section "Group Testing Methods for COVID-19".

As the pandemic began spreading all over the world, a number of research works that used CS estimators for pooled testing emerged in the literature. These techniques include the Tapestry approach [12], the multi-stage approach in [37], the P-BEST method [56] and the initial work in [13]. We survey the algorithms employed in these techniques here below:

1. In the Tapestry approach [12, 57], a number of different CS algorithms were explored and compared—on synthetically generated as well as real data. These

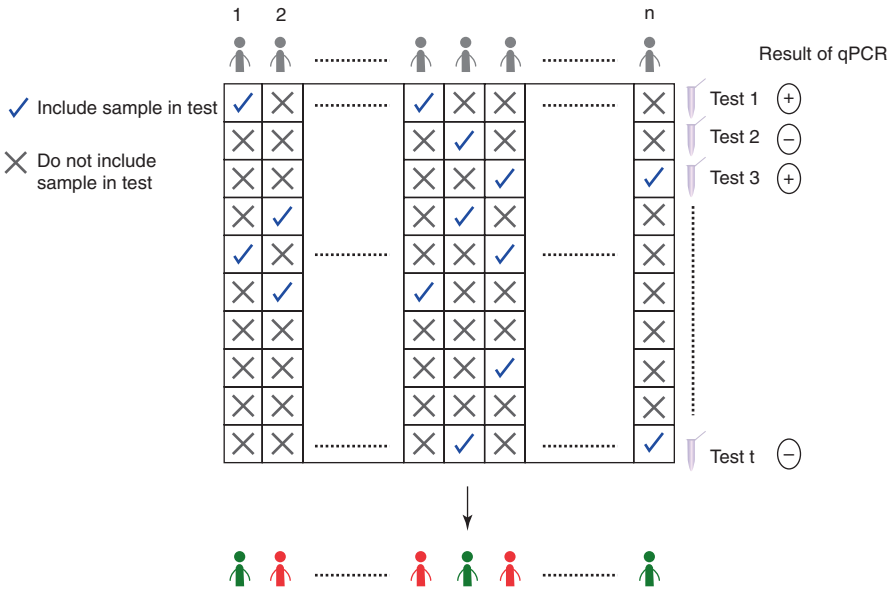


Fig. 2.5 RT-PCR pooling process in a compressive framework. The results of $m = t$ RT-PCR pools are on the RHS in the form of vector $\boldsymbol{y}$, which is a noisy version of $\boldsymbol{\Phi f}$. The entries of $\boldsymbol{\Phi}$ are represented by crosses (0) and ticks (1)

algorithms include the LASSO [53] with non-negativity constraints (NNLASSO) (see Eqn. 2.14), sparse Bayesian learning [58] SBL, non-negative least squares NNLSor non-negative least absolute deviation NNLAD [59], and greedy algorithms such as orthogonal matching pursuit [60] with non-negativity constraints NNOMP [61]. The NNLS and NNLAD approaches were also experimented with in [14]. These algorithms were executed after the traditional group testing algorithm COMP described in section "Group Testing Methods for COVID-19" and outlined in Algorithm 2.3. This was deemed particularly advantageous if there exist some pools with zero viral load, which immediately implies that the viral loads of the samples which contributed to this pool are all zero. Using this COMP step improves prediction performance and also improves computational efficiency as it reduces the problem size. As COMP is only a binary approach, the subsequent steps involving estimation of real-valued viral loads remain very important. Theoretical properties of this combined approach were explored in detail in [12]. It was empirically observed that COMP followed by SBL yielded the best performance among all other approaches such as NNLASSO, NNLS, NNLAD and NNOMP. An adaptive, multi-stage extension of the Tapestry system was presented in [62] (also see section "Comparative Discussion and Summary").

2. The multi-stage approach in [37] adopts a unique multi-stage strategy that beautifully combines three techniques: a variant of \COMP called 'definite defectives' (DD) described in section "Group Testing Methods for COVID-19" and outlined in Algorithm 2.4, followed by a maximum-likelihood approach from [43] (also

called 'Boolean compressed sensing'), followed by an iterative viral loads estimation algorithm. Extensive experiments showcase the beneficial contribution of each of the three stages.
3. The P-BEST approach [56] followed the same strategy as their group's earlier work on pooled testing in agricultural applications [63, 64]. The estimator here is a form of NNLASSO but with the viral loads quantized to three values: 0, 1, 2 (see Eqn. 1 of [63]), representing zero viral loads, intermediate viral loads and high viral loads respectively.
4. To our best knowledge, the earliest piece of work on use of CS for COVID-19 testing appeared in [13] where a basis pursuit denoising (BPDN) approach (see Eqn. 2.15) was adopted.

Details of Algorithms

For the sake of completeness, we briefly explain the various algorithms referred to in the previous section: NNLASSO, BPDN, NNLAD, NNLS, SBL and NNOMP.

1. NNLASSO: This approach works by minimizing the following cost function:

$$J_1(\boldsymbol{f}) \triangleq \| \boldsymbol{y} - \boldsymbol{\Phi f} \|_2^2 + \lambda \| \boldsymbol{f} \|_1 \text{ s.t.} \boldsymbol{f} \succeq \boldsymbol{0}, \tag{2.17}$$

   where $\succcurlyeq$ refers to an element-wise $\geq$ inequality and $\boldsymbol{0}$ refers to a vector of zeros with the same number of elements as in $\boldsymbol{f}$. The minimization procedure proceeds by standard algorithms such as iterative shrinkage and thresholding (ISTA) or its fast variants [65]. The regularization parameter λ ideally depends on the noise level but can be selected using cross-validation.
2. BPDN: This is a constrained version of the earlier NNLASSO problem, and acquires the form of minimizing the following cost function:

$$J_2(\boldsymbol{f}) \triangleq \| \boldsymbol{f} \|_1 \text{ s.t.} \| \boldsymbol{y} - \boldsymbol{\Phi f} \|_2^2 \leq \varepsilon \text{ and } \boldsymbol{f} \succeq \boldsymbol{0}, \tag{2.18}$$

   where ε depends on the noise level.
3. NNLAD and NNLS: These methods work based on minimizing only the data fidelity term with a non-negativity constraint. This non-negativity constraint is enough to guarantee uniqueness and stability of the solution with high probability, in conjunction with a random binary matrix $\boldsymbol{\Phi}$, as analyzed in [14, 59]. The cost function for NNLAD is given as follows:

$$J_3(\boldsymbol{f}) \triangleq \| \boldsymbol{y} - \boldsymbol{\Phi f} \|_1 \text{ s.t.} \boldsymbol{f} \succeq \boldsymbol{0}. \tag{2.19}$$

   The cost function for NNLS is

$$J_4(\boldsymbol{f}) \triangleq \| \boldsymbol{y} - \boldsymbol{\Phi f} \|_2^2 \text{ s.t.} \boldsymbol{f} \succeq \boldsymbol{0}. \tag{2.20}$$

4. NNOMP: This is a variant of the popular orthogonal matching pursuit (OMP) algorithm. OMP is a very well known greedy approximation technique [60], and the variant imposes a non-negativity constraint on the solution. The original algorithm requires the solution of multiple least-squares solution of increasing sizes in every iteration. In the variant, we would require least-squares solutions with non-negativity constraints which adds to the overall computational cost. Fast variants have been developed in [61].
5. Sparse Bayesian Learning (SBL): This [58, 66] is a non-convex optimization algorithm based on Expectation-Maximization (EM). In recent empirical studies on compressed sensing, SBL has demonstrated the best reconstruction performance in comparison to many other algorithms [67]. SBL considers a Gaussian likelihood function for $\boldsymbol{y}$ given $\boldsymbol{\Phi f}$ and a Gaussian prior on every element of $\boldsymbol{f}$ with different unknown variances for each element, in the following manner:

$$p\left(\boldsymbol{y}|\boldsymbol{\Phi f}\right) = \frac{\exp\left(-\|\boldsymbol{y}-\boldsymbol{\Phi f}\|_2^2 / \left(2\sigma^2\right)\right)}{\left(2\pi\sigma^2\right)^{m/2}} \tag{2.21}$$

$$p\left(f_i;\gamma_i\right) = \frac{\exp\left(-f_i^2/\left(2\gamma_i\right)\right)}{\sqrt{2\pi\gamma_i}}; \gamma_i \geq 0. \tag{2.22}$$

In this formulation, note that both $\boldsymbol{x}$ and $\{\gamma_i\}_{i=1}^n$ are unknown. Many theoretical results regarding the properties of the EM procedure in SBL are analyzed in [58]. In particular, it is proved that all local minima are sparse vectors. Unlike the previous algorithms, no non-negativity constraint on the elements of $\boldsymbol{f}$ is explicitly imposed.

Most of the aforementioned use a squared loss for the fidelity function which is essentially related to additive Gaussian noise. However, we know that for CS used for RT-PCR, the noise is multiplicative as seen in section "Mathematical Model" and (2.16). However, following ([68], Sec. 7), if we perform a first order Taylor series approximation on both sides of (2.16), we obtain the following:

$$y_i \approx \boldsymbol{\Phi}^i \boldsymbol{f} + \boldsymbol{\Phi}^i \boldsymbol{f}\left[\log\left(1+q_a\right)\right]e_i. \tag{2.23}$$

This approximation is reasonably accurate because the error term e_i (see section "Mathematical Model") is a Gaussian random variable with a small variance. Given this approximated noise model, one would have preferred a fidelity term of the form $\sum_{i=1}^{m}\left(y_i - \boldsymbol{\Phi}^i\boldsymbol{f}\right)^2 / \left(\boldsymbol{\Phi}^i\boldsymbol{f}\right)^2$. However this function is non-convex and the squared loss $\sum_{i=1}^{m}\left(y_i - \boldsymbol{\Phi}^i\boldsymbol{f}\right)^2$ has been widely used instead.

As explained here, CS-based algorithms infer quantitative viral load information. Although a non-infected person has a viral load of zero, the algorithms may report very small non-zero viral loads for such samples. This is due to small algorithmic biases, which can produce unduly low specificity rates (high false positives). To prevent this, the viral loads must be put through some threshold τ such that any estimated viral load value (call it $\hat{x}_i$ is to be regarded as zero if $\hat{x}_i < \tau$, and non-zero otherwise. Such strategies have been followed in the literature [12]. The optimal choice of the threshold τ needs to be decided a priori by 'learning' on past data comparing pooling to sequential testing so as to optimize one of the performance measures outlined in section "Assessment of Algorithm Performance and Experimental Protocols". The optimal threshold τ may potentially vary based on prevalence rates.

Assessment of Algorithm Performance and Experimental Protocols

The various algorithms in the literature are typically compared with respect to the following criteria, where the $\hat{\boldsymbol{f}}$ refers to an estimate of $\boldsymbol{f}$:

1. RMSE:= $\| \boldsymbol{f} - \hat{\boldsymbol{f}} \|_2 \; / \| \boldsymbol{f} \|_2$
2. Number of false positives (FP):= $\left| \left\{ i : f_i = 0, \hat{\boldsymbol{f}}_i > 0 \right\} \right|$
3. Number of false negatives (FN):= $\left| \left\{ i : f_i > 0, \hat{\boldsymbol{f}}_i = 0 \right\} \right|$
4. Sensitivity (also called Recall or True Positive rate):= # correctly detected positives/# actual positives
5. Specificity (also called True Negative Rate):= # correctly detected negatives/# actual negatives.

The experiments to assess the performance of the various CS algorithms can be classified broadly as experiments on (1) synthetic and (2) real data.

For the prediction of $\boldsymbol{f}$ from $\boldsymbol{y}$, $\boldsymbol{\Phi}$ from synthetically generated data [12–14, 37], the vector $\boldsymbol{f}$ is chosen to be sufficiently sparse and the non-zero entries are drawn uniformly at random from the interval $[0, 32768]$. Alternatively, the distribution of cycle threshold (CT) values ([19], Fig. 2.3) could be used to generate CT data and convert those into viral loads, although such an approach has not been followed in the literature to our best knowledge. The pools have been synthetically simulated using the noise model from (2.16) using different types of $\boldsymbol{\Phi}$ matrices, as surveyed later in section "Choice of Pooling Matrices".

For the experiments on real data, the work in [12] used artificially injected viral RNA to create positive samples. These data were obtained from the Wyss Institute

at Harvard, USA and the National Centre for Biological Sciences, India. In [12], pooling matrices of sizes 16×40 and 24×60 were used and the number of infected samples was varied between 0 and 4 (prevalence rate up to 10%). Experimental results on pooled inference showed mostly 0 and much less commonly 1 or 2 false negatives, and between 0 and 3 false positives with the distribution of false positives skewed heavily towards 0 or 1. The work in [56] used left-over samples that had been previously tested for COVID19. Their experiments considered pooling matrices of size 48×384 with the number of positive samples ranging between 2 to 5 (prevalence rate less than 1%). Their approach yielded no false negatives and an occasional false positive.

Choice of Pooling Matrices

Pooling matrices for compressive recovery in pooled testing must allow for successful recovery of $\boldsymbol{f}$ from $\boldsymbol{y}$. Moreover, unlike typical CS, these matrices must also obey certain additional physical constraints (unique to pooling tasks) which have been explained elaborately in [12, 68]. We briefly explain these constraints here below. These constraint have been mentioned earlier in section "Pooled COVID-19 Tests", but we explain them here again in the context of specifying properties of compressed sensing matrices.

1. The pooling matrix must be binary, for ease of pooling. Thus, $\Phi_{ij} = 0$ means that the jth sample did not contribute to the ith pool, and $\Phi_{ij} = 1$ means that a (small) fixed volume from the jth sample contributed to the ith pool. Note that the contributing volumes are equal across pools and across samples. It is understood that the total sample of a single individual will contribute to multiple pools. Hence while creating a pool, only small (fixed-volume) portions from the contributing samples are collected and mixed together.
2. They must be sparse, again for ease of pooling. Sparsity of $\boldsymbol{\Phi}$ is required to ensure two important properties: ($Z1$) that not too many samples contribute to a given pool, and ($Z2$) that a single sample does not contribute to too many different pools. Property ($Z1$) is important because the sample volume added to an RT-PCR reaction in the RT-PCR machine is fixed. Due to this, an increase in pool size means that each sample would contribute a smaller amount of that volume to any given pool. This can adversely affect the power of RT-PCR to differentiate positive samples from negative ones. Property ($Z2$) is important because the contribution of one sample to a very large number of pools could lead to sample depletion.

Several different types of combinatorial design strategies have been used in the literature to design $\boldsymbol{\Phi}$ matrices so as to obey the aforementioned constraints. The work

in [12] uses Kirkman triple matrices where each sample contributes to exactly 3 pools and each pool is created from a fixed number of samples. These $m \times n$ Kirkman matrices also have additional flexibility in terms of size, namely that n can be any integer multiple of $C(m, 3)$. Kirkman matrices are also adjacency matrices that correspond to certain types of 'expander graphs' which have interesting properties in compressed sensing, including obeying a variant of the RIP as discussed in detail in [12]. As they satisfy the RIP, it follows that they allow for good compressive recovery. The work in [56, 63] uses matrices designed from Reed–Solomon codes seeking inspiration from the rich classical literature on error correcting codes [69], allowing inherent robustness to experimental noise and variations in viral concentration.

In addition, there exists a large body of literature on the combinatorial design of binary sensing matrices [70–72]. However these designs are applicable to only fixed sizes, which may not be relevant to the sizes required for RT-PCR pooling. For example, the proposed designs in [72] have sizes $r^2 \times r^{l+1}$ where r is a prime power, and l is an integer such that $1 < l < r$. This choice of matrix size may not be immediately relevant to RT-PCR and has unnatural restrictions. To address this issue, the recent work in [68] considers sparse balanced binary matrices of arbitrary size. These matrices have the property that all columns of the matrix contain an equal number of ones, and that all rows of the matrix also contain an equal number of ones. These can be designed for a much wider variety of matrix sizes.

Choice of Number of Pools

The choice of m, the number of pools, is an important aspect of the design of $\mathbf{\Phi}$. Let us denote the number of non-zero entries in the underlying viral loads vector $\boldsymbol{f}$ by s. Intuitively speaking, the smaller the value of s, the smaller will be the required number of pools m. As per compressed sensing theory, m is $\Omega(s \log n)$ [52, 73]. However s is usually unknown, and hence configuring an acceptable number of pools is a difficult task. There are two ways to get some rough estimate of s: (1) use knowledge of prevalence rates from a particular city or area where the pooling is being performed, and (2) determine s on the fly directly from the pooled tests. The latter has been accomplished in [12] using binary pooling matrices $\mathbf{\Phi}$ using properties of the binomial distribution, based on a technique proposed in [74, 75]. These techniques, of course, provide only estimates which could be error-prone, but they do give a good idea of the number of pools required. In fact, if it turns out that s is too high, pooling may not longer be a viable option due to loss of accuracy, and sequential testing is then preferable [12]. In the present literature [12, 13, 37, 56], the number of pools is chosen somewhat conservatively assuming low values of s, and without necessarily always computing s as a first step.

Use of Side Information in Pooled Inference

In this section, we cover a very different aspect of pooled inference—namely, the use of extra information, or side information. We first describe the different types of side information, followed by its use in CS algorithms and then its use for improving traditional group testing methods such as Dorfman pooling.

The sparsity of $\boldsymbol{f}$ is a useful and valid assumption given the generally low prevalence rates for COVID-19 [55]. However, there is additional side information (SI) that is available in such applications, which can improve the performance of pooled inference. One form of SI is knowledge of the prevalence rate. This was exploited by means of a probabilistic decoder based on message passing in [16]. Other forms of SI can include knowledge of an individual's symptoms, medical history and travel history. Over and above this, knowledge of family memberships is useful SI, because COVID-19 is likely to spread from one family member to another. Here the notion of a family includes not just the conventional notion, but also includes groups of people in frequent close contact, such as health care professionals working in the same hospital wing, students sharing the same dorm, security officers working at the same checkpoint, etc. Furthermore, contact tracing has been widely employed as a means to control the pandemic [76, 77]. Such information, includes the duration of contact between pairs of individuals and measures of physical proximity (distance). Such information can be collected using one or more of the following methods: Bluetooth in mobile phones [78], the global positioning system [79], manual inquiries by social workers [80, 81], and financial transaction data [77, 79]. There are privacy issues associated with the collection of contact tracing or family information, however it should be noted that these types of side information have one advantage—namely that they do not require collection of medical history information or medical data.

In [68, 82], two classes of estimators are explored to explicitly account for family information and contact tracing information to improve the performance of pooled inference. The first class of techniques involves extensions of the LASSO such as the group LASSO [83] or the overlapping group LASSO [84]. The group LASSO uses the fact that the number of infected *families* is small (as opposed to the number of individuals, as in normal sparsity). This is extended to handle overlapping 'families' as determined from contact tracing information. This class of techniques uses quantitative information and the multiplicative noise model in RT-PCR. The second class of techniques in [68, 82] involves extensions of Bayesian algorithms such as generalized approximate matching pursuit (GAMP) to account for family structure and contact tracing information. This model is explicitly for binary health status vectors (as opposed to viral load vectors) and accounts for binary noise in the pooled results. This latter set of techniques explicitly accounts for person-to-person transmission probabilities. Both classes of techniques demonstrate superior empirical performance (in terms of number of false positives and false negatives) over their counterparts which account for no family or contact tracing SI and use only the sparsity of the vector of viral loads or health statuses. Independent work in [85, 86]

also demonstrated the benefits of community-based SI on the performance of pooling. The emphasis of the work in [85, 86] is to use the community information to design effective encoders, for which the authors provide theoretical analysis. They employ a loopy belief propagation (LBP) decoder which accounts for community information and show that it outperforms the basic LBP variant which does not account for such information. Their decoders work in tandem with binary health status vectors.

There exist many other pieces of work which account for SI, mostly in the form of prevalence rates or prior infection probabilities, but they do not do so in a non-adaptive CS framework but rather in the framework of traditional group testing algorithms such as Dorfman pooling. Such individual prior infection probabilities can be derived based on factors such as medical or travel history, age, gender, etc. In particular in works such as [87–89], there is emphasis on segregating people into different groups based on their individual prior infection probabilities, and performing pooling of samples belonging to individuals from a given group only. This is in tune with recent publications from the WHO [90] which note reduction in the number of tests if pooling is performed on samples of people with similar infection probability values. There also exist works such as [91–94] which account for correlations between samples of different individuals (represented as nodes of a graph). Such correlations can be obtained from various aspects of a social graph or via contact tracing information. However the works in [91–94] propose adaptive algorithms, quite unlike the non-adaptive CS based algorithms considered in [68, 82, 85, 86].

Comparative Discussion and Summary

In this chapter, we surveyed the applications of AI techniques influenced by GT and CS algorithms to pooled testing of COVID-19 RT-PCR samples. Here, we present a comparative discussion of the relative advantages and disadvantages of the two approaches:

1. Traditional approaches such as Dorfman's method are two-stage procedures. In particular, the output of the first stage acts as an input to the second stage. In both stages, RT-PCR testing is required, and the two stages cannot be executed in parallel. Such approaches are termed 'adaptive'. On the other hand, non-adaptive GT and CS based approaches test all pools in *parallel* and the individual health status values are made available. These are only single-stage approaches and also referred to as being 'one-shot'. Note that even if these non-adaptive approaches employ different algorithms in a sequential manner (*e.g.*, [12, 37]), there is only one *testing* stage. As RT-PCR is a time-consuming process requiring about 4 h, a single-stage procedure is more efficient in terms of time.
2. Adaptive approaches typically have shown lower false-negative and false-positive rates as compared to non-adaptive approaches [12, 62], even though the

non-adaptive approaches do perform quite well. However the work in [12] which focuses on non-adaptive CS-based algorithms typically reports a requirement of fewer tests than with Dorfman's method, besides the fact that the former algorithms do not require more than one stage of testing unlike Dorfman's method which requires two stages. A similar set of simulations were run here for different number of infected samples (in simulation) and comparative results for Dorfman pooling and the COMP-SBL algorithm from section "Details of Algorithms" are presented in Figs. 2.6 and 2.7.

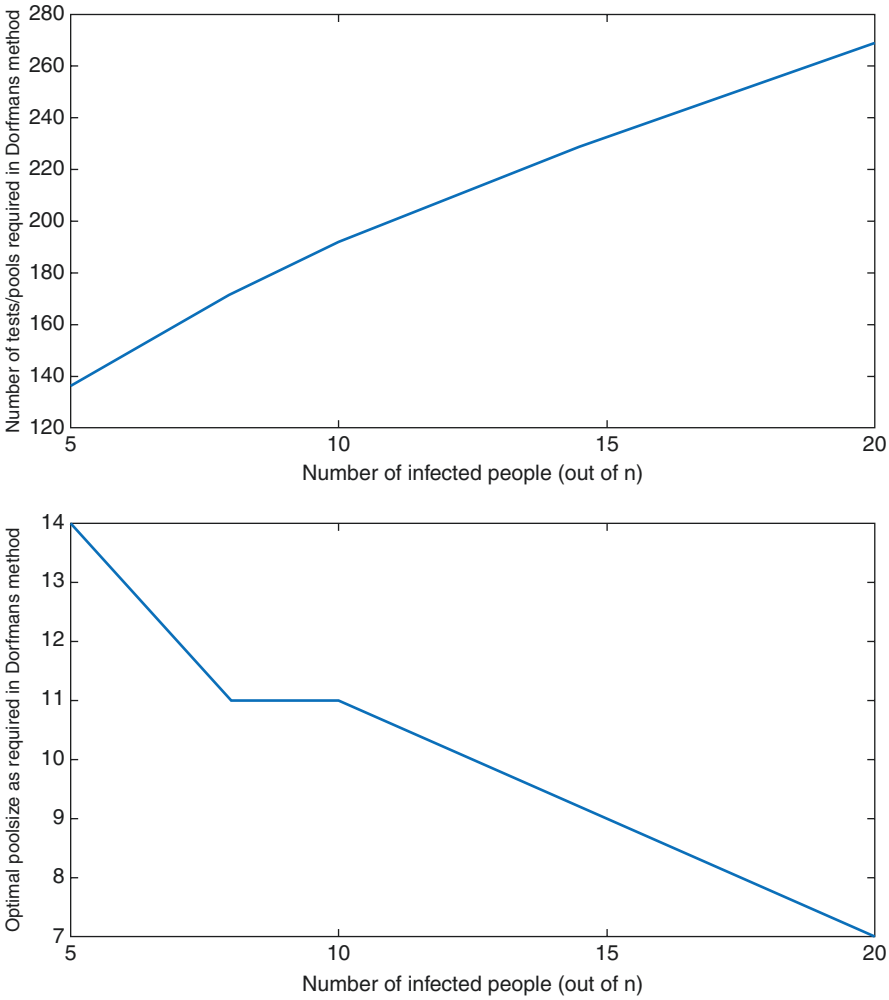


Fig. 2.6 Average number of tests, i.e. number of pools (top sub-figure) and optimal poolsize (bottom sub-figure), required for Dorfman pooling with $n = 961$ samples, with s denoting the number of infected samples. Dorfman pooling yields no false positives or false negatives. Compare these results with the sensitivity and specificity results with the CS algorithm COMP-SBL in Fig. 2.7. These results are similar to Tables 3 and 6 of [12]

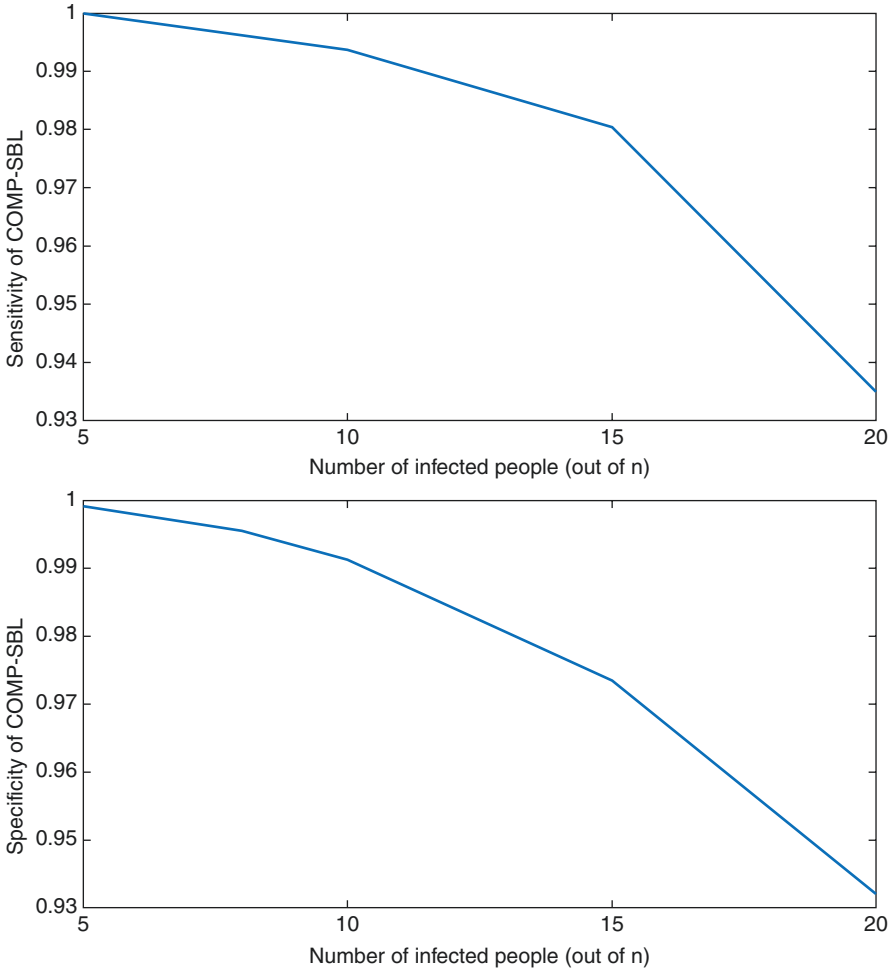


Fig. 2.7 Average sensitivity (top sub-figure) and specificity (bottom sub-figure) results for the CS algorithm COMP-SBL with $m = 93$ pools and $n = 961$ samples, where the number of infected samples is denoted as s. Compare these results with the number of tests required for Dorfman pooling in Fig. 2.6. The variance values for all entries are very small. As can be seen, the number of pools required is smaller than that for Dorfman pooling. These results are similar to Tables 3 and 6 of [12]

3. Both classes of methods require prior knowledge of the number of non-zeros in $\boldsymbol{f}$ to configure the optimal number of tests. Also see section "Choice of Number of Pools". In the existing literature, however, this knowledge has not been explicitly used for configuring the number of required tests.
4. With some exceptions, most adaptive techniques work in a regime where $\boldsymbol{f}$ is binary. The selection of non-adaptive approaches should typically depend on what type of output is required. When one seeks to recover the viral loads for diagnosis (as in Example 2.3), non-adaptive approaches that use CS naturally

treat $\boldsymbol{f}$ as non-negative real-valued [12, 13, 37, 56]. When one is interested in merely identifying the presence of infection, GT methods are natural candidates being designed for regimes where $\boldsymbol{f}$ is binary, though one can also adopt binary variations of CS techniques, see, e.g., [16, 85, 86].

Pooling for RT-PCR has been employed on the ground in several countries and university campuses. An incomplete list of this is presented in [95]. In most of these places, each pool typically contained contributions from about a dozen or two dozen samples. The listing at [95] includes many provinces/states in countries such as USA, UK Switzerland, India, Ecuador, Israel, China, Belgium, Brazil and many others. The listing mentions several large pooling based studies or experiments, involving thousands and occasionally even millions of samples. Most of these experiments employed GT techniques, most commonly Dorfman pooling (Algorithm 2.1), although a few employed CS based methods [56, 96] as explained in section "Compressed Sensing for Pooled Testing for COVID-19". The usage of pooling was shown to yield notable savings and considerable increase in testing throughput [97]. We believe that this widespread usage of pooling amply illustrates its practical relevance during this pandemic. This also illustrates the importance of artificial intelligence and algorithmic tools in combating the spread of COVID-19.

The AI algorithms covered in this survey chapter typically do not require training in the conventional sense that many AI techniques use. However the algorithms can benefit from prior information which can be 'learned' by observing relevant COVID-19 data. This includes inference of prevalence rates as explained in section "Choice of Number of Pools", or the choice of optimal thresholds in compressed sensing algorithms (see section "Details of Algorithms") to decide whether certain viral loads are to be regarded as zero or non-zero, or side information as explained in section "Use of Side Information in Pooled Inference". One could further incorporate ideas from AI or ML techniques that have mined useful knowledge from available data regarding the spread of COVID-19, including information from the epidemiology literature [98]. The incorporation of such information to improve the performance of pooling is a good avenue for future research.

References

1. Salathé M, Althaus CL, Neher R, Stringhini S, Hodcroft E, Fellay J, et al. COVID-19 epidemic in Switzerland: on the importance of testing, contact tracing and isolation. Swiss Med Wkly. 2020;150(11–12):w20225.
2. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al. Fair allocation of scarce medical resources in the time of Covid-19. N Engl J Med. 2020;382(21):2049–55.
3. Lucia C, Federico PB, Alejandra GC. An ultrasensitive, rapid, and portable coronavirus SARS-CoV-2 sequence detection method based on CRISPR-Cas12. bioRxiv. 2020;2020:1127.
4. Ben-Assa N, Naddaf R, Gefen T, Capucha T, Hajjo H, Mandelbaum N, et al. SARS-CoV-2 on-the-spot virus detection directly from patients. medRxiv. 2020;2020:2389.
5. Nolan T, Hands RE, Bustin SA. Quantification of mRNA using real-time RT-PCR. Nat Protoc. 2006;1(3):1559–82.

6. Yelin I, Aharony N, Shaer-Tamar E, Argoetti A, Messer E, Berenbaum D, et al. Evaluation of COVID-19 RT-qPCR test in multi-sample pools. MedRxiv. 2020;2020:2389.
7. Hanel R, Thurner S. Boosting test-efficiency by pooled testing strategies for SARS-CoV-2. 2020. https://arxiv.org/abs/2003.09944.
8. Gilbert AC, Iwen MA, Strauss MJ. Group testing and sparse signal recovery. In: 2008 42nd Asilomar conference on signals, systems and computers. Piscataway, NJ: IEEE; 2008. p. 1059–63.
9. Eldar YC, Kutyniok G. Compressed sensing: theory and applications. Cambridge: Cambridge University Press; 2012.
10. Eldar YC. Sampling theory: beyond bandlimited systems. Cambridge: Cambridge University Press; 2015.
11. Shental N, Levy S, Skorniakov S, Wuvshet V, Shemer-Avni Y, Porgador A, et al. Efficient high throughput SARS-CoV-2 testing to detect asymptomatic carriers. Sci Adv. 2020;6(37):eabc5961.
12. Ghosh S, Agarwal R, Rehan MA, Pathak S, Agarwal P, Gupta Y, et al. A compressed sensing approach to pooled RT-PCR testing for COVID-19 detection. IEEE Open J Signal Process. 2021;2:248–64.
13. Yi J, Mudumbai R, Xu W. Low-cost and high-throughput testing of COVID-19 viruses and antibodies via compressed sensing: system concepts and computational experiments. 2020. https://arxiv.org/abs/2004.05759.
14. Petersen HB, Bah B, Jung P. Practical high-throughput, non-adaptive and noise-robust SARS-CoV-2 testing. 2020. https://arxiv.org/abs/2007.09171.
15. Yi J, Cho M, Wu X, Xu W, Mudumbai R. Error correction codes for COVID-19 virus and antibody testing: using pooled testing to increase test reliability. 2020. https://arxiv.org/abs/2007.14919.
16. Zhu J, Rivera K, Baron D. Noisy pooled PCR for virus testing. 2020. https://arxiv.org/abs/2004.02689.
17. Jawerth N. How is the COVID-19 virus detected using real time RT-PCR? https://www.iaea.org/newscenter/news/how-is-the-covid-19-virus-detected-using-real-time-rt-pcr.
18. Efficiency of Real-Time PCR. https://www.thermofisher.com/in/en/home/life-science/pcr/real-time-pcr/real-time-pcr-learning-center/real-time-pcr-basics/efficiency-real-time-pcr-qpcr.html. Accessed 5 Apr 2021.
19. Buchan B, et al. Distribution of SARS-CoV-2 PCR cycle threshold values provide practical insight into overall and target-specific sensitivity among symptomatic patients. Am J Clin Pathol. 2020;154(4):479–85.
20. Gevertz J, Dunn S, Roth C. Mathematical model of real-time PCR kinetics. Biotechnol Bioeng. 2005;92(3):346–55.
21. Wyllie AL. Saliva or nasopharyngeal swab specimens for detection of SARS-CoV-2. N Engl J Med. 2020;383(13):1283–6.
22. Allen JWL, Verkerke H, Owens J, Saeedi B, Boyer D, Shin S, et al. Serum pooling for rapid expansion of anti-SARS-CoV-2 antibody testing capacity. Transfus Clin Biol. 2021;28(1):51–4.
23. Hughes-Oliver JM, Swallow WH. A two-stage adaptive group-testing procedure for estimating small proportions. J Am Stat Assoc. 1994;89(427):982–93.
24. Dorfman R. The detection of defective members of large populations. Ann Math Stat. 1943;14(4):436–40.
25. Du D, Hwang FK, Hwang F. Combinatorial group testing and its applications, vol. 12. Singapore: World Scientific; 2000.
26. Macula AJ. Probabilistic nonadaptive group testing in the presence of errors and DNA library screening. Ann Comb. 1999;3(1):61–9.
27. Varanasi MK. Group detection for synchronous Gaussian code-division multiple-access channels. IEEE Trans Inf Theory. 1995;41(4):1083–96.
28. Cheraghchi M, Karbasi A, Mohajer S, Saligrama V. Graph-constrained group testing. IEEE Trans Inf Theory. 2012;58(1):248–62.

29. Wu S, Wei S, Wang Y, Vaidyanathan R, Yuan J. Partition information and its transmission over boolean multi-access channels. IEEE Trans Inf Theory. 2014;61(2):1010–27.
30. Bajwa WU, Haupt JD, Sayeed AM, Nowak RD. Joint source–channel communication for distributed estimation in sensor networks. IEEE Trans Inf Theory. 2007;53(10):3629–53.
31. Clifford R, Efremenko K, Porat E, Rothschild A. k-mismatch with don't cares. In: European symposium on algorithms. Berlin: Springer; 2007. p. 151–62.
32. Tsai WT, Chen Y, Cao Z, Bai X, Huang H, Paul R. Testing web services using progressive group testing. In: Advanced workshop on content computing. Berlin: Springer; 2004. p. 314–22.
33. Cormode G, Muthukrishnan S. What's new: finding significant differences in network data streams. IEEE/ACM Trans Netw. 2005;13(6):1219–32.
34. Xuan Y, Shin I, Thai MT, Znati T. Detecting application denial-of-service attacks: a group-testing-based approach. IEEE Trans Parallel Distrib Syst. 2009;21(8):1203–16.
35. Goodrich MT, Atallah MJ, Tamassia R. Indexing information for data forensics. In: International conference on applied cryptography and network security. Berlin: Springer; 2005. p. 206–21.
36. Chan CL, Jaggi S, Saligrama V, Agnihotri S. Non-adaptive group testing: explicit bounds and novel algorithms. IEEE Trans Inf Theory. 2014;60(5):3019–35.
37. Cohen A, Shlezinger N, Solomon A, Eldar YC, Médard M. Multi-level group testing with application to one-shot pooled COVID-19 tests. In: Proc. IEEE ICASSP; 2021.
38. Ben-Knaan EF, Shlezinger N, Eldar YC. Recovery of noisy pooled tests via learned factor graphs with application to COVID-19 testing. In: Proc. IEEE ICASSP; 2022.
39. Ben-Ami R, Klochendler A, Seidel M, Sido T, Gurel-Gurevich O, Yassour M, et al. Large-scale implementation of pooled RNA extraction and RT-PCR for SARS-CoV-2 detection. Clin Microbiol Infect. 2020;26(9):1248–53.
40. Aldridge M, Johnson O, Scarlett J. Group testing: an information theory perspective. Found Trends Commun Inf Theory. 2019;15(3–4):196–392.
41. Sobel M, Groll PA. Group testing to eliminate efficiently all defectives in a binomial sample. Bell Syst Tech J. 1959;38(5):1179–252.
42. Hu M, Hwang F, Wang JK. A boundary problem for group testing. SIAM J Algebraic Discrete Methods. 1981;2(2):81–7.
43. Atia GK, Saligrama V. Boolean compressed sensing and noisy group testing. IEEE Trans Inf Theory. 2012;58(3):1880–901.
44. Kautz W, Singleton R. Nonrandom binary superimposed codes. IEEE Trans Inf Theory. 1964;10(4):363–77.
45. Chan CL, Che PH, Jaggi S, Saligrama V. Non-adaptive probabilistic group testing with noisy measurements: near-optimal bounds with efficient algorithms. In: 2011 49th Annual Allerton conference on communication, control, and computing (Allerton). Piscataway, NJ: IEEE; 2011. p. 1832–9.
46. Aldridge M, Baldassini L, Johnson O. Group testing algorithms: bounds and simulations. IEEE Trans Inf Theory. 2014;60(6):3671–87.
47. Weiss Y, Freeman WT. On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs. IEEE Trans Inf Theory. 2001;47(2):736–44.
48. Kschischang FR, Frey BJ, Loeliger HA. Factor graphs and the sum-product algorithm. IEEE Trans Inf Theory. 2001;47(2):498–519.
49. MacKay DJ. Information theory, inference and learning algorithms. Cambridge: Cambridge University Press; 2003.
50. Shlezinger N, Farsad N, Eldar YC, Goldsmith AJ. Inference from stationary time sequences via learned factor graphs. IEEE Trans Signal Process. 2022.
51. Davenport M, Duarte M, Eldar Y, Kutyniok G. Introduction to compressed sensing. In: Eldar Y, Kutyniok G, editors. Compressed sensing: theory and applications. Cambridge: Cambridge University Press; 2012. p. 1–64.
52. Candes E. The restricted isometry property and its implications for compressive sensing. Comptes Rendus Math. 2008;346(9–10):589–92.

53. Hastie T, Tibshirani R, Wainwright M. Statistical learning with sparsity: the LASSO and generalizations. Boca Raton: CRC Press; 2015.
54. Baraniuk R, Davenport M, DeVore R, Wakin M. A simple proof of the restricted isometry property for random matrices. Constr Approx. 2008;28:253–63.
55. Benatia D, Godefroy R, Lewis J. Estimating COVID-19 prevalence in the United States: a sample selection model approach. medRxiv. 2020;2020:2942. https://doi.org/10.1101/2020.04.20.20072942v1.
56. Shental N, et al. Efficient high throughput SARS-CoV-2 testing to detect asymptomatic carriers. Sci Adv. 2020;6(37):5961.
57. Ghosh S, et al. Tapestry: a single-round smart pooling technique for COVID-19 testing. medRxiv. 2020;2020:727. https://doi.org/10.1101/2020.04.23.20077727v1.
58. Wipf D, Rao BD. Sparse Bayesian learning for basis selection. IEEE Trans Signal Process. 2004;52(8):2153–64.
59. Kueng R, Jung P. Robust nonnegative sparse recovery and the nullspace property of 0/1 measurements. IEEE Trans Inf Theory. 2018;64(2):689–703.
60. Pati Y, Rezaiifar R, Krishnaprasad P. Orthogonal matching pursuit: recursive function approximation with application to wavelet decomposition. In: Asilomar Conference on signals, systems and computing; 1993. p. 40–4.
61. Yaghoobi M, Wu D, Davies M. Fast non-negative orthogonal matching pursuit. IEEE Signal Process Lett. 2015;22(9):1229–33.
62. Heidarzadeh A, Narayanan K. Two-stage adaptive pooling with RT-qPCR for COVID-19 screening. In: ICASSP; 2021.
63. Nida H, Blum S, Zielinski D, Srivastava DA, Elbaum R, Xin Z, et al. Highly efficient de novo mutant identification in a *Sorghum bicolor* TILLING population using the ComSeq approach. Plant J. 2016;86:349–59.
64. Shental N, Amir A, Zuk O. Identification of rare alleles and their carriers using compressed se(que)nsing. Nucleic Acids Res. 2010;38(79):e179.
65. Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J Imaging Sci. 2(1):183–202.
66. Tipping M. Sparse Bayesian learning and the relevance vector machine. J Mach Learn Res. 2001;1:211–44.
67. Crespo Marques E, Maciel N, Naviner L, Cai H, Yang J. A review of sparse recovery algorithms. IEEE Access. 2019;7:1300–22.
68. Goenka R, Cao SJ, Wong CW, Rajwade A, Baron D. Contact tracing information improves the performance of group testing algorithms. 2021. https://arxiv.org/abs/2106.02699.
69. Reed IS, Solomon G. Polynomial codes over certain finite fields. J Soc Ind Appl Math. 1960;8(2):300–4.
70. Naidu RR, Jampana P, Sastry CS. Deterministic compressed sensing matrices: construction via euler squares and applications. IEEE Trans Signal Process. 2016;64(14):3566–75.
71. Naidu RR, Murthy CR. Construction of binary sensing matrices using extremal set theory. IEEE Signal Process Lett. 2017;24(2):211–5.
72. DeVore RA. Deterministic constructions of compressed sensing matrices. J Complex. 2007;23(4):918–25.
73. Candes E, Wakin M. An introduction to compressive sampling. IEEE Signal Process Mag. 2008;25(2):21–30.
74. Bioglio V, Bianchi T, Magli E. On the fly estimation of the sparsity degree in compressed sensing using sparse sensing matrices. In: ICASSP. 2015:3801–5.
75. Ravazzi C, Fosson S, Bianchi T, Magli E. Sparsity estimation from compressive projections via sparse random matrices. EURASIP J Adv Signal Process. 2018;56:578.
76. Center for Disease Control and Prevention. Contact tracing for COVID-19. https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/contact-tracing.html.

77. Case Investigation and Contact Tracing. Part of a multipronged approach to fight the COVID-19 pandemic. 2021. https://www.cdc.gov/coronavirus/2019-ncov/php/principles-contact-tracing.html.
78. Hekmati A, Ramachandran G, Krishnamachari B. CONTAIN: privacy-oriented contact tracing protocols for epidemics. https://arxiv.org/abs/2004.05251.
79. Kleinman R, Merkel C. Digital contact tracing for COVID-19. Can Med Assoc J. 2020;192(24):E653–6.
80. Hohman M, McMaster F, Woodruff SI. Contact tracing for COVID-19: the use of motivational interviewing and the role of social work. Clin Soc Work J. 2021;49(4):419–28.
81. Ross AM, Zerden L, Ruth B, Zelnick J, Cederbaum J. Contact tracing: an opportunity for social work to lead. Soc Work Public Health. 2020;35(7):533–45.
82. Goenka R, Cao SJ, Wong CW, Rajwade A, Baron D. Contact tracing enhances the efficiency of COVID-19 Group Testing. In: ICASSP; 2021; pp. 8168–8172.
83. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. J R Stat Soc Ser B. 2007;68(1):49–67.
84. Jacob L, Obozinski G, Vert JP. Group LASSO with overlap and graph LASSO. In: Proceedings of the 26th Annual International Conference on machine learning. 2009.
85. Nikolopoulos P, Srinivasavaradhan SR, Guo T, Fragouli C, Diggavi S. Group testing for connected communities. In: AISTATS; 2021.
86. Nikolopoulos P, Srinivasavaradhan SR, Guo T, Fragouli C, Diggavi S. Group testing for overlapping communities. In: IEEE International conference on communications 2021; pp. 1–7.
87. Attia M, Chang W, Tandon R. Heterogeneity aware two-stage group testing. IEEE Trans Signal Process. 2021;69:3977–90.
88. Bilder CR, Tebbs JM, Chen P. Informative retesting. J Am Stat Assoc. 2010;105(491):942–55.
89. McMahan CS, Tebbs JM, Bilder CR. Informative Dorfman screening. Biometrics. 2012;68(1):287–96.
90. Deckert A, Barnighausen T, Kyei NN. Simulation of pooled-sample analysis strategies for COVID-19 mass testing. Bull World Health Organ. 2020;98(9):590.
91. Arasli B, Ulukus S. Group testing for overlapping communities. 2021. https://arxiv.org/abs/2101.05792.
92. Ahn S, Chen WN, Ozgur A. Adaptive group testing on networks with community structure. 2021. https://arxiv.org/abs/2101.02405.
93. Lendle SD, Hudgens MG, Qaqish BF. Group testing for case identification with correlated responses. Biometrics. 2012;68:532–40.
94. Lin YJ, Yu CH, Liu TH, Chang CS, Chen WT. Positively correlated samples save pooled testing costs. 2020. https://arxiv.org/abs/2011.09794.
95. List of countries implementing pool testing strategy against COVID-19. https://en.wikipedia.org/wiki/List_of_countries_implementing_pool_testing_strategy_against_COVID-19.
96. IIT-Bombay professor comes up with new tool for Covid testing. It cuts time and cost. 2021. https://theprint.in/health/iit-bombay-professor-comes-up-with-new-tool-for-covid-testing-it-cuts-time-and-cost/684709/.
97. Deka S, Kalita D. Effectiveness of sample pooling strategies for SARS-CoV-2 mass screening by RT-PCR: a scoping review. J Lab Physicians. 2020;12(03):212–8.
98. Gatta VL, Moscato V, Postiglione M, Sperli G. An epidemiological neural network exploiting dynamic graph structured data applied to the COVID-19 outbreak. IEEE Trans Big Data. 2021;7(1):45–55.

Chapter 3
AI for Drug Repurposing in the Pandemic Response

Anh T. L. Truong, Agata Blasiak, Mathias Egermark, and Dean Ho

A. T. L. Truong
The N.1 Institute for Health (N.1), National University of Singapore, Singapore, Singapore

Department of Biomedical Engineering, College of Design and Engineering, National University of Singapore, Singapore, Singapore
e-mail: biettla@nus.edu.sg

A. Blasiak
The N.1 Institute for Health (N.1), National University of Singapore, Singapore, Singapore

Department of Biomedical Engineering, College of Design and Engineering, National University of Singapore, Singapore, Singapore

The Institute for Digital Medicine (WisDM), Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
e-mail: agata.blasiak@nus.edu.sg

M. Egermark (✉)
The N.1 Institute for Health (N.1), National University of Singapore, Singapore, Singapore
e-mail: mathias.egermark@nus.edu.sg

D. Ho
The N.1 Institute for Health (N.1), National University of Singapore, Singapore, Singapore

Department of Biomedical Engineering, College of Design and Engineering, National University of Singapore, Singapore, Singapore

The Institute for Digital Medicine (WisDM), Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

Department of Pharmacology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
e-mail: biedh@nus.edu.sg

N. Lidströmer, Y. C. Eldar (eds.), *Artificial Intelligence in Covid-19*,
https://doi.org/10.1007/978-3-031-08506-2_3

Introduction

The initial outbreak and subsequent emergence of the COVID-19 pandemic triggered a worldwide surge in the research and development of diagnostics, vaccines and therapeutic drugs. The unprecedented speed at which effective and safe products were developed—to diagnose, prevent and treat the new disease—has been an astounding triumph of modern science, engineering and medicine. A critical analysis of both the successes and shortcomings of these efforts can reveal how even higher speeds of development and increased efficacy of treatment might be achieved in the future—to develop further treatments for COVID-19, and to prepare for future epidemics and pandemics. In particular, the deployment of Artificial Intelligence (AI)-based approaches across the continuum of drug development offers unprecedented advantages in terms of *technical flexibility and efficiency* (i.e. speed, resource-efficiency, algorithm adaptability) and *clinical applicability and acceptability* (i.e. scientific rigor, physiological applicability of findings, practical implementation). For this reason, harnessing this enormous potential of AI-based approaches for drug discovery, development and drug repurposing has been a crucial aspect of a strategic response to the COVID-19 pandemic, and will likely remain essential towards effective pandemic preparedness.

Following epidemic and pandemic emergence, the rapid deployment of effective therapeutics is essential while vaccines are developed in parallel. Due to constrained industry operations, stress on healthcare systems, and impacted supply chains that inevitably result from pandemic conditions, drug repurposing has traditionally served as a first option to identify early therapeutic candidates. This was indeed apparent at the onset of the COVID-19 pandemic, where considerable research and medical focus were given to drug repurposing. Already clinically available drugs have the potential to be repurposed against an emerging infectious threat without the need to establish toxicological profiles and conduct pharmacokinetic–pharmacodynamic (PKPD) analyses *de novo*. This approach allows a rapid path to implementation in medical practice. In response to the COVID-19 pandemic, several approaches aimed to use AI to identify drug candidates for repurposing by screening existing data sources: drug libraries, scientific literature, and molecular databases to identify promising mechanisms of targeting SARS-CoV-2, predict binding affinity between drugs and SARS-CoV-2, and simulate interactions between virus, host and drugs in the human and virus interactomes [1–3]. Drug repurposing workflows typically rely on *in vitro* drug sensitivity assays, which yields promising first steps towards further validation in preclinical models and potential clinical studies. However, these approaches have rarely yielded actionable outcomes. In addition, early repurposing efforts have also included attempts to develop drug combinations. However, conventional combination therapy development strategies that are purely mechanism-driven are fraught with limitations introduced when small pools of candidate therapies are used for combination design, among others, which have collectively led to an inability to successfully translate into positive clinical benefit.

These findings illustrate the key difference between drug sensitivity and true optimization, where the systems-level strategies are used to optimally combine therapies at correctly identified doses.

When repurposed drugs serve as the first-line therapeutic intervention against a novel and unknown pathogen, it is important to think beyond single-agent therapy, since repurposed candidates are not developed specifically for the pathogen in question. This is why truly optimizing the design of repurposed combinations may be absolutely essential for an effective initial pandemic response. Combining multiple drugs together is a strategy with potential to further boost the treatment efficacy by tapping into multiple disease and drug mechanisms simultaneously. However, identifying an effective multi-drug combination is not a trivial task even without the unusual circumstances of a pandemic. Importantly, pinpointing the best possible combinations also requires considerations beyond drug synergy, as synergy between compounds alone is unlikely to guarantee a positive clinical response. Conventional screening and combination therapy design methods are commonly based on testing possible combinations one-by-one. For practical reasons, drug doses are usually fixed as an evaluation of n drugs at d doses requires d^n experiments, which means that with as little as four dose levels for 12 different drugs, over 16 million ($= 4^{12}$) experiments would be needed. As a result, some of the combinations that could have been effective at different dose ratios, are discarded without testing. AI is uniquely suited for supporting an approach that identifies optimal combinations of both drugs and dose levels in a high-dimensional interaction space. The AI-based, mechanism- and disease-agnostic platform IDentif.AI has shown promise in rapidly identifying optimized combination therapies that could be applied to new strains of SARS-CoV-2 and other pathogens [4–6].

The use of AI to identify drugs and drug combinations for repurposing bears the promise of exceedingly rapid development and better risk/benefit profiles of therapeutics for COVID-19 and possible future infectious threats. In this chapter, a selection of AI-based approaches that showcase the potential of AI for drug repurposing in COVID-19 are described, and the IDentif.AI platform is reviewed as an applied example in this field. The role of Real-World Data (RWD) for use in these platforms is also discussed. Lastly, an outline of future possible directions in preparing for rapid response to future variants and new pathogens is provided.

Desirable Features of AI for Drug Repurposing in Pandemic Response

Shortly after the announcement of a novel coronavirus in January 2020 by the World Health Organization (WHO), extensive research efforts to discover and develop drugs against COVID-19 were promptly set in motion globally. The traditional drug discovery and development process of novel drugs often requires a long time. As

such, drug repurposing presented a vital strategy as the first response to the disease outbreak while research on the discovery of novel compounds was still underway. To date, a variety of computational approaches have been deployed to expedite drug repurposing efforts, such as deep learning-based knowledge mining, network-based analyses and *in silico* modelling. These techniques can be used to screen large drug databases containing up to thousands of therapeutic drugs and search vast troves of medical literature within days [1, 3, 5, 7–9].

To fully evaluate the suitability of AI for drug repurposing, it is important to first understand key desirable features that allow AI applications to meaningfully support a rapid and sustained response in a pandemic context. In this section, we categorize these desired features into two groups: *technical flexibility and efficiency; clinical applicability and acceptability.*

Technical Flexibility and Efficiency

Flexibility and efficiency, in general, refer to the technical ability of the AI platform or AI-based methodology to be applied for timely, resource-efficient and adaptable usage at different stages of a pandemic. The need for such flexibility and efficiency arises, in part, because of the rapid pace with which infectious pathogens spread and mutate, while overburdening medical infrastructures and causing precipitous damage to economies and societies. Given the time criticality of a pandemic—and hence, the shortened timeline for effectively identifying therapeutic interventions—*speed* is of essence. To this end, AI-based approaches can play a crucial role in rapidly screening large drug databases and real-world clinical data to identify potential therapeutic drugs quickly and efficiently.

Moreover, the sudden onset of disease outbreaks inevitably leads to severe resource constraints and overburdening of biomedical research infrastructure. AI applications therefore need to be characterized by *efficiency*, both in terms of low resource requirements for their implementation, as well as in their ability to improve resource utilization in the overall drug development process. In particular, pandemic responses must contend with *resource constraints*—such as limited access to the virus, cells and appropriate laboratory infrastructure [e.g. Biosafety Level 3 (BSL-3) laboratories that are required for experiments with air-borne infectious pathogens, such as SARS-CoV-2]. AI applications should therefore help alleviate bottlenecks by identifying and prioritizing promising candidates for further experimental validation, thereby helping to ensure that scarce laboratory resources are prioritized for promising leads.

However, even as AI applications can expedite drug repurposing efforts and optimize resource utilization, it is important to ensure that they do not inadvertently create new bottlenecks and operational challenges. It is therefore crucial to ensure that, wherever possible, computational infrastructure, complexity and cost is minimized, so that applications are easy to run, test, and maintain. Moreover, these applications need to be accompanied with clear, straight-forward operational

protocols, so that operators can efficiently make use of them with minimal training. These *technical requirements*—in terms of software, hardware, learning processes and operational protocols—are crucial for ensuring that AI applications are easily accessible and require as low resource utilization as possible in terms of human resources and energy consumption.

Finally, as seen with COVID-19, pandemic responses are dynamic processes due to the constant mutations of pathogens and the quickly evolving knowledge about drugs, targets and pathways. Effective responses must stay abreast of these dynamic developments, to ensure that new variants do not compromise the efficacy of treatments, and that the most up-to-date knowledge about the pandemic is being included. For this reason, the *ability to re-learn*—to continuously respond to evolving information by being able to easily incorporate new knowledge about drugs, targets and pathways—is a key consideration for the effective use of AI-based models in pandemic drug development.

Clinical Applicability and Acceptability

This category of desired features focuses on the *clinical applicability and practicality* of AI applications to facilitate in the discovery of clinically deployable therapeutic drug solutions. Here, one key consideration is to ensure that AI system's findings are delivered to clinical practitioners in a manner that inspires confidence, trust and willingness to include them in a clinical setting. As such, to ensure that AI applications are trustworthy and accepted, their *technical performance* must be consistently reliable, and their findings should be presented with a scientific rigor including clear explanations and well-reasoned justifications.

Moreover, many currently pursued AI applications for drug repurposing produce predictions based on literature mining, experimental datasets on human and virus interactome, and molecular interactions. As a result, such applications often face a translational gap. To ensure their clinical applicability, it is important for users of AI models to incorporate *physiological/biological applicability* of the proposed treatment options. This includes ensuring that drug candidates proposed for clinical use are explainable with regard to the *proposed mechanism of action*, not just reliant on "black-box" predictions, in order to be accepted by the clinician community. It also needs to be clear that drug candidates identified via AI applications still need to undergo prospective experimental validation in clinical trials before wide clinical adoption. For repurposing of drugs already in clinical use, and for monitoring of new drugs in clinical practice, RWD analysis may be employed to verify safety and efficacy profiles.

Finally, another key gap between AI applications and the clinical implementation of their findings centers on *practical and public health considerations*. It includes consideration of whether proposed treatments are practically possible to distribute and administer in the intended care setting, taking into account administration form and other pharmaceutical aspects. For example, developers need to ask if proposed

doses of repurposed drugs are available especially in the face of inevitable supply chain disruption, are the costs of the drugs feasible or prohibitively expensive, can more resource-demanding forms of delivery such as intravenous administration be performed in the intended setting or is only peroral administration possible? To overcome such challenges, the use of AI applications should also consider the social, economic, and regulatory contexts that might hinder the clinical implementation of their findings.

Major AI Applications for Drug Repurposing in Response to COVID-19

Knowledge Mining

Knowledge mining approaches leverage on publicly available biomedical literature and databases to construct comprehensive medical knowledge graphs containing different entities and their relationships, where entities may be drugs, diseases, proteins, genes, pathways and are represented as nodes or vertices, and relationships may be drug–host interaction, drug–target interactions, protein–protein interactions which are represented as edges or links [10]. In this field, AI plays a big role, as a variety of different AI algorithms are often incorporated to facilitate and speed up the knowledge mining process. A widely published application of AI-based knowledge mining during the COVID-19 pandemic is the BenevolentAI's knowledge graph which assembles multiple networks of structured medical information with machine learning support [11] to identify drug target and potential therapies against SARS-CoV-2. The BenevolentAI platform predicted baricitinib, a drug approved for clinical use in rheumatoid arthritis treatment, to be able to modulate immune system response and to have anti-viral effect in patients with COVID-19 [3, 12]. CoV-KGE is another AI-based knowledge graph developed to speed up drug repurposing efforts specifically for COVID-19. In 2020 it was reported to encompass 15 million edges across 39 types of relationships between drugs, diseases, proteins, genes, pathways and expressions, onto which a deep learning graph representation technique was employed to identify drugs with potential for treating COVID-19 [13]. Of 41 drug candidates identified by CoV-KGE, 26 drugs (e.g. dexamethasone, melatonin) have entered prospective clinical trials [14].

A key advantage of knowledge graph applications are their *technical flexibility and efficiency* in extracting and structuring substantial amounts of publicly available scientific literature and data in order to generate new hypotheses within a short period of time. Analysis of existing data on drugs, drug–target interactions, drug–disease interactions and disease pathology of previously known coronaviruses (e.g. SARS-CoV-1, MERS-CoV) helped identify therapeutic drug candidates for

SARS-CoV-2 even in the early days of the outbreak. For instance, BenevolentAI proposed baricitinib in a Correspondence to The Lancet in early February 2020—just 1 month after WHO released a statement on the discovery of SARS-CoV-2 outbreak in China [3]. In addition to the *speed* advantage, applying modern AI techniques to comprehensive knowledge graphs [11, 13] enables systematic screening of large libraries of approved drugs, clinically investigational drugs and/or experimental drugs (BenevolentAI: 378 drugs, CoV-KGE: 3481 drugs) in a *faster and more resource-efficient* manner than traditional experimental high-throughput screening. Thus, going forward, it would be valuable to position knowledge mining applications at the frontline of efforts to prioritize drug candidates for further testing based on mechanistic perspective, especially in cases where understanding of emerging new diseases is still limited. Additionally, knowledge mining algorithms are *able to re-learn* and incorporate new disease-specific evidence as understanding of the disease increases and new datasets are published throughout the pandemic timeline to improve the performance of the approach. For example, Ge et al. [15] observed an improvement in the prediction results after updating their knowledge graph with a newly published dataset of human protein–virus protein interactions. The CoV-KGE team sought to validate their predictions against clinical evidence from ongoing COVID-19 trials and transcriptomic and proteomics data in SARS-CoV-1/2-infected human cell lines in consolidating the final list of 41 drug candidates for repurposing [13].

However, knowledge mining applications are not without limitations. Although these approaches can optimize resource utilization, they may incur high costs and bottlenecks arising from the resource requirements for data management (e.g. data collection, data access and data transfer), and computational operation (e.g. AI computation, data cleaning and filtering). For instance, CoV-KGE uses DGL-KE—a high-performance but expensive deep learning package [13]. Consequently, the team had to operate with a limited quality graph to maintain a low cost while keeping the screening process expedient.

In terms of *clinical applicability and acceptability*, comprehensive knowledge graphs that incorporate pre-clinical and clinical evidence (including COVID-19) indicating the *physiological and biological applicability* of predicted drugs for use in patients may facilitate a faster translation into clinical practice. Shortly after identifying baricitinib using the BenevolentAI platform in early 2020, Stebbing et al. [16] demonstrated its dual anti-viral and anti-inflammatory efficacy in *in vitro* and *in vivo* SARS-CoV-2 models and a case series of 4 patients with bilateral COVID-19 pneumonia. These preliminary findings led to initiation of larger randomized controlled trials (RCTs) such as the Adaptive COVID-19 Treatment Trial 2 (ACTT-2; NCT04401579) [17] and CoV-BARRIER (NCT04421027) [18] to test the safety and efficacy of baricitinib. Baricitinib received the first emergency approval for use in combination with remdesivir in November 2020, and as a monotherapy in December 2021.

Network-Based Analysis

Network-based analysis in drug repurposing uses a variety of network-based algorithms and integrates systems pharmacology with network science to create models of human protein interactomes that represent the molecular networks of the virus–host interactions, protein–protein interactions (PPIs), functional associates and drug–target interactions [8, 10]. Drug candidates are identified based on the extent of their interactions with the protein targets and potential effect on the disease module. Several network-based strategies falling under the broad definition of AI [19, 20] have been applied to identify drug repurposing candidates for COVID-19. Network proximity-based approach predicts drug effect based on the distance between the viral protein target and target proteins of drugs within human interactome [9, 21, 22]. Diffusion-based methods predict drugs based on similarity between the drug targets and viral protein targets [8]. Artificial neural networks such as graph convolutional network [8] and autoencoders [7] can combine multiple data sources to predict drug efficacy. A multimodal workflow of several different network-based strategies can also be combined to cross-validate the prediction results and consolidate a reliable drug list [8]. Also, Zhou et al. [9] used network analysis to identify potential two-drug combinations based on identifying the shortest distance between the protein targets of two drugs in the 'Complementary Exposure' pattern.

The strategy of network analysis for drug repurposing is potentially a powerful approach owing to its *flexible* theory that drugs targeting one disease might work in another in a similar functional PPI network. In the work in early 2020 by Zhou et al. [9], prior knowledge of molecular interactions between drug targets and closely related known pathogens (e.g. SARS-CoV, MERS-CoV, IBV, MHV, HCoV-229E, and HCoV-NL63) was used to quickly identify 16 potential drug candidates, of which 9 drugs are currently tested in clinical trials for COVID-19 treatment [14].

However, although network-based approaches have high flexibility in incorporating data from different domains, their *technical efficiency* may be limited by several factors. First, these approaches are highly reliant on high data availability, and a potential limitation is therefore the lack of the high-quality public datasets needed to construct reliable networks. Consistency and quality of data are also often limited by various factors, including data completeness, differences in experimental assays and cell lines across different studies. Therefore, substantial effort and coordination across laboratories worldwide would be required to generate high-quality experimental datasets and build standardized data resources suitable for comprehensive network-based approaches. Additionally, updating the networks with sufficient experimentally derived SARS-CoV-2-specific omics data is essential to facilitate the *re-learning* process of the algorithms in order to continuously improve the prediction quality throughout the pandemic timeline. However, keeping the databases up-to-date with continuous generation of high-quality experimental SARS-CoV-2 data for network deployment requires time and resources that may create operational challenges and *inefficiency* given the already severe resource constraints worldwide.

In evaluating the *clinical applicability and acceptability* of network-based applications, it is evident that the *physiological/biological applicability* of the prediction results was taken into account by many network-based studies. As the interactome networks assemble experimentally derived PPI data (including for COVID-19), predicted drugs with existing *in vitro* evidence for COVID-19 can be directly tested in subsequent steps such as *in vivo* and human clinical trials. Several studies have also aimed to improve the *applicability* of the outcomes by incorporating experimental validation in various human cell lines [8] or validation against RWD from COVID-19 clinical trial registry [23]. Furthermore, network-based applications may lead towards promising drug candidates, but clinical evidence of an optimal use of these candidates (e.g. co-administration with other candidates at suitable respective doses, dose scheduling, etc.) remains an essential aspect to properly advance their development. Importantly, the clinical validation of a single agent opens the doors to even further enhancing the application of this agent through continued combinatorial optimization. Synergistic drug combinations could be more effective than single-agent therapies while requiring lower drug doses, and therefore, offering a potentially safer, less toxic, and more preferred option by the clinical community for drug repurposing [24]. Considering the *clinical applicability and acceptability* of findings on combinatorial therapies, Zhou et al. [9] also aimed to search for efficacious drug combinations using the abovementioned 'Complementary Explosure' pattern to identify potential synergistic drugs that target the same virus–host network but in separate neighborhoods in the human interactome.

Lastly, it should be kept in mind that network-based approaches should not be regarded as exhaustive when it comes to identifying drug repurposing candidates. Despite the many advantages of network-based analyses, the *clinical applicability* of their results is limited by the scope of the networks used. The types of drugs predicted may be limited to drugs with certain modes of interactions represented in the network (e.g. drug–host interaction only, drug–virus interaction only). Direct virus-targeting drugs such as remdesivir were not identified by drug–host interaction network [9]. In addition, the type of experimental models used to construct the network may also influence the drug ranking or restrict the types of drugs that can be screened. For example, loratadine was not ranked highly in a study by Gysi et al. [8] which was based on an *in vitro* experimental model with African green monkey kidney Vero E6 cells, but showed antiviral effect in human intestinal epithelial Caco-2 cell line [25]. Furthermore, dexamethasone and hydrocortisone, which do not have antiviral effects, were shown to reduce cytokine storm and achieve a lower 28-day mortality in COVID-19 patients [26], but were not identified by networks screening for drugs with antiviral interactions [9]. This experience points to network analysis as a useful approach for identifying promising drug repurposing targets, but not for ruling out candidates. Furthermore, as with all model-based approaches to drug repurposing, drug candidates identified by networks analysis, must still undergo further prospective testing in COVID-19 patients, especially if it based on little or no SARS-CoV-2-specific data.

In Silico *Modelling*

In silico methods rely on predicting the drug–target protein interactions and binding affinities of chemical compounds to target viral proteins at a molecular level to identify potential drugs. Conventional *in silico* techniques include structure-based drug screening, such as virtual drug screening [27] and molecular docking [28] that utilize three-dimensional (3D) protein structures and chemical structures of the drugs to predict the affinities of the drugs to the target protein. Notably, in the COVID-19 pandemic context, AI has significantly contributed to accelerating the *in silico* screening process while overcoming the common limitations of traditional docking-based applications. Incorporating machine learning algorithms (e.g. random forest regression, [29]; Naïve Bayes classification, [30]) and deep learning algorithms (e.g. convolutional neural network and recurrent neural network, [1]; bi-directional long short term memory, [31]) in *in silico* workflows minimized the need for complex modelling of 3D protein structures in predicting the drug–target protein interactions. For instance, Batra et al. [29] trained random forest regression models using high-fidelity expensive docking studies, which were subsequently used to screen much larger chemical databases up to thousands of FDA-approved ligands. Mohapatra et al. [30] utilized Naïve Bayes classifier to prioritize drugs prior to validation with costly docking simulation. Deep learning-based models, such as Molecular Transformer-Drug Target Interaction (MT-DTI; [1]) and DeepDTA [31], eliminated the need for 3D-structural data and only relied on non-structural information such as chemical sequences (SMILES) and amino acid sequences (FASTA) of a target protein to predict binding affinities of drugs to target viral protein.

In silico modelling serves as a promising approach for drug repurposing given its *technical flexibility and efficiency* to *rapidly* screen large and diverse drug libraries up to thousands of therapeutic drugs [1, 31]. Within the first few months of 2020, several studies managed to deliver a list of predicted drugs, some of which are being tested in clinical trials (e.g. atazanavir, sirolimus) or are already authorized for emergency use (e.g. remdesivir). As these deep learning-based models investigate mechanistic drug–target interactions at a molecular level and utilize pre-trained AI models, they can be used to identify drugs based on their predicted binding interactions with a target viral protein (e.g. $3CL^{pro}$, RdRp, PL^{pro}) without the need for prior domain knowledge. On the other hand, traditional structure-based screening requires a large amount of high-quality 3D structural information of drug molecules and viral proteins to train in docking simulation, which often creates operational bottlenecks especially when 3D-structural data may not be available for new viral threats with completely uncharacterized protein targets. Deep learning-based methodologies are able to overcome this technical limitation by using non-structural data in

building drug-interaction models [1, 31] or through an optimized use of docking application in conjunction with AI algorithms [29, 30]. In addition, as opposed to the traditional experimental high-throughput screening, these AI-assisted methodologies do not require access to laboratory facilities and resources to identify drug candidates. Thus, they may be useful as the first response to an unexpected disease outbreak and can help identify and prioritize promising candidates in a way that *optimizes resource utilization*. In addition, *in silico* workflows are adaptable and can also be applied throughout a pandemic timeline by expanding the drug search space and *re-learning* when new viral protein targets are identified.

A potential limitation of *in silico* approaches is the lack of consideration for their *clinical applicability and acceptability*. Published *in silico* studies often focus on reporting their technical methodologies and performance accuracy instead of identifying drugs for repurposing [14]. Drug candidates predicted across studies vary substantially as the studies use different evaluation benchmarks to rank the drugs. As such, prediction results cannot be compared across studies, and there is a clear need to standardize *technical performance* metrics, to allow for cross-examination across studies to increase confidence for the next step of validation of *in silico* predictions. Additionally, these approaches for identifying drugs based on binding interactions with proteins are also limited by their focus on binding mechanism to viral/host proteins. As an example of how this may limit the search for promising targets, one study found that 76 out of 77 drugs identified by a network-based methodology did not have evident binding mechanisms, and therefore could not have been identified using *in silico* docking-based approaches [8].

Furthermore, while molecular interactions can be replicated in *in silico* applications, these approaches are limited in their ability to replicate the physiological conditions surrounding the complex human and virus interactomes [32, 33]. Therefore, *in silico* predictions still require multiple validation steps (*in vitro*, *in vivo* and clinical) before being deployed into the clinical use. Additionally, most studies did not take into account the practical applicability of the drugs (e.g. side effects, administration routes, regulatory, drug access, costs) in suggesting the final drug lists for further validation. In the context of the COVID-19 pandemic, *in silico* predictions were seldom included in selecting drugs for clinical trial testing, highlighting the translational gap between this type of AI technology and its application in clinic use.

In summary, the deployment of various AI platforms in the COVID-19 pandemic has revealed the unprecedented benefits of AI applications for drug repurposing, while also illuminating a range of important considerations that need to be addressed for clinically successful applications of AI. Understanding the strengths and limitations of each of these different platforms is crucial in formulating a holistic strategy to effectively deploy AI at various points along the pandemic timeline.

IDentif.AI Platform for Rapid Identification of Drug Combinations

As discussed above, several AI-based approaches for identifying drug repurposing candidates have been applied and published in the context of COVID-19. In addition to these efforts to identify suitable single-agent therapies, there is a need to identify ways to optimize for potential efficacious drug combinations. Finding the optimal drug combination for a specific indication requires selecting not only the drugs but also their doses in relation to each other to uncover and benefit from unpredictable, dose-dependent drug–drug interactions. With traditional or brute force methods, investigating drug/dose interaction space one-by-one is an insurmountable task. The circumstances of a pandemic—when speed, reagents and laboratory access are of tremendous value—call for a more experimentally prudent approach that can inform about the drug–drug effects in the whole interaction space with low number of experiments.

IDentif.AI is a small data platform that combines carefully designed live cell infection experiments and regression analysis to identify optimal drug treatments among thousands of possible multidrug combinations. At the core of the IDentif.AI operation is an observation that a complex biological system can be described by a smooth quadratic surface with the drug dose input and phenotypic outputs serving as the variables behind that surface, and that surface being clinically applicable [34–46]. This observation set the foundation for a selection of a specific network of data points in the multidimensional drug and dose interaction space for experimental testing. The network of data points can be identified via orthogonal array design (OAD), or OAD combined with a factorial design (FD) to form orthogonal array composite design (OACD) [47]. OAD or OACD selected data points together with the quadratic input–output relationship allow inference about the whole interaction space based on a small experimental set and therefore significantly lower the number of experimental combinations for testing. This property made it particularly suitable for personalizing drug combinations for cancer patient as applied by a quadratic phenotypic optimization platform (QPOP) already validated in cancer patients [48, 49]. In the infectious disease field, this approach was used to identify an optimal drug combination to treat tuberculosis [50–52] and human immunodeficiency virus (HIV; clinical validation) [53], among others.

Project IDentif.AI

IDentif.AI as a pandemic preparedness platform was first demonstrated in Project IDentif.AI study performed in the extra-ordinary circumstances of limited laboratory access and time pressure of the early pandemic days. Abdulla and co-authors demonstrated that it is possible to interrogate a prohibitively large drug-dose parameter interaction space of 12 drug candidates against a virus in as little time as three

days [4]. A549 human lung cell line infected with vesicular stomatitis virus (VSV), was used as the model for experiments. Viral infection rate after 12 h of infection served as the phenotypic output while the drug and dose of the treatments co-administered with the virus served as the inputs. The 12 drugs were selected based on the need to achieve *clinical applicability and acceptability* of the approach, taking into account their safety profile, potential of efficacy either based on prior evidence against other viral pathogens or their hypothesized mechanism-of-action, tolerability by broad patient groups, access, and potential for adoption into clinical practice.

The experimental procedures comprised four steps (Fig. 3.1), with carefully scheduled time points for cell plating, drug(s) and/or virus additions, incubation, and experimental readouts.

The first step optimized viral infection conditions [cell plating density and the amount of virus added reflected by multiplicity of infection (MOI)] to result in a compact experimental infection model where majority of the cells are infected after 12 h of incubation with the virus. The second step comprised testing the toxicity and efficacy of increasing concentrations of single drugs. The monotherapy testing results were the basis for selecting drug levels in the network of data points—drug combinations—for further testing. The third step experimentally tested 72 drug combinations and revealed a wide range of their antiviral properties. The analysis based on a quadratic relationship between the drugs, their doses and infection rate, together with STRICT algorithm analysis to identify quantitative drug interactions [54] pinpointed 5 drugs of high antiviral interaction potential for validation. The final step validated the interaction between the 5 drugs with experimental testing of 30 combinations as designed with the OAD.

The analysis revealed the drug and dose dependence of the combinations' antiviral properties. The most optimal drug–dose combination comprised of 4 drugs and lowered the viral infection rate to 1.5% of the cells. Changing the concentration of one of the drugs lowered the combination's efficacy to 10.7% cell infection rate, while changing the drugs' concentrations and swapping one of the drugs for the fifth drug in the pool lowered the combination's efficacy and resulted in 21.6% cell infection rate. The optimal combination could have been simply overlooked if the concentration range was not considered. Additionally, the analysis allowed to identify smooth surface interactions between three drug pairs that may serve as mechanistic insights and can be further investigated for their clinical potential.

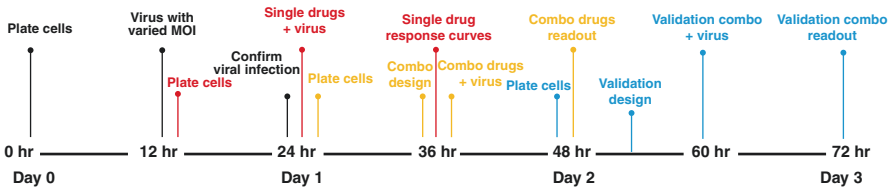


Fig. 3.1 Project IDentif.AI experimental schedule demonstrates a compact workflow that achieves data collection, IDentif.AI analysis and experimental validation within 3 days. (Adapted with no change from [4])

IDentif.AI for Drug Optimization Against SARS-CoV-2

Shortly after Project IDentif.AI demonstrated the potential for robust optimization of combinations from repurposed drugs against a viral infection in the VSV model [4], the IDentif.AI platform was used to identify an optimal drug combination against SARS-CoV-2 [5]. In the presence of only fractional clinical data at the time, 11 of the 12 drugs were included in the study based on their hypothesized anti SARS-CoV-2 properties by affecting a range of pathways: viral entry, viral replication, viral RNA-synthesis, and viral release. The 12th drug, dexamethasone, was added to identify combinations that were aligned with the clinical guidelines at the time that recommended dexamethasone as a therapy for COVID-19-induced acute respiratory distress syndrome. African green monkey kidney Vero E6 cell line infected for 72 h with SARS-CoV-2 isolated from a local patient was used as the model for the efficacy experiments. Drug safety was additionally measured in THLE-2 human liver and AC16 human cardiomyocyte cell lines, also after 72 h incubation.

The study lasted 2 weeks and had 4 steps: (1) drug-dose selection, (2) experimental drug combination treatments, (3) IDentif.AI analysis, and (4) experimental validation of the analysis results (Fig. 3.2).

The first step aimed to identify two concentration levels of each drug to be included in the combinatorial experimentation. The concentration levels were identified based on the experimental efficacy and safety of the monotherapies as well as clinical actionability. The drugs were co-administered with the virus, and the virus'

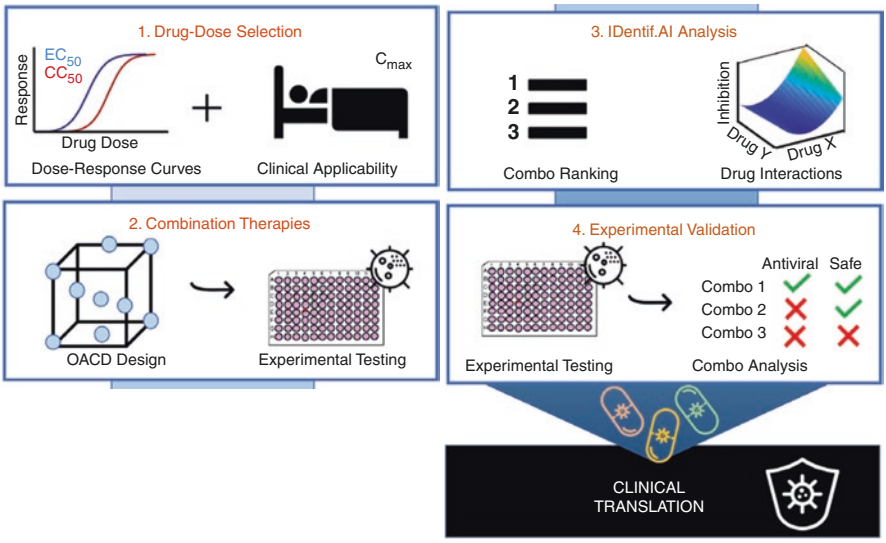


Fig. 3.2 IDentif.AI workflow comprising of two experimental steps with live cell infection model to prospectively collect the data, IDentif.AI analysis step, and an experimental validation step. The whole workflow was performed in 2 weeks. (Adapted from [5])

cytopathic effect (CPE) after 72 h of incubation was used as the phenotypic output. 10% and 20% absolute effective concentration (EC_{10} and EC_{20}) for each drug were selected. Those low concentration levels ensured that none of the drugs was over-represented and overweighted the surface in the combinatorial results analysis. The EC_{10} and EC_{20} values were then compared against maximum plasma concentration in human studies (C_{max}) to avoid using drug concentrations unrealistically high to be physiologically achievable in patients. The study used 10% C_{max} as a universal cut off drug concentration of what the target cells may experience in a human body. This step of the IDentif.AI study stems from the characteristics of repurposed drugs: already tested and known PK like C_{max} and plausibly lower efficacy against a new indication than the efficacy against the original indication the drug was developed for. Inputs that are by design conducive to the result's physiological actionability are critical in ensuring that the AI application yields data of the highest downstream usability in a clinical practice. Accordingly, out of the pool of 12 drugs only one drug, remdesivir (RDV), achieved EC_{20} below the 10% C_{max} cut off. The selected concentrations for combinatorial testing of the remaining drugs were derived from their C_{max} instead of their efficacy data.

In the second step, 100 drug combinations were designed using OACD to probe the 12 drug–dose interaction space. The CPE results of testing the drug combinations were used for multiparameter IDentif.AI analysis. The experimental data—drugs and their doses as inputs and normalized CPE Inhibition as the output—analyzed with a quadratic series revealed a comparative ranking of drug combinations according to their efficacy. The drug combinations comprising of 4 drugs or less accounted for 9968 drug combinations. Among those, IDentif.AI analysis pointed at unforeseen interaction between remdesivir and lopinavir (RDV/LPV) which formed the top two-drug combination (Fig. 3.3). The combination of RDV/LPV co-administered with ritonavir (RDV/LPV/RTV) was the top three-drug combination and was present in all top ten four-drug combinations. RDV/LPV and RDV/LPV/RTV were also higher ranked than RDV on its own. The IDentif.AI analysis ranked LPV/RTV at place 1261 and hydroxychloroquine combined with azithromycin at place 5161—the results that are aligned with suboptimal outcomes of clinical validation of those drug combinations.

Due to its adaptability, IDentif.AI also allows to infer about the unforeseen interactions leading to increased cytotoxicity of the drugs. Drugs were added to the cells without the virus according to the same OACD-identified 100 drug combinations. Drugs and their doses served as inputs, while cytotoxicity (based on cell CPE) was used as the output. The experimental results and the following IDentif.AI analysis indicated that adding additional drugs did not increase cytotoxicity in Vero E6 nor in AC16 cells, but increased cytotoxicity in THLE-2 cells. Interestingly, adding dexamethasone was predicted by the IDentif.AI analysis to lower the cytotoxicity induced by RDV/LPV/RTV in THLE-2 cells.

In the final phase, the IDentif.AI analysis results were experimentally validated. The selected 9 drugs in combination and in monotherapies were added to the Vero E6 cells (with and without co-addition of the virus) and to the AC16 and THLE-2 cells. The *in vitro* experimental results demonstrated a 6.5x increase in efficacy of

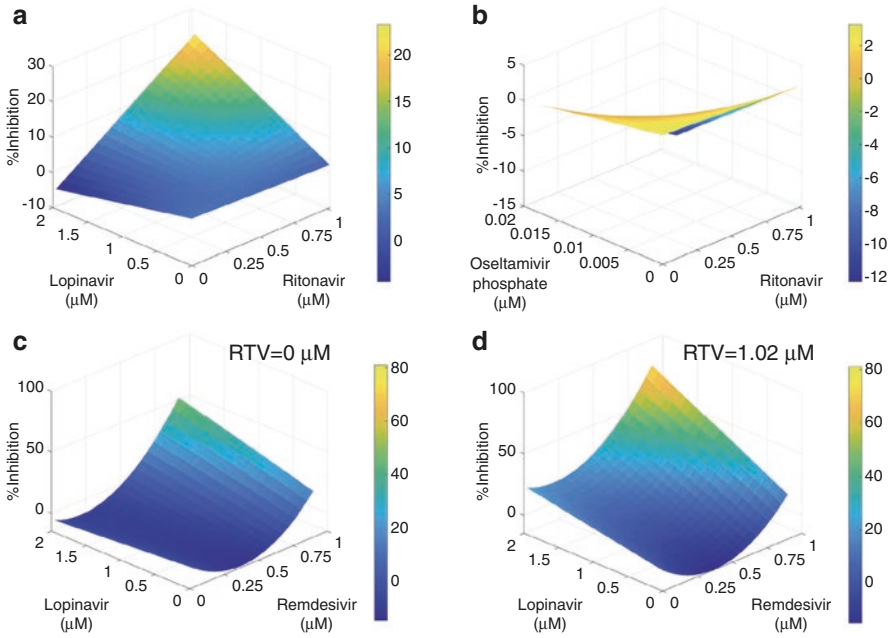


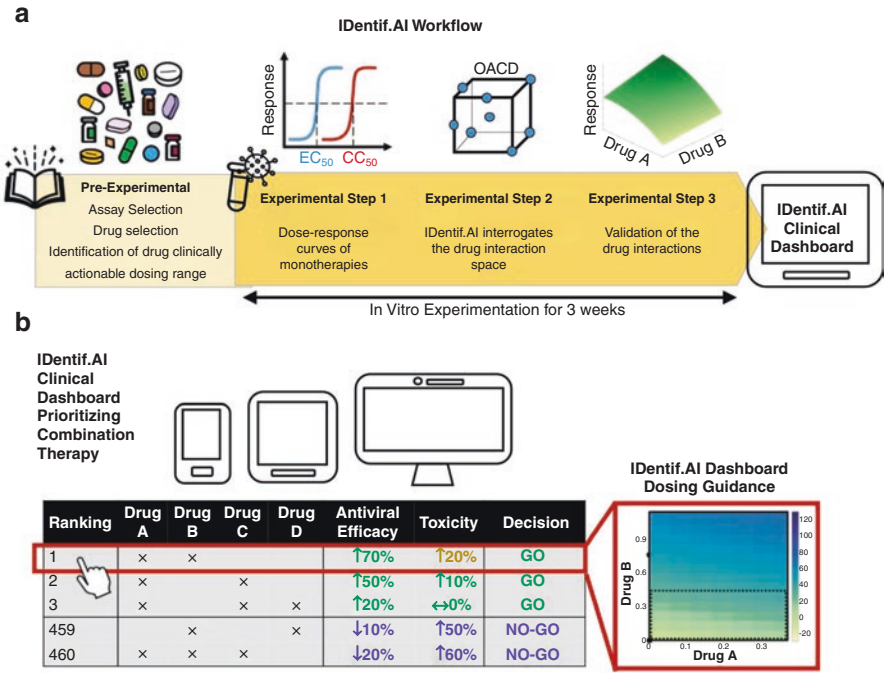
Fig. 3.3 Drug interaction surfaces demonstrating combined efficacy against SARS-CoV-2 generated by IDentif.AI in step 3 of the workflow. (Adapted from [5])

RDV/LPV/RTV when directly compared to the efficacy of RDV alone. RDV/LPV/RTV lead to complete viral inhibition at clinically actionable doses. Direct measurements in the validation set also revealed that when provided to the cells at the clinically actionable doses none of the four regimens considered to be strong clinical candidates at the time exceeded 20% viral Inhibition. Those regimens were independently found to fail the clinical trials.

IDentif.AI 2.0 Platform in an Evolving Pandemic

IDentif.AI platform is well suited in terms of its *ability to re-learn* and include new circumstances to evolve with the evolving pandemic. The platform includes updated data that are generated prospectively by performing a new set of experimental using the latest pathogen variant(s) and the drug candidates based on the latest knowledge.

In early Summer 2021, the IDenif.AI platform—called IDentif.AI 2.0—was modified to the pandemic context of the time and used to identify drug combinations effective against the emerging SARS-CoV-2 variants of concern: B.1.351 (Beta) and B.1.617.2 (Delta) [6]. Like the IDentif.AI study against SARS-CoV-2 from 2020, IDentif.AI 2.0 had 4 steps (Fig. 3.4a) and used Vero E6 cell line infected for 72 h with SARS-CoV-2 isolated from a local patient as the model for the efficacy experiments.



Ranking	Drug A	Drug B	Drug C	Drug D	Antiviral Efficacy	Toxicity	Decision
1	×	×			↑70%	↑20%	GO
2	×		×		↑50%	↑10%	GO
3	×		×	×	↑20%	↔0%	GO
459		×		×	↓10%	↑50%	NO-GO
460	×	×	×		↓20%	↑60%	NO-GO

Fig. 3.4 IDentif.AI 2.0 workflow (**a**) leads to a ranked list of combinations, which can be presented as a clinical dashboard for prioritising drug combinations for further testing and potential clinical deployment (b) IDentif.AI - Clinical dashboard prioritizing combination therapy. (Adapted from [6])

The drugs toxicities were additionally measured in THLE-2 and AC16 human cell lines at the validation stage. The initial 12 drug screening pool included 9 new drugs and RDV, LPV and RTV for a comparison with the 2020 study. IDentif.AI 2.0's adaptability facilitated narrowing the drug screening pool to 6 drugs to align with the laboratory guidelines while minimizing experimental variation. Within 3 weeks the study identified EIDD-1931 (the active substance after metabolism of molnupiravir) to have a potential for becoming a backbone of a range of combinatorial anti-SARS-CoV-2 therapies. IDentif.AI 2.0 was adapted to a specific pandemic and resource context and generated ranked list of combinations envisioned to support decision on which combination therapies to prioritize for further testing (Fig. 3.4b).

IDentif.AI as a Pandemic Preparedness Platform

The IDentif.AI approach for drug discovery has several characteristics that ensure high *technical flexibility* and make it particularly suitable as a pandemic preparedness platform. In the three described examples above, the IDentif.AI workflow took

between 3 days to 3 weeks [4–6]. This is a staggering speed of interrogating an interaction space that for 12 drugs consists of 530,000 drug–dose combinations. An additional advantage is lowering the requirement of resources and workload as the pandemic onset may limit access to resources, laboratory personnel and even laboratory space. IDentif.AI is indication- and mechanism-agnostic. This means that it can be adapted to new infectious disease threats in the form of new pathogen variants or altogether new pathogens. IDentif.AI can also be performed with limited knowledge of the threat typical for the early days from the pandemic onset—it is not reliant on a detailed understanding of the action mechanism of the pathogen or the candidate drugs. It is also able to incorporate new potential drug candidates and new pathogen variants as they emerge over time. Additionally, as all the data inputs for the analysis are prospectively generated, IDentif.AI allows control over data quality and therefore avoids some of the common biases related to data source.

IDentif.AI is not without limitations. As all data come from an *in vitro* experimental model, the results of the analysis should undergo a clinical validation step before broad deployment. Additionally, the selected *in vitro* experimental model limits the detectable interactions to those affecting the pathways present in the model. For example, the Vero E6 cell line used in IDentif.AI and IDentif.AI 2.0 is unable to reflect the effects of immunomodulation that has been demonstrated to alleviate severe COVID-19 symptoms in clinical practice. Similarly, the results may differ depending on the timepoint for adding virus and the drugs, incubation time, and cell culturing conditions. Understanding the limitations of IDentif.AI is necessary for placing it in the right place within the drug discovery workflow—prioritization of the combination therapies for further testing. Nevertheless, it is worth noting that IDentif.AI has independently replicated clinically established drug–drug interactions (e.g. RTV boosting effect on LPV) and the failed anti-SARS-CoV-2 trial results for treatments with RTV/LPV, hydroxychloroquine combined with azithromycin, and the limited clinical efficacy of the RDV monotherapy.

Several considerations were additionally included in the IDentif.AI workflow to increase the *Clinical Applicability and Acceptability* of the generated results. The drugs in the screening pool were selected with considerations on their practical use in the clinic: drug administration route (intravenous vs. peroral), cost, and access, among others. To this end, IDentif.AI can be designed to include drugs identified as accessible and deployable by a specific country or community, so as to make the most out of what is available in different geographical locations and economic settings. In another realization, the initial drug pool for IDentif.AI screening can be selected to make the IDentif.AI results actionable, such as to exclude certain drug classes for the populations requiring their exclusion due to e.g., chronic diseases, risk profiles, pregnancy or cultural aspects and beliefs. An ability to tailor IDentif.AI drug pool to the target population is aligned with the principles of precision public health approach.

A technical consideration included in IDentif.AI to increase *clinical applicability and acceptability* was to constrain the selection of the concentration levels for the drug–dose combination by each drug's maximum plasma concentration in human (C_{max}). This strategy avoided operating with drug concentrations that lead to

a common pitfall of low EC_x:C_{max} ratio, which in turn is known to preclude translation of the *in vitro* results to the *in vivo* setting. Additionally, as IDentif.AI yields not one optimal drug combination, but a ranked list of the combinatory and monotherapy treatments, it allows to compare and select treatments that are feasible for use in a sustainable fashion when fighting pathogen threats. For example, the IDentif.AI-generated ranked list may inform combinatorial strategies with lower drug concentrations but with efficacy comparable to drug monotherapies, which may be of particular interest to reduce risk of the pathogen drug resistance development.

All in all, the IDentif.AI platform offers several unique features to serve as a pandemic preparedness platform. It also stresses the importance of immersing the AI platform in the context of the pandemic, understanding its limitations and a multiprong approach towards identifying the most appropriate input data to ensure maximal value of the generated results in a real-world setting.

Use of Real-World Data to Identify Potential Targets for Drug Repurposing

Although several different definitions of the term RWD are in use [55, 56], the term can generally be said to encompass patient information gathered in the course of routine clinical practice rather than through prospective trials or other forms of research. In the current era of broadly implemented electronic health records (EHR), RWD has the obvious advantage of the abundance and immediate availability of data such as patient demographics, medications and surgical interventions, imaging, and laboratory test results, without the need for organizing prospective data collection. The potential value of RWD in drug development is therefore apparent already under ordinary circumstances, especially in the areas of adverse event monitoring, study recruitment optimization and repurposing [57], and these advantages become even more relevant in the time criticality of a developing pandemic. Consequently, initiatives have been initiated during the COVID-19 pandemic to mine RWD for information that can instruct drug repurposing efforts [10]. For example, analysis using network-based prediction and a propensity score matching of 26,779 individuals in a Cleveland Clinic Health System COVID-19 registry showed that melatonin use was associated with decreased likelihood of SARS-CoV-2 positivity, providing support for further, prospective investigation of melatonin as a potential treatment against COVID-19 [23] in several clinical trials [58].

In addition to its benefit of being available for retrospective analysis of large patient data volumes, RWD also reflect the diversity and heterogeneity of the potential target population and different clinical settings, in contrast to the narrowly defined population and strictly controlled environment typically applied in RCTs.

However, there are many limitations to the applicability of RWD analysis for drug repurposing. Challenges in using RWD include operational feasibility (e.g., data access and cost, availability of relevant data needed, data protection, patients'

consent, availability of hospital data source), governance (e.g., data-sharing policy, transparency, policy towards funding source), sustainability of data collection and analysis, technical feasibility (e.g. extent and completeness of data collected, documentation of time course, consistent use of terminology and data formats), and methodological feasibility (variability across different data sources, adequacy of documentation of potential confounding factors and effect modifiers such as drug dose and disease severity) [59]. Importantly, when considering the aforementioned discussion of using IDentif.AI to simultaneously reconcile extraordinarily large drug and dose parameter spaces, it is essential to note the specificity of data required in order to mediate global optimization. In these scenarios of small data-driven optimization, specific permutations of drug combinations with defined doses for each drug are needed to accurately represent the parameter spaces. As such, the required dataset is carefully designed and prospectively acquired. Therefore, in the context of RWD, the specific datasets needed likely do not exist within current databases and would instead need to be prospectively acquired. These considerations shed light on the need to re-think the role of data acquisition and co-curation alongside all stakeholders in a healthcare workflow to address this key concept: We must move beyond the amount data acquired and think about how the data is acquired.

Indeed, the most prominent benefits of RWD, i.e. data availability and heterogeneity, also present the greatest limitations of using RWD for drug repurposing; lack of standardization of data formats mean that information from different healthcare providers' systems may often not be possible to merge and pool, instead requiring resource-intensive and time-consuming manual data transfer which may often be prohibitively costly and associated with risks of introducing human error into the data. Such limitations might be overcome with AI, specifically with natural language processing for automated identification and mapping of clinical data to common data standards, but thereby instead introducing the risk systematic errors in data transfer. Furthermore, RWD collected outside the controlled environment of prospective studies is notoriously prone to incompleteness and gaps in documentation and large individual variation in how and what individual clinicians actually record in free text health records, for example how symptoms and their severity are described and classified in free text. Additionally, data silos between and within healthcare organizations often lead to a lack of longitudinal data of each patient, i.e. information from different timepoints along the patient journey may be missing as patient events such as complication, re-hospitalization or changes in diagnostic results cannot be tracked over time, leading to incomplete capture of efficacy as well as safety outcomes. Taking these limitations into account mean that RWD approaches for identifying drug repurposing candidates are particularly appropriate in acute settings where the whole patient management sequence of interest can be expected to be completed within one institution (typically in the in-hospital setting covering the patient journey between emergency room, intensive care unit and ward up until discharge) and situations where data from large patient volumes can be accessed in order to compensate for data gaps on the individual patient level.

Furthermore, the utility of RWD as a data source is limited by the lack of balanced, systematic assignment of patients to treatment arms as in RCTs, leading to

the risk of selection bias as a severe confounding factor when evaluating potential drug targets. To address such concerns, patient matching based on factors that might be expected to influence treatment choice is typically applied to achieve comparability between treated and non-treated groups [60]. Propensity score is a commonly used method to calculate the likelihood of a patient receiving a specific treatment, and thereby provides the basis for matching in retrospective study designs. Machine learning applications have also been shown to calculate propensity score more accurately than traditional methods based on logistic regression analysis, thereby providing an important tool in the armamentarium needed to make RWD useful in the identification of candidates for drug repurposing in the pandemic situation [23].

Future Directions

The response to COVID-19 pandemic offers multiple learnings, starting from the early phase response. At the time, with lack of access to conclusive prospective data on effective treatments, clinical practitioners prescribed unproven repurposed drugs to a great degree worldwide, as shown in e.g. multinational network cohort study of RWD data sources in USA, South Korea, Spain, China [61]. Although data from such spontaneous repurposing efforts could have informed decisions to stop, start or accelerate prospective evidence generation for repurposing of available antiviral or adjunct drugs in COVID-19, little evidence was in fact generated in real time regarding the prescribing patterns in routine clinical practice [61]. For the future, greater efforts to build collaborations to exchange data and apply AI approaches in analyzing such prescribing data are warranted, to allow more rapid identification of potential targets for repurposing and design of prospective trials to quickly validate the potential utility of such target drug candidates. It is therefore a welcome development that increased international efforts are going in this direction, for example with the opening of WHO's dedicated hub for epidemic and pandemic intelligence hub with the mission to enable better data sharing and analytics in order to enable the world to respond better to future health emergencies [62].

The developers of the AI platforms used in response to the onset of the COVID-19 pandemic now have a unique opportunity to further develop and validate the software and hardware, generate evidence and align with the applicable regulatory framework (e.g. clinical decision support system (CDSS) if the AI platform can generate results actionable by the clinicians) or industry standards and certifications. The generation of substantial evidence of performance and desired characteristics is also important for instilling trust in the proposed AI platform, such that it is ready for a deployment in the case of a future outbreak and its results are widely acceptable and interpretable by the community.

Additionally, AI can enable truly personalized optimization of dose and drug combinations at the patient level. The concept of N-of-1 models for dose optimization based only on individual patient data has been shown to be a viable alternative to big data approaches based on systems biology, omics and real-world clinical data.

The platform CURATE.AI uses only the relationship between the actual pharmaceutical exposure and the phenotypic response of the patient, typically by recording drug intake and changes in well-validated, disease-specific biomarkers, without the need for population data or extensive PK/PD data from the individual [63–72]. This approach offers benefits especially for the treatment of prolonged conditions and could contribute to personalization of potential future treatment regimens for long-term post-acute COVID-19 syndrome known as long COVID.

The COVID-19 pandemic has had an unprecedented impact on all aspects of people's lives. The 2 years since the pandemic onset offer valuable lessons and have revealed a range of important considerations to be taken into account when formulating a strategic response at different stages of a pandemic timeline. AI-based technologies developed in response to the threat—including AI-based drug discovery and drug repurposing platforms—now have an opportunity to mature and become a part of a preparedness response system ready for future potential epidemic or pandemic threats. Efforts to strengthen both short- and long-term strategies for developing AI technologies, scientific and regulatory framework, global collaborations, and partnership should be prioritized, in order to strengthen our preparedness for future pandemics by pre-emptively mitigating the impact when another crisis strikes.

References

1. Beck BR, Shin B, Choi Y, et al. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug–target interaction deep learning model. Comput Struct Biotechnol J. 2020;18:784–90. https://doi.org/10.1016/j.csbj.2020.03.025.
2. Kadioglu O, Saeed M, Greten HJ, et al. Identification of novel compounds against three targets of SARS CoV-2 coronavirus by combined virtual screening and supervised machine learning. Comput Biol Med. 2021;133:104359. https://doi.org/10.1016/j.compbiomed.2021.104359.
3. Richardson P, Griffin I, Tucker C, et al. Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. Lancet (London, England). 2020;395(10223):e30. https://doi.org/10.1016/S0140-6736(20)30304-4.
4. Abdulla A, Wang B, Qian F, et al. Project IDentif. AI: harnessing artificial intelligence to rapidly optimize combination therapy development for infectious disease intervention. Adv Ther. 2020;3(7):2000034. https://doi.org/10.1002/adtp.202000034.
5. Blasiak A, Lim JJ, Seah SGK, et al. IDentif. AI: Rapidly optimizing combination therapy design against severe Acute Respiratory Syndrome Coronavirus 2 (SARS-Cov-2) with digital drug development. Bioeng Transl Med. 2021a;6(1):e10196. https://doi.org/10.1002/btm2.10196.
6. Blasiak A, Truong AT, Remus A, et al. The IDentif. AI 2.0 pandemic readiness platform: rapid prioritization of optimized COVID-19 combination therapy regimens. medRxiv. 2021b;2021:9321. https://doi.org/10.1101/2021.06.23.21259321.
7. Belyaeva A, Cammarata L, Radhakrishnan A, et al. Causal network models of SARS-CoV-2 expression and aging to identify candidates for drug repurposing. Nat Commun. 2021;12:1024. https://doi.org/10.1038/s41467-021-21056-z.
8. Gysi DM, Do Valle Í, Zitnik M, et al. Network medicine framework for identifying drug-repurposing opportunities for COVID-19. Proc Natl Acad Sci. 2021;118(19):e2025581118. https://doi.org/10.1073/pnas.2025581118.

9. Zhou Y, Hou Y, Shen J, et al. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discov. 2020a;6:14. https://doi.org/10.1038/s41421-020-0153-3.
10. Zhou Y, Wang F, Tang J, et al. Artificial intelligence in COVID-19 drug repurposing. Lancet Digit Health. 2020c;2(12):e667–76. https://doi.org/10.1016/S2589-7500(20)30192-8.
11. Segler MH, Preuss M, Waller MP. Planning chemical syntheses with deep neural networks and symbolic AI. Nature. 2018;555(7698):604–10. https://doi.org/10.1038/nature25978.
12. Stebbing J, Phelan A, Griffin I, et al. COVID-19: combining antiviral and anti-inflammatory treatments. Lancet Infect Dis. 2020b;20(4):400–2. https://doi.org/10.1016/S1473-3099(20)30132-8.
13. Zeng X, Song X, Ma T, et al. Repurpose open data to discover therapeutics for COVID-19 using deep learning. J Proteome Res. 2020;19(11):4624–36. https://doi.org/10.1021/acs.jproteome.0c00316.
14. Galindez G, Matschinske J, Rose TD, et al. Lessons from the COVID-19 pandemic for advancing computational drug repurposing strategies. Nat Comput Sci. 2021;1:33–41. https://doi.org/10.1038/s43588-020-00007-6.
15. Ge Y, Tian T, Huang S, et al. An integrative drug repositioning framework discovered a potential therapeutic agent targeting COVID-19. Signal Transduct Target Ther. 2021;6(1):165. https://doi.org/10.1038/s41392-021-00568-6.
16. Stebbing J, Krishnan V, de Bono S, et al. Mechanism of baricitinib supports artificial intelligence-predicted testing in COVID-19 patients. EMBO Mol Med. 2020a;12(8):e12697. https://doi.org/10.15252/emmm.202012697.
17. Kalil AC, Patterson TF, Mehta AK, et al. Baricitinib plus remdesivir for hospitalized adults with Covid-19. N Engl J Med. 2021;384(9):795–807. https://doi.org/10.1056/NEJMoa2031994.
18. Marconi VC, Ramanan AV, de Bono S, et al. Efficacy and safety of baricitinib for the treatment of hospitalised adults with COVID-19 (COV-BARRIER): a randomised, double-blind, parallel-group, placebo-controlled phase 3 trial. Lancet Respir Med. 2021;9(12):1407–18. https://doi.org/10.1016/S2213-2600(21)00331-3.
19. Khan M, Mehran MT, Haq ZU, et al. Applications of artificial intelligence in COVID-19 pandemic: a comprehensive review. Expert Syst Appl. 2021;185:115695. https://doi.org/10.1016/j.eswa.2021.115695.
20. Lv H, Shi L, Berkenpas JW, et al. Application of artificial intelligence and machine learning for COVID-19 drug discovery and vaccine design. Brief Bioinform. 2021;22(6):bbab320. https://doi.org/10.1093/bib/bbab320.
21. Cheng F, Desai RJ, Handy DE, et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. Nat Commun. 2018;9(1):1–12. https://doi.org/10.1038/s41467-018-05116-5.
22. Guney E, Menche J, Vidal M, et al. Network-based in silico drug efficacy screening. Nat Commun. 2016;7:10331. https://doi.org/10.1038/ncomms10331.
23. Zhou Y, Hou Y, Shen J, et al. A network medicine approach to investigation and population-based validation of disease manifestations and drug repurposing for COVID-19. PLoS Biol. 2020b;18(11):e3000970. https://doi.org/10.1371/journal.pbio.3000970.
24. Sun W, Sanderson PE, Zheng W. Drug combination therapy increases successful drug repositioning. Drug Discov Today. 2016;21(7):1189–95. https://doi.org/10.1016/j.drudis.2016.05.015.
25. Ellinger B, Bojkova D, Zaliani A, et al. A SARS-CoV-2 cytopathicity dataset generated by high-content screening of a large drug repurposing collection. Sci Data. 2021;8(1):1–10. https://doi.org/10.1038/s41597-021-00848-4.
26. RECOVERY Collaborative Group. Dexamethasone in hospitalized patients with Covid-19. N Engl J Med. 2021;384(8):693–704. https://doi.org/10.1056/NEJMoa2021436.
27. Peele KA, Durthi CP, Srihansa T, et al. Molecular docking and dynamic simulations for antiviral compounds against SARS-CoV-2: a computational study. Inform Med Unlocked. 2020;19:100345. https://doi.org/10.1016/j.imu.2020.100345.
28. Muralidharan N, Sakthivel R, Velmurugan D, et al. Computational studies of drug repurposing and synergism of lopinavir, oseltamivir and ritonavir binding with SARS-CoV-2 prote-

ase against COVID-19. J Biomol Struct Dyn. 2021;39(7):2673–8. https://doi.org/10.1080/07391102.2020.1752802.
29. Batra R, Chan H, Kamath G, et al. Screening of therapeutic agents for COVID-19 using machine learning and ensemble docking studies. J Phys Chem Lett. 2020;11(17):7058–65. https://doi.org/10.1021/acs.jpclett.0c02278.
30. Mohapatra S, Nath P, Chatterjee M, et al. Repurposing therapeutics for COVID-19: rapid prediction of commercially available drugs through machine learning and docking. PLoS One. 2020;15(11):e0241543. https://doi.org/10.1371/journal.pone.0241543.
31. Anwaar MU, Adnan F, Abro A, et al. Combined deep learning and molecular docking simulations approach identifies potentially effective FDA approved drugs for repurposing against SARS-CoV-2. Comput Biol Med. 2022;141:105049. https://doi.org/10.1016/j.compbiomed.2021.105049.
32. Gimeno A, Ojeda-Montes MJ, Tomás-Hernández S, et al. The light and dark sides of virtual screening: what is there to know? Int J Mol Sci. 2019;20(6):1375. https://doi.org/10.3390/ijms20061375.
33. Marklund EG, Benesch JL. Weighing-up protein dynamics: the combination of native mass spectrometry and molecular dynamics simulations. Curr Opin Struct Biol. 2019;54:50–8. https://doi.org/10.1016/j.sbi.2018.12.011.
34. Al-Shyoukh I, Yu F, Feng J, et al. Systematic quantitative characterization of cellular responses induced by multiple signals. BMC Syst Biol. 2011;5(88):1–17. https://doi.org/10.1186/1752-0509-5-88.
35. Ding X, Sanchez DJ, Shahangian A, et al. Cascade search for HSV-1 combinatorial drugs with high antiviral efficacy and low toxicity. Int J Nanomedicine. 2012;7:2281. https://doi.org/10.2147/IJN.S27540.
36. Honda Y, Ding X, Mussano F, et al. Guiding the osteogenic fate of mouse and human mesenchymal stem cells through feedback system control. Sci Rep. 2013;3:3420. https://doi.org/10.1038/srep03420.
37. Liu Q, Zhang C, Ding X, et al. Preclinical optimization of a broad-spectrum anti-bladder cancer tri-drug regimen via the Feedback System Control (FSC) platform. Sci Rep. 2015;5:11464. https://doi.org/10.1038/srep11464.
38. Tekin E, White C, Kang TM, et al. Prevalence and patterns of higher-order drug interactions in *Escherichia coli*. NPJ Syst Biol Appl. 2018;4(1):1–10. https://doi.org/10.1038/s41540-018-0069-9.
39. Tsutsui H, Valamehr B, Hindoyan A, et al. An optimized small molecule inhibitor cocktail supports long-term maintenance of human embryonic stem cells. Nat Commun. 2011;2:167. https://doi.org/10.1038/ncomms1165.
40. Valamehr B, Tsutsui H, Ho C-M, et al. Developing defined culture systems for human pluripotent stem cells. Regen Med. 2011;6(5):623–34. https://doi.org/10.2217/rme.11.54.
41. Wang H, Silva A, Ho C-M. When medicine meets engineering—paradigm shifts in diagnostics and therapeutics. Diagnostics. 2013;3(1):126–54. https://doi.org/10.3390/diagnostics3010126.
42. Wei F, Bai B, Ho C-M. Rapidly optimizing an aptamer based BoNT sensor by feedback system control (FSC) scheme. Biosens Bioelectron. 2011;30(1):174–9. https://doi.org/10.1016/j.bios.2011.09.014.
43. Weiss A, Berndsen RH, Ding X, et al. A streamlined search technology for identification of synergistic drug combinations. Sci Rep. 2015;5:14508. https://doi.org/10.1038/srep14508.
44. Wong PK, Yu F, Shahangian A, et al. Closed-loop control of cellular functions using combinatory drugs guided by a stochastic search algorithm. Proc Natl Acad Sci. 2008;105(13):5105–10. https://doi.org/10.1073/pnas.0800823105.
45. Yu F, Al-Shyoukh I, Feng J, et al. Control of Kaposi's sarcoma-associated herpesvirus reactivation induced by multiple signals. PLoS One. 2011;6(6):e20998. https://doi.org/10.1371/journal.pone.0020998.
46. Yu H, Zhang WL, Ding X, et al. Optimizing combinations of flavonoids deriving from astragali radix in activating the regulatory element of erythropoietin by a feedback sys-

tem control scheme. Evid Based Complement Alternat Med. 2013;2013:1436. https://doi.org/10.1155/2013/541436.
47. Xu H, Jaynes J, Ding X. Combining two-level and three-level orthogonal arrays for factor screening and response surface exploration. Stat Sin. 2014;24(1):269–89. https://www.jstor.org/stable/26432543
48. de Mel S, Rashid MB, Zhang XY, et al. Application of an ex-vivo drug sensitivity platform towards achieving complete remission in a refractory T-cell lymphoma. Blood Cancer J. 2020;10(9):1–5. https://doi.org/10.1038/s41408-020-0276-7.
49. Rashid MBMA, Toh TB, Hooi L, et al. Optimizing drug combinations against multiple myeloma using a quadratic phenotypic optimization platform (QPOP). Sci Transl Med. 2018;10(453):eaan0941. https://doi.org/10.1126/scitranslmed.aan0941.
50. Clemens DL, Lee B-Y, Silva A, et al. Artificial intelligence enabled parabolic response surface platform identifies ultra-rapid near-universal TB drug treatment regimens comprising approved drugs. PLoS One. 2019;14(5):e0215607. https://doi.org/10.1371/journal.pone.0215607.
51. Lee B-Y, Clemens DL, Silva A, et al. Drug regimens identified and optimized by output-driven platform markedly reduce tuberculosis treatment time. Nat Commun. 2017;8:14183. https://doi.org/10.1038/ncomms14183.
52. Lee B-Y, Clemens DL, Silva A, et al. Ultra-rapid near universal TB drug regimen identified via parabolic response surface platform cures mice of both conventional and high susceptibility. PLoS One. 2018;13(11):e0207469. https://doi.org/10.1371/journal.pone.0207469.
53. Shen Y, Liu T, Chen J, et al. Harnessing artificial intelligence to optimize long-term maintenance dosing for antiretroviral-naive adults with HIV-1 infection. Adv Ther. 2020;3(4):1900114. https://doi.org/10.1002/adtp.201900114.
54. Sun J, Wang B, Warden AR, et al. Overcoming multidrug-resistance in bacteria with a two-step process to repurpose and recombine established drugs. Anal Chem. 2019;91(21):13562–9. https://doi.org/10.1021/acs.analchem.9b02690.
55. Jarow JP, LaVange L, Woodcock J. Multidimensional evidence generation and FDA regulatory decision making: defining and using "real-world" data. JAMA. 2017;318(8):703–4. https://doi.org/10.1001/jama.2017.9991.
56. Makady A, de Boer A, Hillege H, et al. What is real-world data? A review of definitions based on literature and stakeholder interviews. Value Health. 2017;20(7):858–65. https://doi.org/10.1016/j.jval.2017.03.008.
57. Chen Z, Liu X, Hogan W, et al. Applications of artificial intelligence in drug development using real-world data. Drug Discov Today. 2021;26(5):1256–64. https://doi.org/10.1016/j.drudis.2020.12.013.
58. Wichniak A, Kania A, Siemiński M, et al. Melatonin as a potential adjuvant treatment for COVID-19 beyond sleep disorders. Int J Mol Sci. 2021;22(16):8623. https://doi.org/10.3390/ijms22168623.
59. Cave A, Kurz X, Arlett P. Real-world data for regulatory decision making: challenges and possible solutions for Europe. Clin Pharmacol Ther. 2019;106(1):36–9. https://doi.org/10.1002/cpt.1426.
60. Li Q, Lin J, Chi A, et al. Practical considerations of utilizing propensity score methods in clinical development using real-world and historical data. Contemp Clin Trials. 2020;97:106123. https://doi.org/10.1038/srep11464.
61. Prats-Uribe A, Sena AG, Lai LYH, et al. Use of repurposed and adjuvant drugs in hospital patients with Covid-19: multinational network cohort study. BMJ. 2021;373:n1038. https://doi.org/10.1136/bmj.n1038.
62. World Health Organization. WHO, Germany open hub for pandemic and epidemic intelligence in Berlin. 2021. https://www.who.int/news/item/01-09-2021-who-germany-open-hub-for-pandemic-and-epidemic-intelligence-in-berlin. Accessed 12 Jan 2022.
63. Blasiak A, Kee TW, Rashid MBM, et al. CURATE.AI-optimized modulation for multiple myeloma: an N-of-1 randomized trial [abstract]. AACR. Cancer Res. 2020a;80:CT268. https://doi.org/10.1158/1538-7445.am2020-ct268.

64. Blasiak A, Khong J, Kee T. CURATE.AI: optimizing personalized medicine with artificial intelligence. SLAS Technol. 2020b;25(2):95–105. https://doi.org/10.1177/2472630319890316.
65. Ho D. Artificial intelligence in cancer therapy. Science. 2020a;367(6481):982–3. https://doi.org/10.1126/science.aaz3023.
66. Ho D. Addressing COVID-19 drug development with artificial intelligence. Adv Intell Syst. 2020b;2(5):2000070. https://doi.org/10.1002/aisy.202000070.
67. Pantuck AJ, Lee D-K, Kee T, et al. Modulating BET bromodomain inhibitor ZEN-3694 and enzalutamide combination dosing in a metastatic prostate cancer patient using CURATE. AI, an artificial intelligence platform. Adv Ther. 2018;1(6):1800104. https://doi.org/10.1002/adtp.201800104.
68. Tan BKJ, Teo CB, Tadeo X, et al. Personalised, rational, efficacy-driven cancer drug dosing via an artificial Intelligence SystEm (PRECISE): a protocol for the PRECISE CURATE. AI pilot clinical trial. Front Digit Health. 2021;3(16):5524. https://doi.org/10.3389/fdgth.2021.635524.
69. Truong AT, Tan LW, Chew KA, et al. Harnessing CURATE. AI for N-of-1 optimization analysis of combination therapy in hypertension patients: a retrospective case series. Adv Ther. 2021;4(10):2100091. https://doi.org/10.1002/adtp.202100091.
70. Zarrinpar A, Lee D-K, Silva A, et al. Individualizing liver transplant immunosuppression using a phenotypic personalized medicine platform. Sci Transl Med. 2016;8(333):333ra49. https://doi.org/10.1126/scitranslmed.aac5954.
71. Gentile F, Agrawal V, Hsing M, et al. Deep docking: a deep learning platform for augmentation of structure based drug discovery. ACS Cent Sci. 2020;6(6):939–49. https://doi.org/10.1021/acscentsci.0c00229.
72. WHO Rapid Evidence Appraisal for COVID-19 Therapies (REACT) Working Group, Sterne JA, Murthy S, et al. Association between administration of systemic corticosteroids and mortality among critically ill patients with COVID-19: a meta-analysis. JAMA. 2020;324(13):1330–41. https://doi.org/10.1001/jama.2020.17023.

Chapter 4
AI and Point of Care Image Analysis for COVID-19

Michael Roberts, Oz Frank, Shai Bagon, Yonina C. Eldar, and Carola-Bibiane Schönlieb

Introduction

Point of care image analysis for hospitalised patients remains a largely manual process. Typically, in Europe and the US, when a patient suspected to have COVID-19 is admitted to the hospital, a Chest X-ray (CXR) is acquired and a reverse transcription polymerase chain reaction (RT-PCR) test is performed. This CXR is then interpreted by radiologists and other clinicians who will identify particular patterns which may indicate a likely initial diagnosis and prognosis. When the results of the RT-PCR test are returned, it will replace, or confirm, this initial image-based diagnosis. If the CXR or the patient display any particular complications then further imaging, such as computed tomography (CT) or ultrasound (US) can be requested, which are then interpreted by clinical staff once more.

M. Roberts (✉)
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK

Oncology R&D, AstraZeneca, Cambridge, UK
e-mail: mr808@cam.ac.uk; michael.roberts2@astrazeneca.com

O. Frank · S. Bagon
Faculty of Mathematics and Computer Science, Weizmann AI Center (WAIC), Weizmann Institute of Science, Rehovot, Israel
e-mail: oz.frank@weizmann.ac.il; shai.bagon@weizmann.ac.il

Y. C. Eldar
Weizmann Institute of Science, Rehovot, Israel
e-mail: yonina.eldar@weizmann.ac.il

C.-B. Schönlieb
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK
e-mail: cbs31@cam.ac.uk

N. Lidströmer, Y. C. Eldar (eds.), *Artificial Intelligence in Covid-19*,
https://doi.org/10.1007/978-3-031-08506-2_4

In a pandemic, when every resource in the hospital system is under strain, it is imperative to make manual processes as efficient as possible to reduce the time for decisions to be taken and actions made. Artificial intelligence (AI) algorithms and methods promise great potential for automating many routine tasks for clinicians and thereby the promise to improve clinical care. This includes, but is not limited to (a) identifying regions of pathology on images, (b) tracking disease burden in longitudinal imaging and (c) measuring regions and volumes of interest.

As of 2021, we remain a long way from point of care imaging that is informed by AI techniques being available routinely—although it feels tantalisingly close to being a reality. The COVID-19 pandemic has highlighted how AI-based algorithms could have a significant impact—if only they were available to call upon during the height of the pandemic. In particular, if COVID-19 diagnosis were possible based on e.g. an admission CXR, it would have been possible to triage patients immediately to the "green" wards of non-COVID-19 patients and the "red" wards of COVID-19 cases. Unfortunately, when RT-PCR testing capacity was restricted, clinicians would wait anything from 24 to 48 h for a positive or negative result to be returned. This led to inevitable cross-infection and even more extreme stress on the health system.

Similarly, prognostication for COVID-19 patients using AI-based methods has the promise to provide huge improvement in clinical care and allow for better resource management. Examples of prognostication tasks are: (a) prediction of ventilation requirement and the level of ventilation required, (b) prediction of response to treatments (such as dexamethasone) and the ideal time to administer for optimal response, (c) prediction of the patients that will experience acute respiratory distress syndrome (ARDS).

In this chapter we focus on the reality, rather than the promise, of how AI was applied to point of care imaging (CXR, CT and US) in the COVID-19 pandemic. We will review each imaging modality separately and highlight models described in the literature along with common themes, pitfalls and recommendations for how future models can be developed following best practice. We conclude this chapter by providing some success stories and reasons for optimism, highlight some of the dangers of developing models which do not perform as expected along with lessons learned from this pandemic that we can take forward to be better prepared for the next one.

Motivation for Using Imaging

As the COVID-19 pandemic swept through China in early 2020, chest imaging was used locally as the primary initial diagnostic tool. Meanwhile, European and American radiological societies did not initially support the use of CT and CXR imaging for diagnosis in early March 2020 [1, 2] with the ACR stating

CT should not be used to screen for or as a first-line test to diagnose COVID-19 but this position softened towards the end of March as the pandemic took hold and testing capacity was limited with an update in late March 2020 stating

> The ACR strongly urges caution in taking this approach [...] Clearly, locally constrained resources may be a factor in such decision making.

Several studies also indicate that, in addition to imaging being a potential diagnostic tool, it also encodes prognostic information about the disease. For example, the extent of opacification in the lungs of COVID-19 patients is a significant prognostic marker of mortality [3].

Figure 4.1, courtesy of [6], displays common presentations of COVID-19 in CT scans and CXRs. In both the CXR and CT imaging we see ground-glass opacities in the regions affected by COVID-19, with the CT scans showing a crazy-paving style pattern inside those ground-glass opacities.

Motivation for Using AI with Imaging

The ground-glass opacification of the lungs with a 'crazy-paving' pattern seen in CT scans for COVID-19 patients motivates the idea that pattern recognition algorithms hold the potential to aid clinicians in the diagnosis and prognostication of COVID-19 via chest imaging [7].

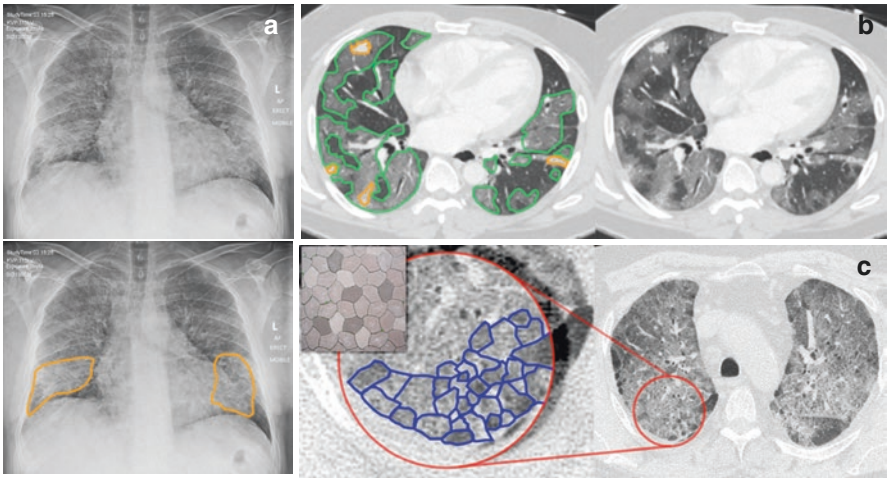


Fig. 4.1 Annotated examples of COVID-19 scans. (**a**) Chest X-ray (CXR) with ground-glass opacification in both lungs and consolidation (outlined in orange). (**b**) A CT scan that shows ground-glass opacification (green) and consolidation (orange). (**c**) A CT scan that indicates severe COVID-19 with a crazy-paving pattern. (Images from the NCCID [4], and inset from [5])

The COVID-19 pandemic is the first of the machine learning era and, given recent developments in the application of machine learning models to medical imaging problems [7–10], there is fantastic promise for applying machine learning methods to COVID-19 radiological imaging for improving the accuracy of diagnosis, compared to the gold-standard RT-PCR, whilst also providing valuable insight for prognostication of patient outcomes. These models have the potential to exploit the large amount of multi-modal data collected from patients and could, if successful, transform detection, diagnosis, and triage of patients with suspected COVID-19. One model of huge potential utility is a model which can not only distinguish COVID-19 from non-COVID-19 patients but also discern alternative types of pneumonia such as those of bacterial or other viral aetiologies. For prognostication, it would is desirable to develop models that predict responses to therapies and clinical pathways for patients. This would allow for resource forecasting, patient triage and ultimately improved care.

Integration of Imaging with Other Modalities

In developing machine learning models for COVID-19 the model input can be from a variety of sources, including but not limited to electronic health record records, full blood count data, chest imaging, audio recording of coughs, symptom diaries, genetics, and more. A single patient will have data recorded about them in several modalities and at different levels of granularity. We aim to demonstrate why it is important to combine/fuse data from multiple sources to get models that make predictions using holistic understanding of the patient's condition.

In late February 2020, the Diamond Princess cruise ship had the largest cluster of positive COVID-19 cases outside of China. A study of 104 of these COVID-19-positive patients found that 73% (76 out of 104) were asymptomatic. However, 54% (41 out of 76) of these asymptomatic individuals displayed lung opacities on their CT scans. The converse was also true, as roughly 21.5% (6 out of 28) of symptomatic patients had normal CT findings [11]. Imaging features alone are clearly not sufficient for accurate diagnosis and neither are the clinical features alone enough to understand the degree of disease in a patient.

It is also important to recognise that clinicians routinely use multiple data sources to develop their judgment for a patient's likely diagnosis and their likely clinical outcome. However, although it is simple for the clinician to do this, fusing multi-modal data is not trivial in a machine learning framework.

Literature Overview

There are a huge number of papers which discuss machine learning models for COVID-19 diagnosis or prognosis using point of care imaging. Indeed, for 2020 and 2021, a basic search of Arxiv, BioRxiv, MedRxiv and Pubmed for papers which mention machine/deep learning and COVID-19 and CT/CXR/US imaging returns

848 results with 294 of these being preprints. The three journals publishing the most papers on the subject were Scientific Reports (17), IEEE Journal of Biomedical and Health Informatics (16) and PLoS ONE (14).

In keeping with this large corpus of literature and the fast moving nature of the pandemic, there have also been many systematic reviews, 27 in total with 16 of them published. The systematic reviews most relevant to this Chapter are [12–34].

Several of the authors of this chapter performed the systematic review [12] which examines the entire literature from January 1, 2020 to October 3, 2020 and identifies 320 published and preprint manuscripts that develop machine learning models using chest CT or radiographs for COVID-19 diagnosis or prognostication. Unfortunately, as we will discuss throughout this Chapter, many of the papers contained systematic issues pertaining to image sourcing, quality, and documentation that introduce bias in developed models, ultimately making them unlikely to perform well in practice [6].

In our search of the literature, we found that 637/848 (75.1%) of papers mention diagnosis, detection, diagnostics, screening, recognition, discrimination, identification or classification of COVID-19 in the title of the manuscript whilst only 94/848 (11.1%) consider prognostication (17 papers consider both). For completion, the remaining papers are 59/848 (7.0%) which design segmentation models for COVID-19 patterns in imaging (15 of which also perform diagnosis) and the remaining 82 papers are literature reviews, introduce new COVID-19 datasets or discuss methodologies such as image reconstruction or denoising.

This significant bias towards the development of diagnosis models is understandable as at the start of the pandemic there was a hunger for a solution to the slow processing of RT-PCR tests globally (typically 24–48 h) which would allow for rapid triage of patients into isolation if needed. COVID-19 imaging and clinical data was also scarce and precious at the start of the pandemic, where non-COVID-19 data was relatively plentiful, and it was therefore easier to develop methods which aim to detect the COVID-19 data among a sea of other data. Prognostic models require well curated data, with imaging linked to standardised outcomes which is harder to collect. This is a challenge with AI methods, which rely on large, diverse data sets, in order to succeed.

While X-ray and CT are commonly used during the treatment of respiratory conditions, lung ultrasound (LUS) is not an obvious choice due to the unique acoustical properties of the lungs. We argue that despite its challenging properties LUS can and should be considered as a point-of-care modality, and further show how AI can assist in achieving this goal. In the rest of the chapter, we discuss AI methods for X-ray, CT and ultrasound imaging, respectively.

Chest X-Ray Imaging

In Europe, the US and most countries of the world chest X-ray imaging is used as the first-line imaging given to any COVID-19 patient once they enter the hospital. Chest X-rays are fast to acquire and relatively low cost, so a patient is highly likely

to have several in a normal admission. Early into the pandemic, when RT-PCR results were commonly only returned 24 h or more after the sample was taken, the CXR images were used in combination with clincal symptoms to make a likely diagnosis of the patients. In addition to the initial diagnosis of patients, CXRs can be used to determine a prognosis for the patient based on location and extent of the ground-glass opacities and consolidation in the lungs. In this section, we discuss many of the approaches taken to diagnosis and prognosis of COVID-19 using CXR images along with specific focus on those which use longitudinal imaging in their models and fuse data from non-imaging modalities into their input features. We conclude by highlighting many of the issues that are common across the literature.

Diagnosis Models

Machine learning models for the diagnosis of COVID-19 using CXR imaging and tend to pose the problem as either a two class problem of COVID-19 vs. Non-COVID-19 or a more complex multi-class problem, such as COVID-19 vs. bacterial pneumonia vs. other viral pneumonia vs. healthy. The latter is more nuanced as it requires careful definition and identification of the non-COVID-19 classes to which the model is likely to be applied in practice whereas a two-class model collates all the non-COVID-19 data into one class. Most papers classify images into the three classes COVID-19, non-COVID-19 pneumonia and normal while several consider an extra class by dividing non-COVID-19 pneumonia into viral and bacterial pneumonia.

Commonly, the CXR images are pre-processed by segmenting the lungs to remove biases in the images, such as labels imprinted on the image, and artefacts around the border of the CXR. In [35], the authors consider a segmentation path input together with the image itself, and show that this enhances performance. They also introduce further pre-processing techniques based on augmentations inspired by clinical input which further improves detection performance.

Most papers directly apply deep learning methodologies to the question of diagnosis. Simply using the images and associated labels to train networks which learn their own features. For this, most authors chose to use transfer learning, taking models trained on existing CXR datasets such as CheXpert and fine-tuning them to the COVID-19 data. Most papers use off-the-shelf network architectures based on the root model being used for transfer learning, including ResNet-18 or ResNet-50, DenseNet-121, VGG-16 or VGG-19, Inception and EfficientNet. Few papers develop their own custom architectures. ResNet and DenseNet architectures reported better performance than the others, with accuracies ranging from 0Â·88 to 0·99. However, we caution against direct comparison since the papers use different training and testing settings (e.g. different datasets and data partition sizes) and consider a different number of classes.

Some papers use a more traditional approach, extracting hand engineered features from the images and linking these to the outcomes by fitting a machine learning model, such as a random forest or neural network.

Prognosis Models

As discussed, the literature for machine learning algorithms that prognosticate for COVID-19 patients is small compared to diagnostic algorithms. This is due to difficulty in identifying, anonymising and extracting datasets of images and clinical outcomes which are linked to one another and of high quality.

Outcomes which are typically predicted by these models are those such as: death or need for ventilation, a need for ICU admission, progression to acute respiratory distress syndrome, the length of hospital stay, likelihood of conversion to severe disease and the extent of lung infection.

Predictors from radiological data were extracted using either handcrafted radiomic features or deep learning. Clinical data included basic observations, serology and comorbidities. Most papers used models based on a multivariate Cox proportional hazards model, logistic regression, linear regression, random forest or compare a huge variety of machine learning models such as tree-based methods, support vector machines, neural networks and nearest neighbour clustering.

Use of Longitudinal Imaging

As CXR imaging is the most common imaging modality for POC in Europe, the USA and most countries of the world, it is common for patients to have multiple images captured in a single hospital stay. If a model is able to read in several images in sequence for a patient then it is possible to capture the changes in features of the disease as it progresses and this should allow for more accurate prediction of their outcomes. This approach introduces some methodological challenges, such as: images acquired on different machines and of different quality, images acquired at non-equidistant intervals and that the baseline images for all patients will be acquired at different stages of their disease depending on when they presented to hospital.

In the literature, there are several examples of papers which consider longitudinal CXR imaging for COVID-19 prognostication, in particular [36, 37]. These papers employ different techniques to fuse the features extracted from the longitudinal images. In [36], the authors use a self-supervised approach using momentum and contrastive learning (MoCo) [XXX] to train a feature extractor for the CXR

images. This feature extractor is applied to each image in the sequence and the relative time of the image acquistion is concatenated to the resulting feature vector. Therefore, for each image we have one feature vector which encodes the image features and the relative timing. After this, a Transformer [XXX] is applied to the sequence of feature vectors from all of the images and classification is made for whether an adverse event (death, ICU admission or intubation) takes place within 24, 48, 72 or 96 h of the CXR.

In [37], the authors use a Cox proportional hazards model to predict the probability of experiencing death, ICU admission, ICU discharge, hospital admission and hospital discharge before the observing time *t*. They extract features from the images using a convolutional LSTM, concatenate them and use fully connected layers to obtain a risk score.

The literature consistently suggests that using multiple CXR images gives better model performance than using a single time point.

Fusion with Other Data Modalities

Very few papers discuss models which integrate both radiological and clinical data for COVID-19 diagnosis and prognostication. Fusing data of these different modalities is a particular challenge, as these data are of different types and dimensionalities. In particular, clinical data tend to be vectors or tables of features for each patient, potentially with multiple time points per patient. Imaging is not only extremely high-dimensional, with millions of pixels in a normal CXR, but there are also spatial relationships between pixels and structures in the image to consider. Methodologies for fusing this data of different modalities (which is also potentially longitudinal) are being actively studied in the literature [38].

There are some examples in the literature, such as [39, 40], which combine CXR imaging features with clinical data in a machine learning model to predict outcomes for COVID-19 patients. In both cases, these papers state that incorporation of clinical data, in addition to imaging features, improved the performance of the model. The papers take very different approaches to fusing the different modalities.

In [39], the authors manually graded severity of the disease in different zones of the CXR and use these hand-engineered features in combination with the clinical data in a variety of machine learning models (support vector machines, random forest, linear regression and XGBoost). They find that the model developed using a combination of the CXR and clinical features performs better than those developed using only the CXR and clinical features. In [40], the authors use a convolutional neural network to extract the features from the CXR imaging and fuse the clinical data into the final fully connected layer by concatenation. They similarly find that a model developed using the fusion of imaging and clinical data is better than those developed on the modalities separately.

Common Issues with AI and Chest X-Ray Imaging

With no standardisation, AI algorithms for COVID-19 have been developed with a very broad range of applications, data collection procedures and performance assessment metrics. This creates various challenges such as (1) bias in small data sets; (2) variability of large internationally-sourced data sets; (3) poor integration of multi-stream data, particularly imaging data; (4) difficulty of the task of prognostication, and (5) necessity for clinicians and data analysts to work side-by-side to ensure the developed AI algorithms are clinically relevant and implementable into routine clinical care. Since the pandemic began in early 2020, researchers have answered the 'call to arms' and numerous machine learning models for diagnosis and prognosis of COVID-19 using radiological imaging have been developed and hundreds of manuscripts have been written.

In this section, we highlight some of the key systemic issues found in [12] and other systematic reviews.

Duplication and Quality Issues

Many papers rely on public COVID-19 datasets, such as [41–45]. However, there is no restriction for a contributor to upload COVID-19 images to many of these public repositories. There is high likelihood of duplication of images across these sources and no assurance that the cases included in these datasets are confirmed COVID-19 cases (authors take a great leap to assume this is true) so great care must be taken when combining datasets from different public repositories. Also, most of the images have been pre-processed and compressed into non-DICOM formats leading to a loss in quality and a lack of consistency/comparability.

Source Issues

In the literature, many papers use the pneumonia dataset of Kermany et al. [8] as a control (i.e. non-COVID-19) group. However, they commonly fail to mention that this consists of paediatric patients aged between one and five. Developing a model using adult COVID-19 patients and very young pneumonia patients is likely to overperform as it is merely detecting children vs. adults. This dataset is also erroneously referred to as the Mooney dataset in many papers (being the Kermany dataset deployed on Kaggle [46]).

Another significant source issue identified in the systematic reviews of the literature is that data of different classes tend to come from different sources, e.g. RSNA [47] contains only non-COVID-19 pneumonia CXRs, Kermany [8] contains paediatric non-COVID-19 pneumonia CXRs and CheXpert [48] contains CXRs for a range of non-COVID-19 lung diseases. The issue here arises when a machine

learning model learns the source of the data rather than imaging features unique to the diseases of interest.

It is demonstrated by Maguolo et al. [49] that by excluding the lung region entirely, the authors could identify the source of the images in the Cohen et al. [41] and Kermany et al. [8] datasets with an AUC between 0.9210–0.9997 and 'diagnose' COVID-19 with an AUC = 0.68.

Frankenstein Datasets

The previously discussed issues of duplication and sourcing of data become compounded when public 'Frankenstein' datasets are used. These are datasets assembled from other datasets and redistributed under a new name. For instance, one dataset [50] combines several other datasets [8, 41, 43] without realising that one of the component datasets [41] already contains another component [43]. This repackaging of datasets, although pragmatic, inevitably leads to problems with authors developing algorithms (in good faith) which are being trained and tested on identical or overlapping datasets which they believed to be from distinct sources.

Implicit Biases in the Source Data

Images uploaded to a public repository and those extracted from publications [41] are likely to have implicit biases due to the contribution source. For example, it is likely that more interesting, unusual or severe cases of COVID-19 appear in publications.

The urgency of the pandemic led to many studies using datasets that contain obvious biases or are not representative of the target population, e.g. paediatric patients. Before evaluating a model, it is crucial that authors report the demographic statistics for their datasets, including age and sex distributions. Diagnostic studies commonly compare their models' performance to that of RT-PCR. However, as the ground-truth labels are often determined by RT-PCR, there is no way to measure whether a model outperforms RT-PCR from accuracy, sensitivity, or specificity metrics alone. Ideally, models should aim to match clinicians using all available clinical and radiomic data, or to aid them in decision making.

Artificial Limitations Due to Transfer Learning

Many papers utilise transfer learning in developing their model, which assumes an inherent benefit to performance. However, it is unclear whether transfer learning offers significant performance benefit due to the over-parametrisation of the models [51]. Many publications used the same resolutions such as 224-by-224 or 256-by-256 for training, which are often used for ImageNet classification, indicating that the pre-trained model dictated the image rescaling used rather than clinical

judgement. This is particularly hard to justify, given that the features which differentiate COVID-19 pnuemonia from other diseases are likely subtle and not appreciable at such coarse resolutions.

Computed Tomography Imaging

Computed tomography (CT) images are far more resource intensive to acquire and analyse than CXR imaging. This is due to the high relative cost of machines, requirement for cleaning between patients, technicians to maintain the equipment and the high level of skill required for their analysis. The acquisition of a scan also exposes a patient to a high dose of radiation. Therefore, in many parts of the world, CT scans were reserved for the most complex clinical COVID-19 cases after an initial CXR. However, in China and Russia, CT was used as a first line imaging modality for COVID-19 patients.

There are fewer papers in the literature which focus on CT imaging, however there are a sizeable number. Just as the pandemic was starting to take hold in Europe in March 2020, its effects were starting to ease in China. As a consequence, many of the earliest papers were describing models developed using Chinese data and focused specifically on CT imaging.

Diagnosis Models

Diagnosis of COVID-19 from a CT scan is most relevant and useful for those countries in which CT scans are used as a first-line imaging modality. In countries which use CXR as first-line, a CT scan is used to assess more complex clinical disease but less for diagnosis of disease. Countries which use CT as a first-line modality will also have more CT scans for COVID-19 cases of mild disease, when it is most useful to accurately diagnose COVID-19 to allow for triage and isolation.

The majority of the papers in the literature apply deep learning directly to the CT diagnosis problem and frame it as a classification task, distinguishing COVID-19 from other lung pathologies such as (viral or bacterial) pneumonia, interstitial lung disease and/or a non-COVID-19 class.

In most of these papers, authors consider isolated 2D slices or even 2D patches taken from the 3D volume. This is usually due to computational and storage constraints, as each CT image is typically between 200 and 500 MB and tens or hundreds of millions of voxels. In most 2D models, authors employed transfer learning, with networks pre-trained on ImageNet [52]. Almost all models used lung segmentation as a pre-processing step.

Recognising that there was a relatively small cohort of COVID-19 CT scan data available early in the pandemic, some authors e.g. [53] use Generative Adversarial Network (GAN) [54] approach to create synthetic COVID-19 imaging.

Outside of deep learning based models, other papers in the literature considered more traditional machine learning methods for COVID-19 diagnosis relying on hand-engineered features or CNN-extracted features. Software such as PyRadiomics [XXX] was commonly used to extract a large number of radiomic features from delineated regions in the CT scans and, after feature reduction, a classifier was fit to the remaining features, with most authors using logistic regression.

The performance of models in the literature is highly variable and optimistic with many reporting AUC, sensitivity and specificity values over 0.95. For reference, the gold standard RT-PCR test has a sensitivity of around 80% [XXX].

Prognosis Models

The literature for prognostic models using CT imaging, as with CXR imaging, is relatively small compared to the diagnostic models but approaches have been developed using similar models and features. There is a bias for prognostic models to focus on CT imaging rather than CXR.

As for CXR, models were developed for predicting severity of outcomes including: death or need for ventilation, a need for ICU admission, progression to acute respiratory distress syndrome, the length of hospital stay, likelihood of conversion to severe disease and the extent of lung infection. The features are either handcrafted radiomic features or learned features extracted from a CNN. Most papers fit models based on a multivariate Cox proportional hazards model, logistic regression, linear regression, random forest or compare a huge variety of machine learning models such as tree-based methods, support vector machines, neural networks and nearest neighbour clustering.

Applications to Regions Away from the Lungs

In the literature, almost all papers for CT prognosis and diagnosis of COVID-19 focus on either the full CT scan or the segmented lungs and use these as inputs to the models. However, this ignores many of the other important structures of the body which can be captured in a CT scan.

Most importantly, as poor cardiovascular health one of the largest risk factors for a poor outcome for COVID-19 patients [55], it is highly relevant to consider the heart features in CT images for prognostic models. For example, identifying and quantifying atherosclerosis, tracking changes in the features of the heart through the course of disease and quantifying the extent and volume of epicardial adipose tissue (EAT), i.e. fat around the heart.

In [56], for example, the authors use a semi-automated tool to quantify the EAT around the heart tissue in CT scans for COVID-19 patients and find that there is a link between EAT and the burden of the COVID-19 pneumonia in the lung (ground-glass opacities and consolidation).

In [57], the authors find a higher prevalence of pulmonary embolisms (PEs) in COVID-19 patients and in [58] the authors discuss a deep learning algorithm for identifying PEs. Further study of the vascular structure could also be performed along with a study of changes in morphology and how these link to outcomes.

In the literature considering lung diseases and imaging, airway features have been postulated as markers of disease progression [59, 60]. Not only can one consider the diameter changes and total volume features, but the overall morphology of this complex structure can also be considered and changes monitored. There is not currently existing literature exploring this.

Finally, we also mention that as obesity is known to be a risk factor for poor COVID-19 outcomes [55], it is not unreasonable to consider the features of the liver if an abdominal CT is acquired alongside the thoracic CT. Hepatic steatosis is an accumulation of fat inside the liver, which is closely linked to obesity, that is found to be of higher prevalence in COVID-19 patients [61]. It is unexplored whether features of this fatty tissue, and the liver as a whole could harbour prognostic imaging features.

Use of Longitudinal Imaging

Acquisition of CT imaging requires exposing the patient to a non-trivial amount of radiation and therefore CT imaging for COVID-19 is only recommended when clinically necessary in the case of seriously ill patients [62] or where the patient condition is worse than would be expected based on the CXR. In our search of the literature, no papers were found which discussed using longitudinal CT imaging and machine learning models. This is potentially due to the fact that the added value of using multiple CT scans to monitor COVID-19 pneumonia has been questioned [63, 64] and found to be of little contribution. Therefore, CTs are not used for routine disease monitoring, and the population will be highly biased. This contrasts with the use of CXR for routine disease monitoring.

Therefore, the patient population with multiple scans is a highly biased cohort with changing serious illness and it is hard to build predictive models which generalise using this.

Fusion with Other Data Modalities

Fusing CT imaging features from another modality is particularly challenging as CT imaging is extremely high-dimensional, commonly with 50–100 million voxels or more. Therefore, to allow for fusion with much lower dimensional data e.g. clinical variables, feature extraction is typically applied to the CT images by calculation of hand-engineered features or use of a CNN to extract learned features. These lower dimensional radiomic features can be easily combined with the lower dimensional clinical data e.g. by concatenation.

In [65], the authors use a convolutional neural network to extract the features from the CT image and encode the symptoms by convolution. These image features are then combined with the encoded clinical features by multiplication and the result is concatenated with the image features. This is then passed through several convolutional layers before a prediction is made.

In [66], the authors use a semi-automated algorithm to segment the lung parenchyma into lobes and then threshold to identify the tissue unaffected by ground-glass opacities (GGO) or consolidation. Two radiologists also graded the images for severity. The authors then build a model using both imaging derived and clinical features, finding that a model using C-reactive protein (an inflammatory marker) along with the imaging features performed better than imaging features alone.

In [67], the authors train a deep learning based segmentation algorithm for GGO and consolidation patterns in CT scans. They fit a model combining radiomic features from these extracted regions with the clinical metadata. This results in a significant boost in performance over using the imaging features alone, increasing the AUC from 0.811 to 0.878.

Common Issues with AI and Computed Tomography Imaging

Computed tomography (CT) imaging is a common 3D imaging modality which is typically used for more complex clinical diagnoses due to the high-resolution of the images, a typical voxel in the images is 0.7 mm × 0.7 mm × 1.0 mm. A CT image is reconstructed from 2D X-ray projections and there is no standard algorithm for performing this reconstruction, therefore the images from different scanners can appear quite different qualitatively. Radiomic features extracted from CT images are known to be sensitive to the reconstruction that is used [68].

CT images are commonly enhanced by adding contrast that absorbs X-rays and is added to exaggerate the difference in intensity between adjacent tissues. Therefore, with intensities vastly different between enhanced and non-enhanced images, radiomic features developed for one are unlikely be applicable to the other. Authors must be clear in manuscripts whether enhanced, non-enhanced or a mixture of these are used to train the models (and in what proportions they occur if the latter).

During the COVID-19 pandemic, CT was used in Europe in the more complex cases of COVID-19 where a patient was likely to benefit from it. Therefore, there is an inherent bias in the population of patients who underwent CT screening as they must be ill enough to justify its use (not mild disease) but not so ill that they are ineligible for a CT scan. Also, in the UK and elsewhere, each time a COVID-19 patient used a CT scanner, the machine and the room required a deep clean for 1–2 h which limited the capacity of hospitals to perform CT scans on COVID-19 patients and hence there is a bias in the data due to this.

CT was used more commonly as a first line imaging modality in China and Russia. Therefore, there are many cases of mild disease within datasets from China and Russia.

Ultrasound Imaging

The diagnosis and treatment of respiratory diseases rely on the use of various imaging modalities. Chest CT is considered the imaging gold standard for pulmonary diseases, as described in section "Computed Tomography Imaging"; however, it is expensive and non-portable. Another standard imaging modality utilized to investigate the lung is chest X-ray, which is discussed at length in section "Chest X-Ray Imaging". Both modalities involve ionizing radiations, which are potentially harmful to the patient. This is particularly significant for specific patient populations such as children, pregnant women, and patients who require repeated examinations over a short period of time. Moreover, CT is generally not available in every hospital nor applicable at bedside, thus requiring patients' mobility. When dealing with a highly infectious disease, this last aspect further increases the risk of contamination within the hospital. Compared to these imaging technologies, ultrasound imaging is safe, cost-effective, more widely available, and transportable, thus has the potential of reaching a much larger population, including non-hospitalized patients. More importantly, there is growing evidence showing that lung ultrasound (LUS) can be used effectively as an imaging modality for pulmonary diseases (e.g., [69–73]).

However, despite all its advantages LUS has not yet taken a major role in point-of-care protocols. One of the reasons hindering wide-spread use of LUS is the difficulty in interpreting it, which is more challenging compared to other imaging modalities, and even with respect to ultrasonography of other organs. Lungs pose a unique challenge for ultrasound as, normally, ultrasound waves are not transmitted through anatomic structures filled with gas. Consequently, the lung parenchyma is not visible beyond the pleura [69]. This phenomenon is shown in Fig. 4.2: While for other organs (left) ultrasound provides detailed visualizations of the organs, when it comes to the lungs (right) the ultrasound waves are not transmitted through the aerated alveoli and thus no anatomical structure below the pleura is visible in LUS.

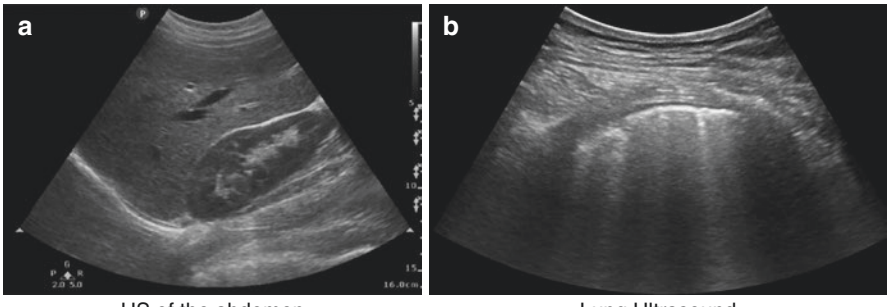


Fig. 4.2 The challenge of lung ultrasound (LUS): (**a**) an ultrasound image of the abdomen, showing the liver and one of the kidneys[a]. (**b**) A LUS frame showing the pleural line as a bright white curve in the middle of the frame. The entire lung cavity below the pleura is not visible and shows only fog-like noise. ([a]image credit quizlet.com/262841714/liverkidney-interface-diagram)

Another key difference between ultrasound and other imaging modalities is the narrow field of view provided by ultrasound. While X-ray and CT can image entire organs and anatomical structures, ultrasound provides only a partial and superficial field of view. As a result, protocols for examining large organs, such as the lungs, require scanning each patient at multiple points, to ensure a complete scan of the organ in question. For LUS protocols vary from, e.g., 6 points [74] to 14 points [72, 75].

What Can be Observed in LUS

Despite the limitations of LUS, it has been observed that although the visual signal below the pleura fails to show any anatomical structures, it still holds valuable clinical information in the shape of the sonographic artifacts visible in the frame. Figure 4.3 exemplifies some of these sonographic artifacts.

For example, in normal aerated lung, *A-lines*, hyperechoic, horizontal lines arising at regular intervals from the pleural line can be seen. Figure 4.3a shows an example of A-lines indicated by a blue arrow. A-lines are reverberation artifacts that arise when the ultrasound beam reflects off of the pleura, instead of being absorbed or transmitted through the aerated lung cavity below it. Multiple reverberations result in multiple A-lines, at multiples of the pleural depth. Observing these horizontal A-lines in LUS indicates a well aerated healthy lung.

A different sonographic artifact are vertical *B-lines*, indicated by yellow arrows in Fig. 4.3b. According to recent developments in LUS [77, 78], vertical artifacts are sonographic signs caused by complex interaction of the multiple scattering phenomena that may form in the presence of an alteration occurring at the lung surface.

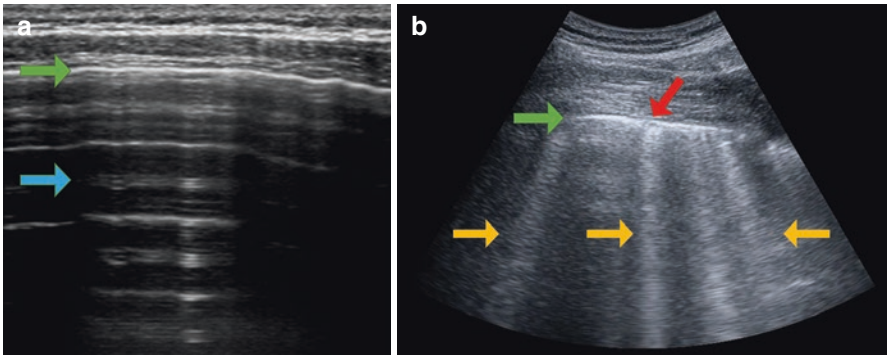


Fig. 4.3 Examples illustrating what can be observed in LUS using either a linear or convex probe. (**a**) The scan shows the pleural line (green) and a sequence of horizontal bright A-lines below it (blue). (**b**) The scan shows the pleural line (green) with some consolidations (red). Several vertical B-lines are visible under the pleural line (yellow). (Figure adapted from [76])

When forming the LUS frame, the ultrasound signals produced by multiple scattering events, in case of resonance phenomena, are interpreted as a bright vertical line emitting from the pleural line and aligned along the ultrasound beam axis [79]. That is, observing B-lines in LUS indicates the presence of fluids just below the pleura. This is, in general, an indication of some pathological condition. The more B-lines observed in a scan or, the wider they are, the more severe is the patient's condition.

Although LUS only shows sonographic artifacts and no anatomical structure below the pleural line, LUS can still be used to inspect the pleural line itself. In healthy patients a smooth and continuous pleural line, as shown in Fig. 4.3a, is usually observed. In contrast, when there are pleural effusions or pulmonary consolidations the pleural line is observed as irregular and discontinuous, as indicated by a red arrow in Fig. 4.3b. Based on these observations, it has been sown that LUS can still be very beneficial clinically. For example, [69] showed how LUS can be used for diagnosing the main lung pathologic entities in patients with ARDS, replacing bedside CXR. More recently, the papers [74, 80] demonstrate that LUS may be used to guide COVID-19 patients' management strategies, as well as resource allocation.

To conclude, its widespread availability and safety are putting LUS in a position to take a more substantial role as a POC modality. The main challenge hindering its use is the difficulty in interpreting the acquired frames, inferring the underlying invisible condition indirectly from the visible sonographic artifacts. This usually requires well-trained and highly specialized radiologists. Nevertheless, recent advancements in AI techniques may assist in bridging the gap by processing and analyzing LUS frames, allowing even novice and inexperienced clinicians to benefit from this ubiquitous POC modality.

Models Assisting in Interpreting LUS

Correctly identifying sonographic artifacts, such as A-lines and B-lines, and correctly locating the pleural line is of great importance when one wishes to evaluate a patient's condition from LUS scans. Consequently, several algorithms were developed to detect, segment, and classify these unique LUS features [79, 81–83], highlighting them to assist clinicians in interpreting LUS scans. Yet, for some of these methods, identifying LUS features is not the end goal but an intermediate stage towards achieving more complicated tasks such as prognosis or diagnosis.

Some algorithms for locating LUS features take advantage of the unique geometric properties of the pleural line, the A- and B-lines and use methods such as Radon-transform [84, 85], dynamic programming [86] or morphological operations [87, 88] to locate them. In contrast, more advanced AI methods rely on training data to accomplish these tasks. These methods range from supervised methods [79, 89–91], to semi-supervised frameworks [90, 92]. These approaches train deep neural networks for either object detection or semantic segmentation to directly infer the location of the various LUS features. Alternatively, [92] proposes to use gradient-weight

class-activation mapping (grad-CAM) [93] on top of a classification network to identify regions of interest in LUS frames.

Ultimately, these approaches automatically highlights the pleural or the A- and B-lines for the clinician during her bed-side examination of the patient, allowing even for a non-expert to benefit from POC LUS modality.

Diagnosis Models

AI can benefit POC LUS beyond just improving human interpretability of the scans. Like in X-ray and CT, AI tools can be used to assist in making differential diagnosis decisions based on LUS. The ultimate goal of these AI tools is assisting in diagnosing a *patient*. This task requires integrating information from several LUS scanning points around the chest, each comprised of many LUS frames. However, most current AI algorithms focus on a less ambitious goal: analyzing only a single LUS frame at a time, leaving the task of integrating the predictions over a sequence of frames and multiple scan points to future works. We focus the discussion in this section and the next one on AI algorithms designed to perform per-frame predictions. In section "Use of Longitudinal Imaging" we cover approaches for integrating these per-frame predictions.

To facilitate training of AI models for the task of differential diagnosis of various pulmonary conditions based on LUS, Born et al. [94] curated the POCUS dataset. This dataset contains LUS records from mainly 3 different classes, COVID-19, bacterial pneumonia and healthy patients. They gathered LUS scans from different online sources and made it available online.[1] The dataset was constantly updated between April 2020 and January 2021. By January 2021, the dataset contained LUS recordings from 216 patients, where the majority of which acquired with convex transducers and the rest with linear probes. Table 4.1 provides more details about the latest version of the POCUS dataset.

Due to the limited number of LUS scans acquired using a linear probes, and LUS scan of viral pneumonia, the majority of the works that used the POCUS dataset focused only on LUS recordings acquired by convex probes and ignored the non-COVID viral pneumonia [94–98]. These works focused on developing deep neural networks for LUS frame classification: either classifying each frame into the corresponding diagnosis or providing a binary decision if the frame presents with COVID-19 associated markers or not. Similar to the CXR literature, few papers develop their own custom architectures [98], while most opt to do transfer learning based on existing trained models, from ResNet and Inception backbones to the more efficient MobileNet [94, 95, 99]. Others utilize those models as feature extractors for support vector machine classifier or other classifiers based on fully connected layers [97]. ResNet and Xception architectures reported better performance than the

[1] https://github.com/jannisborn/covid19_ultrasound

Table 4.1 Current POCUS dataset [94]: number of videos and images per class and probe type

	Convex		Linear	
	Vid.	Img.	Vid.	Img.
COVID-19	64	18	6	4
Bacterial Pneu.	49	20	2	2
Viral Pneu.	3	–	3	–
Healthy	66	15	9	–
Total	**182**	**53**	**20**	**6**

others, with accuracies ranging from 0.83 to 0.99 for frame based diagnosis tasks. However, we caution against direct comparison since the papers not only used different versions of the POCUS dataset but also employed different sampling rate for extracting frames from the dataset's LUS videos.

Prognosis Models

In order to stratify COVID-19 patients using LUS, it has been proposed to score LUS scans according to well-defined physiological findings. Similar scoring scales were independently proposed by [74, 100], suggesting a 4-level scoring system with scores ranging from 0 to 3. Score 0 indicates a healthy lung characterised by a continuous pleural-line and visible A-lines artifacts. In contrast, score 1 indicates first signs of abnormality mostly related to small alterations in the pleural-line, and the appearance of few vertical artifacts. Scores 2 and 3 are representative of a more advanced pathological state, with the presence of small or large consolidations, respectively, and significant presence of vertical artifacts (B-lines and "white lung"). Figure 4.4 shows example frames representative of each score.

In order to facilitate the development of AI systems to automatically score LUS scans according to their clinical severity, Roy et al. [91] curated the Italian COVID-19 Lung Ultrasound dataset (ICLUS).[2] The ICLUS dataset contains 277 LUS videos of 35 patients, from 5 different Italian medical centers with a total of approximately 60,000 frames acquired by convex and linear probes. All frames in the ICLUS dataset were manually annotated into one of the four severity scores. See Table 4.2 for more details. In addition, video level annotation as well as pixel level annotations for the bio-markers indicative of each score, were provided for a subset of the data.

The ICLUS dataset give rise to several AI methods aiming at scoring LUS frames, thus assisting in the prognosis of the disease. Roy et al. [91] utilized the ICLUS dataset for frame-based, video-based and pixel-based severity score prediction. They used a special neural architecture in order to achieve unified processing of LUS frames obtained by either linear or convex probe. In contrast [79] showed

[2] https://iclus-web.bluetensor.ai/

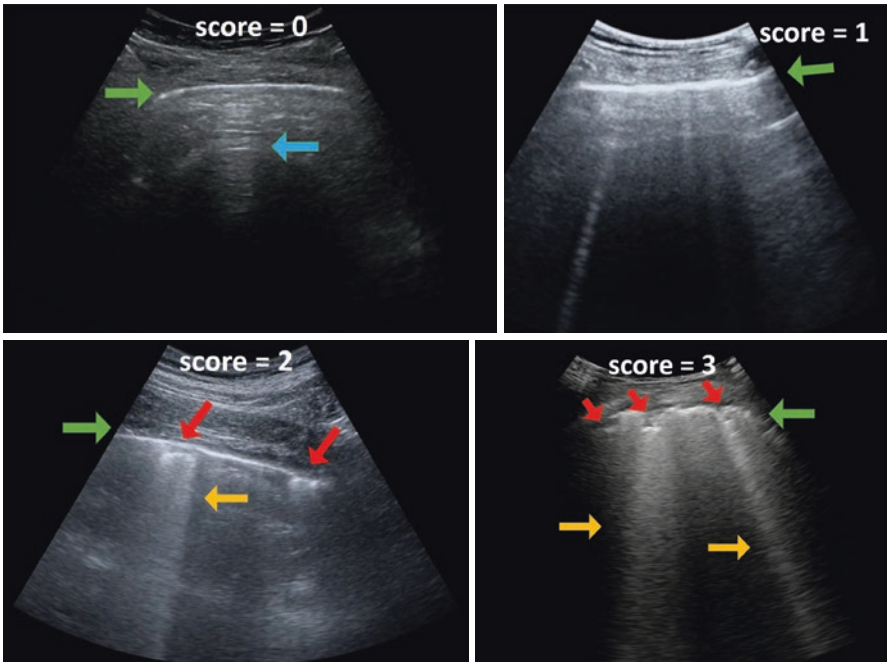


Fig. 4.4 COVID-19 severity scores. LUS frames exemplifying the severity score of [73] from healthy (score = 0, top left) to severe (score = 3, bottom right). One can observe the pleural line (green), A-lines (blue), subpleural consolidations (red) and vertical artifacts (e.g., B-lines and "white lung") (yellow). While the pleural line and consolidations are anatomical features, the A-lines and vertical artifacts are sonographic echoes. (Figure taken from [79])

Table 4.2 ICLUS dataset: number of LUS frames per severity score and probe type [91]

	Convex	Linear	Total
Score = 0	14,690	5283	19,973
Score = 1	11,131	3164	14,295
Score = 2	15,772	3200	18,972
Score = 3	3967	1717	6335
Total	**45,560**	**13,364**	**58,924**

that by injecting domain knowledge into the inputs of standard image classification models, following the model-based AI philosophy [101], one can still handle linear and convex frames in a unified manner. They also achieved a boost in performance on two domain specific tasks: severity classification and segmentation. Specifically, they injected LUS domain knowledge in the form of B-line and pleural line masks along with the raw LUS frame. Another use of LUS features for prognosis purposes was demonstrated in [86]. Localization of the pleural line was used for extraction of

hand-crafted features, such as discontinuities in the pleural line, which in turn were used as an input for a support vector machine classifier of COVID-19 severity score.

A recent dataset comprised over 18,000 LUS frames, acquired by convex probes, from 450 patients (COVID-19 and healthy) was gathered and used for frame based severity score prediction by [102]. The frames where annotated with the 4 scores severity system proposed in [73] and with a new 7 scores severity system, which gives extra weight to pleural line irregularities. Using Resnet-18 and Resnet-50 they achieved F1 score of above 97% for both models and for both scoring systems. However, this dataset was not made publicly available.

Use of Longitudinal Imaging

As explained at the beginning of this section, the ultrasound modality in general and LUS, in particular, provides only a partial and limited field of view. Thus, in order to have a complete and comprehensive assessment of the lungs, clinical studies require 6 [74] to 14 scanning points [72, 75]. These multiple scanning points, in turn, amounts to numerous LUS videos, each comprising hundreds of frames. This is in contrast to other modalities in which each scan provides a comprehensive view of the clinical condition. Therefore, when considering longitudinal imaging in the context of LUS, one needs to account for more fundamental levels of information aggregation: The first is accumulating predictions from individual LUS frames to a coherent estimation of the clinical condition at each scanning point. The second level of aggregation is combining the prediction of each scanning point to provide a coherent forecast for the patient as a whole. We discuss these more fundamental levels of longitudinal imaging in LUS next.

The lack of sufficient LUS videos in the available datasets forced most existing AI methods to focus on frame-level predictions. If made, aggregation is done mainly through naïve averaging of the frames' prediction. One exception was introduced by [91], where a lightweight learned approach based on uninorms was used for aggregation of frame-level predictions into video-level prediction showing better performance compared to simple averaging.

The caveat of most current AI methods for LUS, of processing one LUS frame at a time and then aggregating the predictions, is the discarding of temporal information existing in the LUS video sequence. However, exploiting temporal information can be very beneficial in analyzing LUS. Specifically, such information can be helpful in the identification of B-lines due to their flickering nature induced by the motion of the lungs during respiration. For that aim, the use of 3D convolutions, optical flow [103] or long short-term memory (LSTM) layers [96] where the most common approaches.

Common Issues with AI and Ultrasound Imaging

LUS is acquired using both convex and linear probes. An observation made by [76, 104] suggests that using both in automatic LUS analysis system can be problematic. They noticed that although both anatomic structure and sonographic artifacts can be observed in both probes, the orientation of the sonographic artifacts changes between probe type. While in a linear probe, B-lines appear "axis-aligned", when using a convex probe, B-lines appear "tilted" as if emitted from the focus point of the probe. This difference has little effect on a human observer, but can be confusing for an automatic LUS analysis system. This observation calls for making an explicit adjustment to the various models, either by new neural architectures (e.g., [76, 91]) or by domain-specific pre-processing of the data (e.g., [79]) in order to make these models suitable across probe types. Alternatively, developing dedicated AI systems restricted to a single probe type hinders the use of other probe types' available training data.

Another issue arising in developing AI methods for LUS is the relatively small amount of available training data. Existing datasets, ICLUS [91] and POCUS [94], have only several thousands of LUS frames. When it comes to LUS videos, or indeed multiple scans of patients, the numbers are significantly lower, in the range of only a few hundreds at best. This quantity does not allow, at the moment, the development of elaborated schemes for information integration over frames in a LUS video or between different videos of the same patient. On the other hand, trying to enrich the datasets by curating LUS data from publicly available sources may also be problematic. LUS frames were available online long before the COVID-19 outbreak, thus acquired by older US machines than the COVID-19 ones. As a result, AI systems trained to perform, e.g., diagnosis, on such data may learn to classify idiosyncrasies related to LUS machines rather than actual clinical findings. This might be the case with some of the results reported on the POCUS dataset.

To conclude this section, LUS has the potential of playing a more significant role as a point-of-care modality. Currently, its interpretability is a challenge that allows only expert clinicians to take advantage of LUS effectively. AI models can potentially bridge the gap and make LUS more widely used. The COVID-19 pandemic has drawn attention to this gap and initiated a surge in the development of relevant AI techniques. These AI models are just starting to emerge; hence they tend to be simplistic, working on single frames rather than aggregating information across frames and scanning points to provide a more holistic prediction for each patient. However, considering the ubiquity of the LUS modality vis-a-vis the difficulty and the expertise needed to gain useful clinical information, it seems like AI can significantly bridge this gap.

Conclusions

In this section we summarise findings made from the literature, including the success stories of machine learning for the COVID-19 pandemic, the pitfalls which must be given more focus along with the lessons learned and recommendations.

Success Stories

The COVID-19 pandemic highlighted, in many ways, how unprepared the world was to tackle and respond to a highly infectious virus. In particular, for the machine learning and imaging communities, although we had a plethora of literature regarding chest imaging and machine learning, we did not have a robust method to apply to a new disease quickly, that could be validated rapidly. In addition to the modelling issues, data acquisition was also difficult, with lengthy detailed ethics requests required for each hospital to obtain high quality data—along with commonly requiring data extraction by the very same clinicians who are on the front line treating patients. However, we highlight here several of the success stories of this pandemic which serve to provide a blueprint for the next pandemic response.

Large, rapid dataset collection. Most imaging studies outside of COVID-19 are hampered by having small datasets at different sites with images acquired on different machines. This leads to many biases appearing in the development process. The pandemic presented a unique opportunity to allow researchers to collate imaging data at a scale and pace which is commonly not possible with large datasets assembled for many different instruments. The UK's National COVID-19 Chest Imaging Database (NCCID) assembled by NHSX, the British Society of Thoracic Imaging (BSTI), Royal Surrey NHS Foundation Trust and Faculty is a great example of an initiative to collate data in a systematic way which machine learning researchers can use to develop and validate their algorithms. The NCCID required only one umbrella ethics submission for all centers involved and images were collated at a single center for anonymisation and upload into the cloud. Crucially, the burden of the work fell on non-clinical staff to manage the data collation—rightly keeping more clinicians on the front line.

Managing expectations. The pandemic has highlighted areas which were lacking focus in the machine learning for imaging community, namely how to rapidly learn a new class of data from limited examples. It highlighted also how a short-circuit exists in the research community, whereby access to biased public datasets and widely accessible machine learning frameworks (e.g PyTorch, Tensorflow,

Keras) allowed a large number of researchers to simultaneously fit models fairly easily—without appreciation for the clinical biases in the data, biases in their methodologies and biases in how models had been evaluated. Through systematic reviews such as [12, 13], the clinical and machine learning community is more aware of the issues pervasive in the machine learning literature, in particular for COVID-19, and can make a fairer assessment of the models being developed, keep their expectations grounded in realty and allow them to ask informed questions.

Appreciation of the need to collaborate. Collaborations across disciplines allow for one community to appreciate the others challenges and understand each others limitations. During the COVID-19 pandemic this has been highlighted the tangible benefits of collaboration. Fundamentally, understanding the clinical pathway for COVID-19 patients is critical to allow for development of high quality models to diagnose and prognosticate for the disease. Understanding when and how ventilators are adjusted, the criteria for patients to be admitted to the intensive care units, when and where CXR/CT and US imaging are requested and when followup imaging is required are all essential to allow for appreciation of biases in the datasets and methodologies used. Healthy collaborations allow these to be explored in detail and for clinicians to appreciate the limitations of existing machine learning methodologies.

Pitfalls to Focus On

In this section we focus on some of the common and continuing areas in which we need to learn and improve for the application of machine learning to images in the time of a pandemic.

Willingness to share data and resources. Crucial to any machine learning task is high-quality data and therefore hospitals and other data contributors need to have a willingness to share data with researchers. For imaging, even if there is willingness, an infrastructure to extract, anonymise and share images is also required. Unfortunately, in the COVID-19 pandemic, even with the NCCID initiative, the hospitals who could share data were those who had capacity for clinicians and informatics staff to extract it. This inclusion bias is critical to tackle as it is necessary that the algorithms which are developed can also apply in those resource stretched hospitals which could not provide data. Federated learning techniques, which do not require sharing of imaging data outside of each institution, such as [105, 106] may be critical in the future to ensure that complications due to sharing and moving of data between sites is not a hindrance to research.

Clearly defining the diagnosis control group. Throughout the literature, diagnosis models for COVID-19 based on machine learning methods rely on distinguishing COVID-19 from a control class or classes. However, it is unclear what this control group should be in the context of COVID-19. Commonly, in the higher quality manuscripts, authors have identified different non-COVID-19 viral pneumonias, bacterial pneumonias and COVID-19 pneumonia as separate classes and

trained models to separate these pneumonias. In poorer quality models, healthy patients have been used as controls, or even exclusively paediatric patients with non-COVID-19 pneumonias. The key aspect here is identifying the population in which the model will be applied and ensuring the data used to develop the model reflects this population. For example, if the model is only to be applied by radiologists attempting to distinguish different pneumonias, then training on these alone is reasonable. However, if the model is to be used as a screening tool, the control group should include all different pathologies which are seen in the clinic.

Regulatory considerations. In the literature, many models have been developed and discussed but there is minimal consideration of how these models would pass through regulatory processes to be adopted in the clinic. Many models are described with insufficient documentation to allow for regulation to even be considerable. It is also a reality that regulating software as a medical device is a long and expensive process, which is not only beyond the resources (and skillsets) of many research groups but means that urgent solutions developed during the pandemic cannot hope to be useful in the course of it due to time delays. It is for policy makers to consider how this particular area of regulation could be improved for the next pandemic.

Missing entries in the data. Missing data is a reality of most real-world clinical datasets and most machine learning models require complete training data. Therefore, it is typical to use data imputation techniques to replace this missing data with intelligent guesses. However, the type of missingness affects the quality of this imputation, i.e. whether the data are missing completely at random, missing at random (with missingness dependent on observed values) or missing not at random [107]. It is important to understand the type of missingness in the data used to develop the methods and ensure that this is also expected in the data to which the model will be applied.

Non-standardised variables across sites. When models are developed for use across multiple sites it is imperative that the variables are comparably recorded at each site (both in terms of the concept being encoded and the units). As machine learning progresses towards more federated approaches with multiple institutions, it is important that variables are mapped into a common standard, e.g. using OMOP dictionaries. Not only does this allow for federated approaches to be adopted more easily, but it ensures that models developed at one site can be rapidly applied at other sites.

Lessons Learned and Recommendations

The COVID-19 pandemic has highlighted many issues for the community using machine learning with imaging and clinical data. In this section we will highlight several of the lessons learned and recommendations (many of which are covered in [12]).

Recommendations for study design. It is unfortunate that the same degree of care which is applied to clinical trial design is not applied more extensively to

machine learning. In particular, exploratory analysis on a small dataset (analagous to a phase 1 study), scaling to include more data with more diversity to determine whether the model is identifying signals in the date (phase 2) before allowing for development on the entirety of the dataset (phase 3). It is most common that authors jump to phase 3, short-cutting a lot of important exploratory steps which can identify biases.

Recommendations for data. Public repositories of imaging for COVID-19 patients should be used with extreme caution. Due to the lack of verification procedures to ensure patients are RT-PCR positive or negative, along with the ability for anyone globally to contribute images, this leads to significant risks of bias (e.g. source issues and Frankenstein datasets) as discussed earlier. Authors must also aim to match demographics across cohorts, an often neglected but significant potential source of bias (which may even be impossible with public datasets that do not include demographic information). Many public datasets obtain their images from preprints and published manuscripts which are in low-resolution or compressed formats (e.g. JPEG and PNG), rather than their original DICOM format. If this reduction in resolution is biased across the different image classes, this leads to a serious issue for those models reliant on convolutions and hand engineered features which may simply learn to identify the new resolutions.

For CXRs in particular, researchers should be aware that the view (front-to-back vs. back-to-front) that has been used to acquire that CXR is important as, for example, in sick, immobile patients, an front-to-back CXR view is used for practicality rather than the standard back-to-front CXR projection. The most useful algorithms are those that can diagnose disease at an early stage however many datasets will include an overrepresentation of severe disease which will likely reduce the models applicability.

In the literature, the timing between imaging and RT-PCR tests was also largely undocumented, which has implications for the validity of the ground truth used. A negative RT-PCR test does not necessarily mean that a patient does not have COVID-19 and we must encourage authors to evaluate their algorithms on datasets from the pre-COVID-19 era, such as performed by [108], to validate any claims that the algorithm is isolating COVID-19-specific imaging features. In many papers, it is common for non-COVID-19 diagnoses to be determined from the imaging alone, with those same images used to develop the model. This is known as incorporation bias and leads to an over optimistic model performance.

Recommendations for evaluation. The importance of using a well-curated external validation dataset of appropriate size in order to assess generalizability to other cohorts cannot be overstated. Any useful model for diagnosis or prognostication must be robust enough to give reliable results for any sample from the target population rather than just on the sampled population. Calibration statistics should be calculated for the developed models to inform predictive error and decision curve analysis [109] performed for assessing clinical utility. If a model outputs a prediction of death at $p = 0.6$ vs. $p = 0.8$, clinical judgment is likely to change so it is

important to know how well calibrated the model is. Authors must also disclose how they ensured that images from the same patient were not included in the different dataset partitions, such as describing patient-level splits. It is primarily an issue for datasets containing multiple images from each patient or those which process 3D volumes as independent 2D samples.

It is important to include confidence intervals, when reporting results, to reflect the uncertainty in the estimate, especially when training models on the small sample sizes commonly seen with COVID-19 data. Moreover, it is important and not an onerous task to demonstrate model interpretability. Examples of interpretability techniques include: (1) informing the clinician of which features in the data most influenced the prediction of the model, (2) linking the prognostic features to the underlying biology and (3) overlaying an activation/saliency map on the image to indicate the region of the image which influenced the model's prediction and (4) identifying patients which had a similar clinical pathway. Many papers derive their performance metrics from the test data alone with an unstated operating point to calculate sensitivity and specificity. Clinical judgment should be used to identify the desired sensitivity or specificity of the model and the operating point should be derived from the development data. The differences in the sensitivity and specificity of the model should be recorded separately for the validation and test data.

Recommendations for replicability. It is not possible to reproduce many existing models due to updating of publicly available datasets or codes since the publication of the manuscripts. Therefore, we recommend that a cached version of the public dataset be saved, or the date/version quoted, and specific versions of data or code be appropriately referenced. We acknowledge that although perfect replication is potentially not possible, details such as the seeds used for randomness and the actual partitions of the dataset for training, validation and testing would form very useful supplementary materials. Furthermore, it is necessary that the manuscript states any image resizing, cropping and normalisation used to ensure the work is reproducible.

Recommendations for authors. It is recommended that authors assess their manuscript against appropriate established frameworks, such as RQS, CLAIM, TRIPOD, PROBAST and QUADAS [110–114]. This will ensure reproducibility, and that models are developed in a careful manner.

Recommendations for reviewers. For reviewers, we also recommend the use of the checklists, discussed in the previous point, in order to better identify common weaknesses in reporting the methodology. The most common issues in the papers considered in [12] was the use of biased datasets and/or methodologies. For non-public datasets, it may be difficult for reviewers to assess possible biases if an insufficiently detailed description is given by the authors. We strongly encourage reviewers to ask for clarification from the authors if there is any doubt about bias in the model being considered. Finally, we suggest using reviewers from a combination of both medical and machine learning backgrounds, as they can judge the clinical and technical aspects in different ways.

The Next Pandemic

For the next pandemic, we must be better prepared, with rapid data collection and sharing, and models which are purpose built and trainable to include new classes quickly and robustly. We need an infrastructure to share data at scale, and we need regulatory improvements which allow for validated algorithms to translate into clinic rapidly.

As a community, we need a blueprint for how to optimally respond to the next pandemic. According to the UK National Audit Office, the COVID-19 pandemic has cost the country over £370 billion as of January 2022 [115] with the testing and contact tracing program alone allocated £37 billion [116]. This astonishing amount highlights the trade-off if governments do not invest in pandemic response research and development to ensure we are not taken by surprise again. A particularly encouraging social enterprise is the Trinity Challenge [117] which ran a competition in 2021 for teams to win funding for their ideas "to ensure we are better prepared against health emergencies". There were 8 prize winning teams, all developing tools using machine learning and data analytics to increase global preparedness for future pandemics.

Acknowledgements The authors want to thank the whole AIX-COVNET team (see https://covid19ai.maths.cam.ac.uk) and in particular, Sören Dittmer, Julian Gilbey, Matthew Thorpe, Jacobus Preller, Ian Selby, Effrossyni Gkrania-Klotsas, James Rudd, John Aston and Evis Sala, for the many interesting and fruitful discussions and for their many insights on AI methodologies at the interface to clinical practice which have greatly contributed to this article. The authors also thank all their co-authors in papers [35, 76, 79] who worked with them on AI methodologies for COVID-19 detection using Xray and Ultrasound. MR and CBS also acknowledge support from the DRAGON consortium and Intel. MR further acknowledges support from AstraZeneca, and CBS further acknowledges support from her Royal Society Wolfson fellowship, her Philip Leverhulme prize and EPSRC projects EP/N014588/1, EP/T017961/1. SB is a Robin Chemers Neustein Artificial Intelligence Fellow. YE is supported by Miel de Botton, Jean and Terry de Gunzburg Coronavirus Research, a research grant from the Corona Response Fund, and the Manya Igel Centre for Biomedical Engineering and Signal Processing.

References

1. American College of Radiology. ACR recommendations for the use of chest radiography and computed tomography (CT) for suspected COVID-19 infection. 2020. https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection.
2. The Royal College of Radiologists. The role of CT in patients suspected with COVID-19 infection. 2020. https://www.rcr.ac.uk/college/coronavirus-COVID-19-what-rcr-doing/rcr-position-role-ct-patients-suspected-COVID-19.
3. Cleverley J, Piper J, Jones MM. The role of chest radiography in confirming Covid-19 pneumonia. Br Med J. 2020;370:m2426.

4. Jacob J, Alexander D, Baillie JK, Berka R, Bertolli O, Blackwood J, Joshi I. Using imaging to combat a pandemic: rationale for developing the UK national COVID-19 chest imaging database. Eur Respir J. 2020;56(2):1809.
5. Harvey DM. Example of wallpaper group type p3. Computer-enhanced photograph of a street pavement in Zakopane, Poland. 2005.
6. Driggs D, Selby I, Roberts M, Gkrania-Klotsas E, Rudd JHF, Yang G, Babar J, Sala E, Schönlieb C-B. Machine learning for COVID-19 diagnosis and prognostication: lessons for amplifying the signal while reducing the noise. Radiol Artif Intell. 2021;3(4):e210011.
7. Rajpurkar P, Joshi A, Pareek A, Chen P, Kiani A, Irvin J, Lungren MP. CheXpedition: Investigating generalization challenges for translation of chest X-ray algorithms to the clinical setting. 2020. https://arxiv.org/abs/2002.11379.
8. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, Zhang K. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell. 2018;172(5):1122–31.
9. van Sloun RJG, Cohen R, Eldar YC. Deep learning in ultrasound imaging. Proc IEEE. 2020;108(1):11–29.
10. Mischi M, Lediju Bell MA, van Sloun RJG, Eldar YC. Deep learning in medical ultrasound—from image formation to image analysis. IEEE Trans Ultrason Ferroelectr Freq Control. 2020;67(12):2477–80.
11. Inui S, Fujikawa A, Jitsu M, Kunishima N, Watanabe S, Suzuki Y, Uwabe Y. Chest CT findings in cases from the cruise ship Diamond Princess with coronavirus disease (COVID19). Radiol Cardiothorac Imaging. 2020;2(2):e200110.
12. Roberts M, Driggs D, Thorpe M, Gilbey J, Yeung M, Ursprung S, Schönlieb C-B. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nat Mach Int. 2021;3:199–217.
13. Wynants L, et al. Prediction models for diagnosis and prognosis of COVID-19: systematic review and critical appraisal. BMJ. 2020;369:m1328.
14. Hamzeh A, et al. Artificial intelligence techniques for containment COVID-19 pandemic: a systematic review. Res Sq. 2020;2020:30432. https://doi.org/10.21203/rs.3.rs-30432/v1.
15. Albahri OS, et al. Systematic review of artificial intelligence techniques in the detection and classification of COVID-19 medical images in terms of evaluation and benchmarking: taxonomy analysis, challenges, future solutions and methodological aspects. J Infect Public Health. 2020;13(10):1381–96. https://doi.org/10.1016/j.jiph.2020.06.028.
16. Feng S, et al. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19. IEEE Rev Biomed Eng. 2021;14:4–15.
17. Gillman AG, Lunardo F, Prinable J, Belous G, Nicolson A, Min H, Terhorst A, Dowling JA. Automated COVID-19 diagnosis and prognosis with medical imaging and who is publishing: a systematic review. Phys Eng Sci Med. 2021;45:13–29.
18. Laino ME, Ammirabile A, Posa A, Cancian P, Shalaby S, Savevski V, Neri E. The applications of artificial intelligence in chest imaging of COVID-19 patients: a literature review. Diagnostics. 2021;11(8):1317.
19. Bottino F, Tagliente E, Pasquini L, Napoli AD, Lucignani M, Figà-Talamanca L, Napolitano A. COVID mortality prediction with machine learning methods: a systematic review and critical appraisal. J Pers Med. 2021;11(9):893.
20. Ahuja V, Nair LV. Artificial intelligence and technology in COVID era: a narrative review. J Anaesthesiol Clin Pharmacol. 2021;37(1):28.
21. Aishwarya T, Kumar VR. Machine learning and deep learning approaches to analyze and detect COVID-19: a review. SN Comput Sci. 2021;2(3):1–9.
22. Alsharif MH, Alsharif YH, Yahya K, Alomari OA, Albreem MA, Jahid A. Deep learning applications to combat the dissemination of COVID-19 disease: a review. Eur Rev Med Pharmacol Sci. 2020;24:11455–60.

23. Benameur N, Mahmoudi R, Zaid S, Arous Y, Hmida B, Bedoui MH. SARS-CoV-2 diagnosis using medical imaging techniques and artificial intelligence: a review. Clin Imaging. 2021;76:6–14.
24. Bhargava A, Bansal A. Novel coronavirus (COVID-19) diagnosis using computer vision and artificial intelligence techniques: a review. Multimed Tools Appl. 2021;80(13):19931–46.
25. Hariri W, Narin A. Deep neural networks for COVID-19 detection and diagnosis using images and acoustic-based techniques: a recent review. Soft Comput. 2021;25(24):15345–62.
26. Heidari A, Navimipour NJ, Unal M, Toumaj S. The COVID-19 epidemic analysis and diagnosis using deep learning: a systematic literature review and future directions. Comput Biol Med. 2021;141:105141.
27. Kaur J, Kaur P. Outbreak COVID-19 in medical image processing using deep learning: a state-of-the-art review. Arch Comput Methods Eng. 2021;29:2351–82.
28. Mohammad-Rahimi H, Nadimi M, Ghalyanchi-Langeroudi A, Taheri M, Ghafouri-Fard S. Application of machine learning in diagnosis of COVID-19 through X-ray and CT images: a scoping review. Front Cardiovasc Med. 2021;8:185.
29. Mondal M, Bharati S, Podder P. Diagnosis of COVID-19 using machine learning and deep learning: a review. Curr Med Imaging. 2021;17(12):1403–18. https://arxiv.org/abs/2110.14910
30. Montazeri M, ZahediNasab R, Farahani A, Mohseni H, Ghasemian F. Machine learning models for image-based diagnosis and prognosis of COVID-19: systematic review. JMIR Med Inform. 2021;9(4):e25181.
31. Ozsahin I, Sekeroglu B, Musa MS, Mustapha MT, Uzun Ozsahin D. Review on diagnosis of COVID-19 from chest CT images using artificial intelligence. Comput Math Methods Med. 2020;2020:9756518.
32. Rezaei M, Shahidi M. Zero-shot learning and its applications from autonomous vehicles to COVID-19 diagnosis: a review. Intell Based Med. 2020;3:100005.
33. Soomro TA, Zheng L, Afifi AJ, Ali A, Yin M, Gao J. Artificial intelligence (AI) for medical imaging to combat coronavirus disease (COVID-19): a detailed review with direction for future research. Artif Intell Rev. 2021;55(2):1409–39.
34. Salehi AW, Baglat P, Gupta G. Review on machine and deep learning models for the detection and prediction of Coronavirus. Mater Today Proc. 2020;33:3896–901.
35. Keidar D, Yaron D, Goldstein E, Shachar Y, Blass A, Charbinsky L, Aharony I, Lifshitz L, Lumelsky D, Neeman Z, Mizrachi M, Hajouj M, Eizenbach N, Sela E, Weiss C, Levin P, Benjaminov O, Bachar G, Tamir S, Rapson Y, Suhami D, Atar E, Dror A, Bogot N, Grubstein A, Shabshin N, Elyada Y, Eldar YC. COVID-19 classification of X-ray images using deep neural networks. Eur Radiol. 2021;31:9654–63.
36. Sriram A, Muckley M, Sinha K, Shamout F, Pineau J, Geras KJ, Azour L, Aphinyanaphongs Y, Yakubova N, Moore W. COVID-19 prognosis via self-supervised representation learning and multi-image prediction. 2020. https://arxiv.org/abs/2101.04909.
37. M Shu, RS Bowen, C Herrmann, G Qi, M Santacatterina, R Zabih. Proceedings of the IEEE/CVF international conference on computer vision (ICCV). 2021. pp. 4046–4055.
38. Huang SC, Pareek A, Seyyedi S, et al. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. NPJ Digit Med. 2020;3:136. https://doi.org/10.1038/s41746-020-00341-z.
39. Aljouie AF, Almazroa A, Bokhari Y, Alawad M, Mahmoud E, Alawad E, Alsehawi A, Rashid M, Alomair L, Almozaai S, Albesher B, Alomaish H, Daghistani R, Alharbi NK, Alaamery M, Bosaeed M, Alshaalan H. Early prediction of COVID-19 ventilation requirement and mortality from routinely collected baseline chest radiographs, laboratory, and clinical data with machine learning. J Multidiscip Healthc. 2021;30(14):2017–33. https://doi.org/10.2147/JMDH.S322431.
40. Kwon YJ, Toussie D, Finkelstein M, Cedillo MA, Maron SZ, Manna S, Voutsinas N, et al. Combining initial radiographs and clinical variables improves deep learning prognostication in patients with COVID-19 from the emergency department. Radiol Artif Intell. 2020;3(2):e200098.

41. Cohen JP, Morrison P, Dao L. COVID-19 image data collection. 2020. http://arxiv.org/abs/2003.11597.
42. Radiopaedia. COVID-19: radiology reference article. https://radiopaedia.org/articles/covid-19-4?lang=gb. Accessed 29 July 2020.
43. COVID-19 Database (SIRM). https://www.sirm.org/en/category/articles/covid-19-database/. Accessed 29 July 2020.
44. CORONACASES.org (RAIOSS.com). https://coronacases.org/. Accessed 30 July 2020.
45. Eurorad (ESR). https://www.eurorad.org/. Accessed 29 July 2020.
46. Chest X-Ray Images (Pneumonia) (Kaggle). https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia. Accessed 29 July 2020.
47. RSNA. Pneumonia Detection Challenge (Kaggle). https://www.kaggle.com/c/rsna-pneumonia-detection-challenge. Accessed 29 July 2020.
48. Irvin J, et al. CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison. Proc AAAI Conf Artif Intell. 2019;33:590–7.
49. Maguolo, G. & Nanni, L. A critic evaluation of methods for COVID-19 automatic detection from X-ray images. 2020. http://arxiv.org/abs/2004.12823.
50. Chowdhury MEH, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, Islam KR, Khan MS, Iqbal A, Al-Emadi N, Reaz MBI, Islam MT. Can AI help in screening viral and COVID-19 pneumonia? IEEE Access. 2020;8:132665–76.
51. Zhang C, Bengio S, Hardt M, Recht B, Vinyals O. Understanding deep learning requires rethinking generalization. In: 5th International conference on learning representations ICLR. Cham: Springer; 2017.
52. Deng J, et al. ImageNet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. Berlin: Springer; 2010. p. 248–55. https://doi.org/10.1109/cvpr.2009.5206848.
53. Acar E, Şahin E, Yilmaz İ. Improving effectiveness of different deep learning-based models for detecting COVID-19 from computed tomography (CT) images. medRxiv. 2020;2020:9643. https://doi.org/10.1101/2020.06.12.20129643.
54. Goodfellow IJ, et al. Generative adversarial nets. In: Proceedings of the 27th international conference on neural information processing systems, vol 2. Cambridge: MIT Press; 2014. p. 2672–80.
55. Bae S, Kim SR, Kim M, et al. Impact of cardiovascular disease and risk factors on fatal outcomes in patients with COVID-19 according to age: a systematic review and meta-analysis. Heart. 2021;107:373–80.
56. Grodecki K, Lin A, Razipour A, Cadet S, McElhinney PA, Chan C, Pressman BD, Julien P, Maurovich-Horvat P, Gaibazzi N, Thakur U, Mancini E, Agalbato C, Menè R, Parati G, Cernigliaro F, Nerlekar N, Torlasco C, Pontone G, Slomka PJ, Dey D. Epicardial adipose tissue is associated with extent of pneumonia and adverse outcomes in patients with COVID-19. Metabolism. 2021;115:154436. https://doi.org/10.1016/j.metabol.2020.154436.
57. Jevnikar M, Sanchez O, Chocron R, Andronikof M, Raphael M, Meyrignac O, Fournier L, Montani D, Planquette B, Soudani M, Boucly A. Prevalence of pulmonary embolism in patients with COVID 19 at the time of hospital admission. Eur Respir J. 2021;58(1):2100116.
58. Baeza S, Domingo R, Salcedo-Pujantell M, Deportós J, Moragas G, Garcia-Olivé I, Sanchez C, Gil D, Rosell A. Artificial intelligence to optimize pulmonary embolism diagnosis during Covid-19 pandemic by perfusion SPECT/CT, a pilot study. Eur Respir J. 2021;58(suppl 65):PA359. https://doi.org/10.1183/13993003.congress-2021.PA359.
59. Stancil IT, Michalski JE, Davis-Hall D, et al. Pulmonary fibrosis distal airway epithelia are dynamically and structurally dysfunctional. Nat Commun. 2021;12:4566. https://doi.org/10.1038/s41467-021-24853-8.
60. Roberts M, Kirov K, Mclellan T, Morgan E, Kanavati F, Gallagher D, Molyneaux P, Schönlieb C-B, Ruggiero A, Thillai M. Late Breaking Abstract—Fully automated airway measurement correlates with radiological disease progression in idiopathic pulmonary fibrosis. Eur Respir J. 2021;58(suppl 65):OA3951. https://doi.org/10.1183/13993003.congress-2021.OA3951.

61. Medeiros AK, Barbisan CC, Cruz IR, de Araújo EM, Libânio BB, Albuquerque KS, Torres US. Higher frequency of hepatic steatosis at CT among COVID-19-positive patients. Abdom Radiol (New York). 2020;45(9):2748–54. https://doi.org/10.1007/s00261-020-02648-7.
62. https://www.bsti.org.uk/media/resources/files/BSTI_COVID-19_Radiology_Guidance_version_2_16.03.20.pdf.
63. Chen L, Wang Q, Wu H, Hu J, Zhang J. Repeat chest CT scans in moderate-to-severe patients' management during the COVID-19 pandemic: observations from a single centre in Wuhan, China. Radiat Prot Dosimetry. 2020;190(3):269–75. https://doi.org/10.1093/rpd/ncaa106.
64. Gao Y, Hu Y, Zhu J, Liu H, Qiu R, Lin Q, He X, Lin HB, Cheng S, Li G. The value of repeated CT in monitoring the disease progression in moderate COVID-19 pneumonia. Medicine. 2021;100(10):e25005. https://doi.org/10.1097/MD.0000000000025005.
65. Huang Z, Liu X, Wang R, Zhang M, Zeng X, Liu J, Yang Y, Liu X, Zheng H, Liang D, Hu Z. FaNet: fast assessment network for the novel coronavirus (COVID-19) pneumonia based on 3D CT imaging and clinical symptoms. Appl Intell. 2021;51(5):2838–49.
66. Burian E, Jungmann F, Kaissis GA, Lohöfer FK, Spinner CD, Lahmer T, Treiber M, Dommasch M, Schneider G, Geisler F, Huber W, Protzer U, Schmid RM, Schwaiger M, Makowski MR, Braren RF. Intensive care risk estimation in COVID-19 pneumonia based on clinical and imaging parameters: experiences from the Munich cohort. J Clin Med. 2020;9:1514. https://doi.org/10.3390/jcm9051514.
67. Zhang K, Liu X, Shen J, Li Z, Sang Y, Wu X, Zha Y, Liang W, Wang C, Wang K, Ye L. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. Cell. 2020;181(6):1423–33.
68. Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of a method to compensate multicenter effects affecting CT radiomics. Radiology. 2019;291(1):53–9.
69. Lichtenstein D, Goldstein I, Mourgeon E, Cluzel P, Grenier P, Rouby J. Comparative diagnostic performances of auscultation, chest radiography, and lung ultrasonography in acute respiratory distress syndrome. J Am Soc Anesthesiol. 2004;100(1):9–15.
70. Allinovi M, Parise A, Giacalone M, Amerio A, Delsante M, Odone A, Franci A, Gigliotti F, Amadasi S, Delmonte D, et al. Lung ultrasound may support diagnosis and monitoring of COVID-19 pneumonia. Ultrasound Med Biol. 2020;46:2908–17.
71. Kameda T, Mizuma Y, Taniguchi H, Fujita M, Taniguchi N. Point-of-care lung ultrasound for the assessment of pneumonia: a narrative review in the COVID-19 era. J Med Ultrason. 2021;48:1–13.
72. Smargiassi A, Soldati G, Torri E, Mento F, Milardi D, Giacomo P, De Matteis G, Burzo M, Larici A, Pompili M, et al. Lung ultrasound for COVID-19 patchy pneumonia: extended or limited evaluations? J Ultrasound Med. 2020;40(3):521–8.
73. Soldati G, Smargiassi A, Inchingolo R, Buonsenso D, Perrone T, Briganti D, Perlini S, Torri E, Mariani A, Mossolani E, et al. Proposal for international standardization of the use of lung ultrasound for patients with COVID-19: a simple, quantitative, reproducible method. J Ultrasound Med. 2020;39(7):1413–9.
74. Lichter Y, Topilsky Y, Taieb P, Banai A, Hochstadt A, Merdler I, Oz A, Vine J, Goren O, Cohen B, et al. Lung ultrasound predicts clinical course and outcomes in COVID-19 patients. Intensive Care Med. 2020;46(10):1873–83.
75. Mento F, Perrone T, Macioce V, Tursi F, Buonsenso D, Torri E, Smargiassi A, Inchingolo R, Soldati G, Demi L. On the impact of different lung ultrasound imaging protocols in the evaluation of patients affected by coronavirus disease 2019: how many acquisitions are needed? J Ultrasound Med. 2020;40(10):2235–8.
76. Yaron D, Keidar D, Goldstein E, Shachar Y, Blass A, Frank O, Schipper N, Shabshin N, Grubstein A, Suhami D, Bogot NR, Weiss CS, Sela E, Dror AA, Vaturi M, Mento F, Torri E, Inchingolo R, Smargiassi A, Soldati G, Perrone T, Demi L, Galun M, Bagon S, Elyada YM, Eldar YC. Point of care image analysis for COVID-19. In: IEEE International conference on acoustics, speech and signal processing. Berlin: Springer; 2021.

77. Mento F, Demi L. On the influence of imaging parameters on lung ultrasound B-line artifacts, in vitro study. J Acoust Soc Am. 2020;148(2):975–83.
78. Demi L, Demi M, Prediletto R, Soldati G. Real-time multi-frequency ultrasound imaging for quantitative lung ultrasound—first clinical results. J Acoust Soc Am. 2020;148(2):998–1006.
79. Frank O, Schipper N, Vaturi M, Soldati G, Smargiassi A, Inchingolo R, Torri E, Perrone T, Mento F, Demi L, Galun M, Eldar YC, Bagon S. Integrating domain knowledge into deep networks for lung ultrasound with applications to COVID-19. IEEE Trans Med Imaging. 2021;41(3):571–81.
80. Perrone T, Soldati G, Padovini L, Fiengo A, Lettieri G, Sabatini U, Gori G, Lepore F, Garolfi M, Palumbo I, et al. A new lung ultrasound protocol able to predict worsening in patients affected by severe acute respiratory syndrome coronavirus 2 pneumonia. J Ultrasound Med. 2020;40(8):1627–35.
81. McDermott C, Lacki M, Sainsbury B, Henry J, Filippov M, Rossa C. Sonographic diagnosis of COVID-19: a review of image processing for lung ultrasound. Front Big Data. 2021;4:2.
82. Mento F, Soldati G, Prediletto R, Demi M, Demi L. Quantitative lung ultrasound spectroscopy applied to the diagnosis of pulmonary fibrosis: the first clinical study. IEEE Trans Ultrason Ferroelectr Freq Control. 2020;67:2265–73.
83. Wang Y, Zhang Y, He Q, Liao H, Luo J. A semi-automatic ultrasound image analysis system for the grading diagnosis of COVID-19 pneumonia. 2021. https://arxiv.org/abs/2111.02676.
84. Anantrasirichai N, Hayes W, Allinovi M, Bull D, Achim A. Line detection as an inverse problem: application to lung ultrasound imaging. IEEE Trans Med Imaging. 2017;36(10):2045–56.
85. Karakuş O, Anantrasirichai N, Aguersif A, Silva S, Basarab A, Achim A. Detection of line artifacts in lung ultrasound images of COVID-19 patients via nonconvex regularization. IEEE Trans Ultrason Ferroelectr Freq Control. 2020;67(11):2218–29.
86. Carrer L, Donini E, Marinelli D, Zanetti M, Mento F, Torri E, Smargiassi A, Inchingolo R, Soldati G, Demi L, et al. Automatic pleural line extraction and COVID-19 scoring from lung ultrasound data. IEEE Trans Ultrason Ferroelectr Freq Control. 2020;67(11):2207–17.
87. Moshavegh R, Hansen K, Sørensen H, Hemmsen M, Ewertsen C, Nielsen M, Jensen J. Novel automatic detection of pleura and B-lines (comet-tail artifacts) on in vivo lung ultrasound scans. Medical imaging 2016: ultrasonic imaging and tomography. Int Soc Opt Photonics. 2016;9790:97900K.
88. Brusasco C, Santori G, Bruzzo E, Trò R, Robba C, Tavazzi G, Guarracino F, Forfori F, Boccacci P, Corradi F. Quantitative lung ultrasonography: a putative new algorithm for automatic detection and quantification of B-lines. Crit Care. 2019;23:1–7.
89. Kulhare S, Zheng X, Mehanian C, Gregory C, Zhu M, Gregory K, Xie H, Jones J, Wilson B. Ultrasound-based detection of lung abnormalities using single shot detection convolutional neural networks. In: Simulation, image processing, and ultrasound systems for assisted diagnosis and navigation. Cham: Springer; 2018. p. 65–73.
90. Mason H, Cristoni L, Walden A, Lazzari R, Pulimood T, Grandjean L, Wheeler-Kingshott C, Hu Y, Baum Z. Lung ultrasound segmentation and adaptation between COVID-19 and community-acquired pneumonia. In: International workshop on advances in simplifying medical ultrasound. Cham: Springer; 2021. p. 45–53.
91. Roy S, Menapace W, Oei S, Luijten B, Fini E, Saltori C, Huijben I, Chennakeshava N, Mento F, Sentelli A, et al. Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound. IEEE Trans Med Imaging. 2020;39(8):2676–87.
92. Sloun R, Demi L. Localizing B-lines in lung ultrasonography by weakly supervised deep learning, in-vivo results. IEEE J Biomed Health Inform. 2019;24:957–64.
93. Selvaraju R, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. Berlin: Springer; 2017.
94. Born J, Wiedemann N, Cossio M, Buhre C, Brändle G, Leidermann K, Aujayeb A, Moor M, Rieck B, Borgwardt K. Accelerating detection of lung pathologies with explainable ultrasound image analysis. Appl Sci. 2021;11(2):672.

95. Awasthi N, Dayal A, Cenkeramaddi L, Yalavarthy P. Mini-COVIDNet: efficient lightweight deep neural network for ultrasound based point-of-care detection of COVID-19. IEEE Trans Ultrason Ferroelectr Freq Control. 2021;68:2023–37.
96. Barros B, Lacerda P, Albuquerque C, Conci A. Pulmonary Covid-19: learning spatiotemporal features combining cnn and lstm networks for lung ultrasound video classification. Sensors. 2021;21:5486.
97. Al-Jumaili S, Duru A, Uçan O. Covid-19 ultrasound image classification using SVM based on kernels deduced from convolutional neural network. In: 2021 5th International symposium on multidisciplinary studies and innovative technologies (ISMSIT). Berlin: Springer; 2021. p. 429–33.
98. Muhammad G, Hossain M. COVID-19 and non-COVID-19 classification using multi-layers fusion from lung ultrasound images. Inf Fusion. 2021;72:80–8.
99. Diaz-Escobar J, Ordóñez-Guillén N, Villarreal-Reyes S, Galaviz-Mosqueda A, Kober V, Rivera-Rodriguez R, Lozano Rizk J. Deep-learning based detection of COVID-19 using lung ultrasound imagery. PLoS One. 2021;16:e0255886.
100. Soldati G, Smargiassi A, Inchingolo R, Buonsenso D, Perrone T, Briganti D, Perlini S, Torri E, Mariani A, Mossolani E, et al. Is there a role for lung ultrasound during the COVID-19 pandemic? J Ultrasound Med. 2020;39(7):1459–62.
101. Shlezinger N, Whang J, Eldar YC, Dimakis AG. Model-based deep learning. IEEE Trans Signal Process. 2022;2022:3403.
102. La Salvia M, Secco G, Torti E, Florimbi G, Guido L, Lago P, Salinaro F, Perlini S, Leporati F. Deep learning and lung ultrasound for Covid-19 pneumonia detection and severity classification. Comput Biol Med. 2021;136:104742.
103. Krishnaswamy D, Ebadi S, Bolouri S, Zonoobi D, Greiner R, Meuser-Herr N, Jaremko J, Kapur J, Noga M, Punithakumar K, et al. A novel machine learning-based video classification approach to detect pneumonia in COVID-19 patients using lung ultrasound. Int J Noncommun Dis. 2021;6:69.
104. Bagon S, Galun M, Frank O, Schipper N, Vaturi M, Zalcberg G, Soldati G, Smargiassi A, Inchingolo R, Torri E, Perrone T, Mento F, Demi L, Eldar YC. Assessment of COVID-19 in lung ultrasound by combining anatomy and sonographic artifacts using deep learning. J Acoust Soc Am. 2020;148(4):2736.
105. Bai X, Wang H, Ma L, et al. Advancing COVID-19 diagnosis with privacy-preserving collaboration in artificial intelligence. Nat Mach Intell. 2021;3:1081–9. https://doi.org/10.1038/s42256-021-00421-z.
106. Gafni T, Shlezinger N, Cohen K, Eldar YC, Poor HV. Federated learning: a signal processing perspective. IEEE Signal Process Mag. 2022;39(3):14–41.
107. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med. 2011;30(4):377–99.
108. Banerjee I, et al. Was there COVID-19 back in 2012? Challenge for AI in diagnosis with similar indications. 2020. https://arxiv.org/abs/2006.13262.
109. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565–74.
110. Lambin P, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14:749–62.
111. Mongan J, Moy L, Kahn CE. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell. 2020;2:e200029.
112. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162:55–63.
113. Wolff RF, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170:51.

114. Whiting PF, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529.
115. https://www.nao.org.uk/covid-19/cost-tracker/. Accessed 31 Jan 2022.
116. https://committees.parliament.uk/publications/4976/documents/50058/default/. Accessed 31 Jan 2022.
117. https://thetrinitychallenge.org/. Accessed 31 Jan 2022.

Chapter 5
Machine Learning and Laboratory Values in the Diagnosis, Prognosis and Vaccination Strategy of COVID-19

Anna Carobene, Lorenzo Famiglini, Eleonora Sabetta, Assunta Naclerio, and Giuseppe Banfi

Introduction

In March 2020, the World Health Organization (WHO) declared a pandemic status for a new disease, called COVID-19, caused by infection with the novel beta-coronavirus SARS-CoV-2. Nearly 2 years later, there have been more than two hundred 70 million cases of infection, approximately five and a half million deaths worldwide, and more than eight billion vaccine doses have been administered [1].

In the management of the state of emergency, early diagnosis undoubtedly plays a key role both for patients affected by COVID-19, whose prognosis may benefit from early treatment, and identifying of infected but asymptomatic or pauci-symptomatic subjects, which is essential in the containment of contagions [2].

To date, the gold-standard tool for the diagnosis of SARS-CoV-2 infection is the amplification by rRT-PCR (Reverse Transcription Polymerase Chain Reaction) of viral genomic material (RNA) taken from the upper airways, by oro-pharyngeal and/or nasopharyngeal swab [3]. However, this molecular approach has significant

A. Carobene
IRCCS San Raffaele Scientific Institute, Milan, Italy
e-mail: carobene.anna@hsr.it

L. Famiglini
DISCo, Università Degli Studi di Milano-Bicocca, Milan, Italy
e-mail: l.famiglini@campus.unimib.it

E. Sabetta · A. Naclerio
Università Vita-Salute San Raffaele, Milan, Italy
e-mail: e.sabetta@studenti.unisr.it; a.naclerio@studenti.unisr.it

G. Banfi (✉)
Università Vita-Salute San Raffaele, Milan, Italy

IRCCS Istituto Ortopedico Galeazzi, Milan, Italy
e-mail: banfi.giuseppe@hsr.it

N. Lidströmer, Y. C. Eldar (eds.), *Artificial Intelligence in Covid-19*,
https://doi.org/10.1007/978-3-031-08506-2_5

limitations due to the method's sensitivity, estimated to be around 85% and the time and cost required for the analysis [4].

To improve the pandemic management, the need for enhanced diagnostic capacity for SARS-CoV-2 infections soon emerged, with rapid, accurate, and easily accessible methods. To address this need, some researchers focused their efforts on developing machine learning (ML) models that could help define the diagnosis, and in some cases even the prognosis, of patients with COVID-19 [5].

ML is a branch of artificial intelligence (AI) . It is a term that refers to several computational methods that allow the machine to learn from experience and develop algorithms on a data set, providing it with the ability to perform tasks [6–8]. There is no doubt that IA/ML techniques are finding wide space in all fields of medicine, as shown by the exponential growth of publications in IA/ML/Deep Learning (DL) in recent years (from 203 papers published in 2005, up to 12,563 papers in 2019 indexed on PubMed and more than 31,000 in 2021) [9].

Among the specialties in which IA/ML applications find more space, radiology, oncology, and surgery certainly stand out, fields in which models are developed with the support of imaging, and, as for cardiology and neurology, also of cardiological and neurological diagnostic examinations. In contrast, as Ronzio et al. [10] point out, laboratory medicine is still not well represented.

However, even in this area, the number of publications is growing. Indeed, if the review by Cabitza et al., published in 2018 [11] had identified 37 ML papers published between 2007 and 2017 in the laboratory medicine field, of which only three were published in medical journals, the review by Ronzio et al. published in 2021 identifies 44 papers published in only 4 years, between 2017 and 2020, 71% of which were published in the last 2 years [10]. Let's consider the number of publications in the previous 10 years, not only on ML models but more generally on AI studies with laboratory medicine data, the increase of papers follows an exponential trend [12].

Herman et al. [13] have recently reviewed ML systems that are currently in clinical laboratory practice or are being proposed for such use in recent literature, ML systems that use laboratory data outside the clinical laboratory, challenges to the adoption of ML, and future opportunities for ML in laboratory medicine. Authors described how AI and ML have and will continue to influence the practice and scope of laboratory medicine dramatically, pointing out how this has been made possible by advancements in modern computing and the widespread digitization of health information. These technologies are being rapidly developed and described, but in comparison, their implementation thus far has been modest. The concern of Herman and co-authors is that the participation of the clinical laboratory community is essential to ensure that laboratory data are sufficiently available and incorporated conscientiously into robust, safe, and clinically effective ML supported clinical diagnostics [13].

Early ML models developed for diagnostic purposes for COVID-19 are based on computed tomography (CT) or chest radiography data, often combined with conventional molecular diagnostic findings [5, 14]. Promising results have emerged from these studies but also important critical issues, mainly due to the high number

of false negatives obtained with chest radiography or the impossibility of using CT for screening considering, for example, the high radiation dose, high costs, and a limited number of available instruments [15].

Efforts were then focused on the development of ML models based on the results of blood tests routinely performed in the biomedical laboratory, starting from the scientific evidence that some blood parameters are significantly altered in patients with COVID-19 and can therefore be used as good markers of disease [16–19].

Laboratory examinations have the advantage of having reduced time and costs and are minimally invasive for the patient, allowing their repetition at close intervals.

Due to the low cost and no need for specific assay equipment, laboratory test remains a simple, accessible, near real-time, and cost-effective biomarker, representing the basic routine blood examinations, usually also available in low-resource settings (such as in some developing countries or in a geographical area that is, in the meantime, affected by other socio-sanitary and humanitarian emergencies). Moreover, it is possible to obtain an enormous amount of information from a laboratory test; accordingly, laboratory medicine represents an excellent field of application for ML systems [20].

Therefore, in this context, there has been a strong interest in exploiting the potential of ML-based on hematochemical parameters to obtain new accurate, rapid, inexpensive, and easily accessible tools to improve the management of COVID-19 patients. The development of these approaches in the current health emergency aims to meet the need to have valuable models not only for early diagnosis, supporting the clinician in the management of the patient and allowing a timely therapeutic intervention but also for prognostic/predictive purposes, i.e. in the prediction of the evolution of the disease to identify those patients at higher risk of serious adverse events and therefore to be subjected to closer monitoring and to be started to a more aggressive treatment [11, 21, 22].

Another aspect to consider about the infection with SARS-CoV-2 regards the induced antibody response targeting multiple antigens that changes over time. While not suitable for diagnosing clinical cases, serology is a promising tool for identifying individuals with a previous infection by detecting antibodies generated in response to SARS-CoV-2 infection [23]. Antibody levels are not constant and change over time, so the utility of serological testing depends on the kinetics of the anti-SARS-CoV-2 antibody response during and after infection. Taking advantage of this complexity, a mathematical model of antibody kinetics and ML classifiers to identify serological signatures of SARS-CoV-2 disease generated using multiplex assays has been recently applied [23].

However, there are some problems in recognizing simple patterns in large, complex data sets. It is critical not to make false classifications that lead to incorrect diagnosis and treatment decisions [24]. Many factors can influence the representativeness of datasets obtained from hematochemical parameters for assessing the robustness of a model: differences in testing equipment (cf. the concepts of harmonization), in reference ranges/ethnic variability, in disease manifestations/phenotypic variability and how humans react to contextual factors (cf., biological variation) make the reference population incredibly vast and various, from which

also very different datasets can be drawn to challenge the model's performance [24, 25]. This could explain why ML based on lab data has not yet caught on with other medical disciplines [11].

In this regard, some evidence shows the use of the only CBC data in algorithms developed by ML for diagnostic and prognostic purposes as appropriate for which the leading aspect does not reside on the practicability and economic issues, even, if necessary, but also and especially on the robustness of the model based on its reproducibility [26–30]. And the CBC data compared to the other hematochemical parameters are characterized by a restrained within and between-subjects biological variation [31–34], which supports the reproducibility of results for the same patient at different times, and by a negligible analytical variation [35] which guarantees the reproducibility of the same data between laboratories in various settings, through other equipment, or on heterogeneous populations.

The purpose of this chapter is to consider the multiple published studies on ML approaches developed in this area since the beginning of the pandemic to date, focusing on the heterogeneity in terms of selection of input data and accuracy of results obtained from the different models produced. To determine this heterogeneity, there are some aspects, such as the importance of the selection and standardization of input data or the external validation of the model, which are essential in applying ML in laboratory medicine. The robustness and accuracy of the developed model depend, in fact, on these aspects, to the point of making it or not a valuable tool in clinical practice.

COVID-19, Machine Learning and Laboratory Values: The State of the Art

This chapter considers only studies combining laboratory test results with ML models. To identify which ML models are most effective as a diagnostic and prognostic support tool for COVID-19 has been reviewed the literature present in PubMed (a free search engine of biomedical scientific literature that is continuously updated [9]) and in Scopus (a database of abstracts and citations for articles of research-related publications, created in 2004 by the Elsevier publishing house [36]).

The search was carried out up to 28/12/2021, filtering the results for the period 2020–2021

- PubMed: the following keywords were included in the site's search engine: Title/Abstract ("blood tests" OR "blood exams" OR "laboratory tests" OR "laboratory exams") AND ("COVID-19" OR "COVID" OR "SARS-CoV-2" OR "coronavirus") AND ("machine learning" OR "deep learning" OR "artificial intelligence").
- Scopus: the following keywords were included in the search engine: Title-ABS-KEY ("blood tests" OR "blood exams" OR "laboratory tests"OR "laboratory exams") AND ("covid-19" OR "covid" OR "sars-cov-2" OR "coronavirus") AND ("machine learning" OR "deep learning" OR "artificial intelligence").

Documents meeting the following inclusion criteria were included: English language, available online in letters, articles or conference paper, articles presenting models based on laboratory data only or laboratory data accompanied by vital signs/symptoms/comorbidities using techniques of ML, articles published from 2020 to 2021 that presented ML models for diagnostic and prognostic purposes.

Literature Search Results

The literature search in PubMed, years 2020–2021, identified 123 articles, and in Scopus 156. Of the selected articles, 115 were reported in both databases, and 98 were excluded because they did not meet the selection criteria. Therefore 68 studies have been specified and included in this review (Fig. 5.1). Of the 68 selected

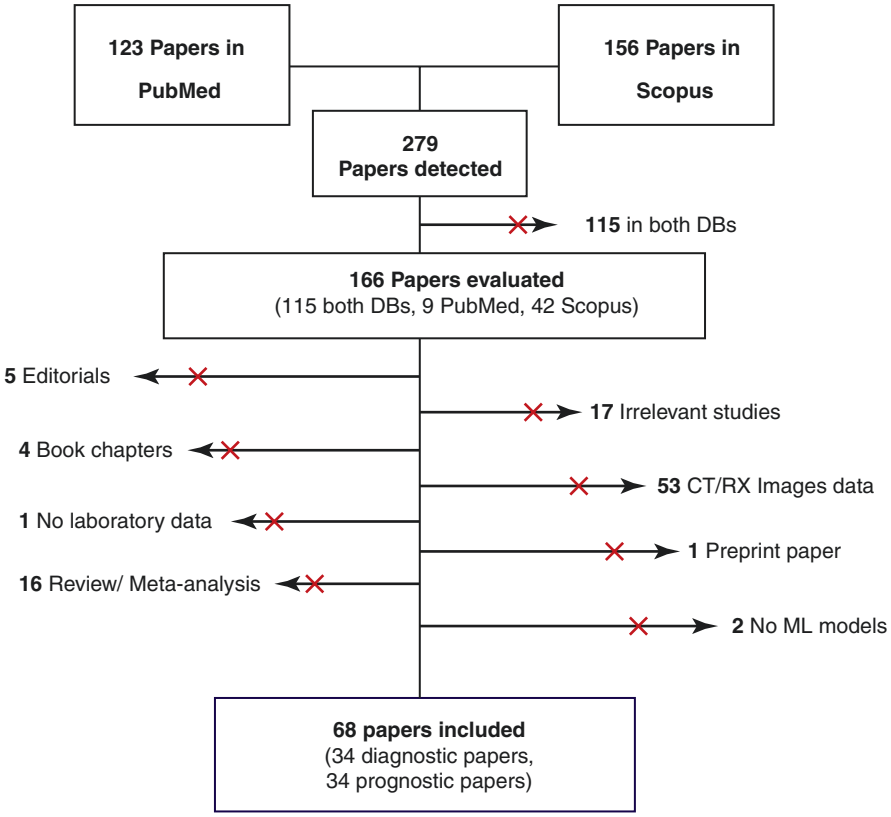


Fig. 5.1 Initial number of publications on diagnosis and prognosis for COVID-19 patients using artificial intelligence techniques identified by PubMed and Scopus (years 2020–2021), number of publications excluded because of selection criteria and final number of papers included in the review

publications, 34 describe the development of ML models for diagnostic purposes and 34 for prognostic purposes.

The complete reference list of the studies selected is shown in Appendix 1. For the sake of clarity, the reference list in Appendix 1 is structured as follows: "D" (studies pointed out as D1–D34) is for studies related to diagnostic purposes. While "P" for prognostic purposes (studies pointed out P1–P34).

The literature search revealed the publication of the first ML diagnostic study, applied to laboratory medicine for the management of COVID-19 patients, starting in June 2020 (D6), and the publication of the first prognostic study just one month later, July 2020 (P3) (Appendix 1, Table 5.1).

Of the 68 papers selected for this review, 39 are published in biomedical journals, 23 in IT-specific journals (18 out of 23 are also reported in PubMed), while seven are IT conference papers (reported only in Scopus) (Table 5.1).

In most studies (63/68), the Materials and Methods section lacks a complete description of the analytical method and/or instrumentation with which the laboratory analyses considered for developing ML models were performed (Table 5.1).

The instruments used from which it would be possible to trace the analytical method are reported only in five studies, whereas the unit of measurement is reported in 34 publications. Among the medical journals, 12 do not report any information related to the laboratory test used (Table 5.1).

The characteristics reported in the papers evaluated (dataset, the purpose of the study, features selected, ML models) for diagnostic and prognostic studies, respectively, are summarized in Figs. 5.2 and 5.3.

Diagnostic Studies

Description of the Population Publications for diagnostic purposes are based on cohorts of subjects characterized by highly variable numerosity (ranging from 106 to 115,394). Study D2, for example, includes 171 subjects, of whom 24.6% are COVID-19 patients, but the cohort is well characterized. In contrast, the paper D10, published in an IT-specific journal, has a very large cohort (115,394 subjects), of which only 0.3% are COVID-19 positive. The prevalence of COVID-19 positive subjects in the investigated population is highly variable (0.3–66.0%, in D10 and D29, respectively). This feature is probably the most critical aspect of comparing performances across the various studies. All diagnostic studies were developed on data collected during the first phase of the pandemic, i.e., within the first 7 months of the year 2020. Of these, four were set in the USA (D1, D3, D5, D9), 5 in Italy on different cohorts (D2, D4, D13–14, D28), 2 in the UK (D10, D18), 1 in Egypt (D29), 1 in Iran (D26), 4 in other European countries [Spain (D16), Turkey (D29), Slovenia (D17) and Austria (D12)] and 17 papers (D6–D8, D11, D15, D19, D21–D25, D27, D30–D34) referred to the same Brazilian cohort and dataset [37] (Fig. 5.2).

Table 5.1 List of publications selected and discussed in this review, from which bibliographic characteristics are given

ID paper	First Author	Journal	Month/year of publication	PubMed	Scopus	Study (D/P)	Availability of information on analytical methods (analyzer, reagents, and units of measurement)
D1	Joshi et al.	Journal of Clinical Virology	08/2020	X	X	D	No information
D2	Formica et al.	Clinical Medicine Journal	07/2020	X	-	D	Units of measurement only
D3	Yang et al.	Clinical Chemistry	11/2020	X	X	D	Analyzer only
D4	Cabitza et al.	Clinical Chemistry and Laboratory Medicine	10/2020	X	X	D	Analyzer/reagents/unit of measurement
D5	Plante et al.	Journal of medical Internet research	12/2020	X	X	D	No information
D6	Avila et al.	PeerJ	06/2020	X	X	D	No information
D7	Banerjee et al.	International Immunopharmacology	09/2020	X	X	D	No information
D8	AlJame et al.	Informatics in medicine unlocked	10/2020	X	X	D	No inf ormation
D9	Bayat et al.	Clinical Infectious Diseases	08/2020	X	-	D	Units of measurement only
D10	Soltan et al.	The Lancet Digital Health	02/2021	X	X	D	Units of measurement only
D11	Freitas Barbosa et al.	Research on Biomedical Engineering	01/2021	-	X	D	No information
D12	Tschoellitsch et al.	Laboratory medicine	12/2020	X	-	D	Units of measurement only
D13	Carobene et al.	Biochimica Clinica	09//2021	-	-X	D	Analyzer and units of measurement
D14	Brinati et al.	Journal of Medical Systems	8/2020	X	X	D	No information
D15	Alves et al.	Computers in Biology and Medicine	5/2021	X	X	D	Units of measurement and reference values only
D16	Planchuelo-Gómez et al.	Frontiers in Neurology	12/2020	X	X	D	Units of measurement and reference values only
D17	Kukar et al.	Scientific Reports	12/2021	X	X	D	No information
D18	Baktash et al.	QJM	11/2021	X	-	D	Units of measurement only
D19	Babaei Rikan et al.	Biomed Signal Process Control	11/2021	X	-	D	No information
D20	Çubukçu et al.	Am J Clin Pathol	11/2021	X	-	D	Analyzer/reagents/unit of measurement
D21	Cobre et al.	Computers in Biology and Medicine	7/2021	X	X	D	No information
D22	Darapaneni et al.	2020 IEEE 15th International Conference on Industrial and Information Systems, ICIIS 2020 - Proceedings	11/2020	-	X	D	No information
D23	Almansoor et al.	2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy, ICDABI	10/2020	-	X	D	No information
D24	Doewes et al.	World Journal of Engineering	5/2021	-	X	D	No informati on
D25	AlJame et al.	Scientific Reports	12/2021	X	X	D	No information
D26	Marateb et al.	Frontiers in Medicine	11/2021	X	X	D	Units of measurement only
D27	Wu et al.	Computer Methods and Programs in Biomedicine	11/2021	X	X	D	No information
D28	Cabitza et al.	Computer Methods and Programs in Biomedicine	9/2021	X	X	D	Analyzer/reagents/unit of measurement
D29	Hany et al.	2021 International Mobile, Intelligent, and Ubiquitous Computing Conference, MIUCC 2021	5/2021	-	X	D	Units of measurement and reference values only

(continued)

Table 5.1 (continued)

D30	Dairi et al.	IEEE Transactions on Instrumentation and Measurement	12/2021	-	X	D	No information
D31	Barbosa et al.	Journal of Biomolecular Structure and Dynamics	8/2021	X	X	D	Units of measurement and reference values only
D32	Ahmad et al.	Complexity	5/2021	-	X	D	No information
D33	Souza et al.	Soft Computing	5/2021	X	X	D	No information
D34	Czako et al.	Proceedings- 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing, ICCP 2020	9/2020	-	X	D	No information
P1	Assaf et al.	Internal and emergency medicine	11/2020	X	X	P	Units of measurement only
P2	Booth et al.	Modern Pathology	10/2020	X	X	P	No information
P3	Zhu J. S. et al.	Journal of the American College of Emergency Physicians	07/ 2020	X	-	P	No information
P4	Chow et al.	PLoS One	12/2020	X	X	P	Units of measurement only
P5	Jimenez-Solem et al.	Scientific Reports	02/2021	X	X	P	Units of measurement only
P6	Razavian et al.	Digital medicine	10/2020	X	X	P	Units of measurement only
P7	KO at al.	Journal of medical Internet research	12/2020	X	X	P	No information
P8	Li et al.	PeerJ	11/2020	-	X	P	Units of measurement only
P9	Hou et al.	International Journal of Medical Sciences	8/2021	X	X	P	Units of measurement only
P10	Schöning et al.	Journal of Translational Medicine	1/2021	X	X	P	Units of measurement only
P11	Aktar et al.	JMIR Med ical Informatics	4/2021	X	X	P	No information
P12	Karthikeyan et al.	Frontiers in Public Health	5/2021	X	X	P	Units of measurement only
P13	Castro et al.	Journal of the Academy of Consultation-Liaison Psychiatry	5/2021	X	X	P	Units of measurement only
P14	Chen et al.	Journal of Medical Virology	4/2021	X	X	P	Units of measurement and reference values only
P15	Marcos et al.	PLoS ONE	4/2021	X	X	P	Units of measurement only
P16	Statsenko et al.	BMJ Open	2/2021	X	X	P	Units of measurement only
P17	Woo et al.	American Journal of the Medical Sciences	10/2021	X	X	P	Units of measurement only
P18	Luo et al.	PLoS ONE	6/2021	X	X	P	Units of measurement only
P19	Yu et al.	PLoS ONE	4/2021	X	X	P	No information
P20	Famiglini et al.	Proceedings- IEEE Symposium on Computer-Based Medical Systems	6/2021	-	X	P	Analyzer and units of measurement
P21	Pulgar-Sánchez et al.	Computers in Biology and Medicine	9/2021	X	X	P	Units of measurement only
P22	Chen et al.	Journal of Medical Internet Research	4/2021	X	X	P	Units of measurement only
P23	Nguyen et al.	Scientific Reports	12/2021	X	X	P	Units of measurement only
P24	Lee et al.	Journal of the American Geriatrics Society	12/2021	X	X	P	Units of measurement for just a subgroup of Lab paramenters
P25	Chung et al.	Frontiers in Physiology	11/2021	X	X	P	No information
P26	Gabbay et al.	Applied Sciences (Switzerland)	11/2021	-	X	P	No information
P27	Qomariyah etr al.	8th International Conference on ICT for Smart Society: Digital Twin for Smart Society, ICISS 2021 - Proceeding	8/2021	-	X	P	Units of measurement and reference values only
P28	Mahboub et al.	Frontiers in Medicine	5/2021	X	X	P	No information
P29	Darapaneni et al	2021 IEEE International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021 - Proceedings	4/2021	-	X	P	No information
P30	Tang et al.	Frontiers in Cellular and Infection Microbiology	3/2021	X	X	P	Units of measurement only
P31	Zhu et al.	J Am Coll Emerg Physicians Open	7/2020	X	-	P	Units of measurement only
P32	Lu et al.	PeerJ	4/2021	X	X	D	Units of measurement only
P33	Chung et al.	Journal of Medical Internet Research	4/2021	X	X	D	Units of measurement only
P34	Ahmed et al.	Applied Sciences (Switzerland)	7/2021	-	X	D	No information

Articles published in IT-specific and conference papers, non-biomedical journals, are shaded in gray and light green respectively

D/P: *D* diagnostic study, *P* prognostic study

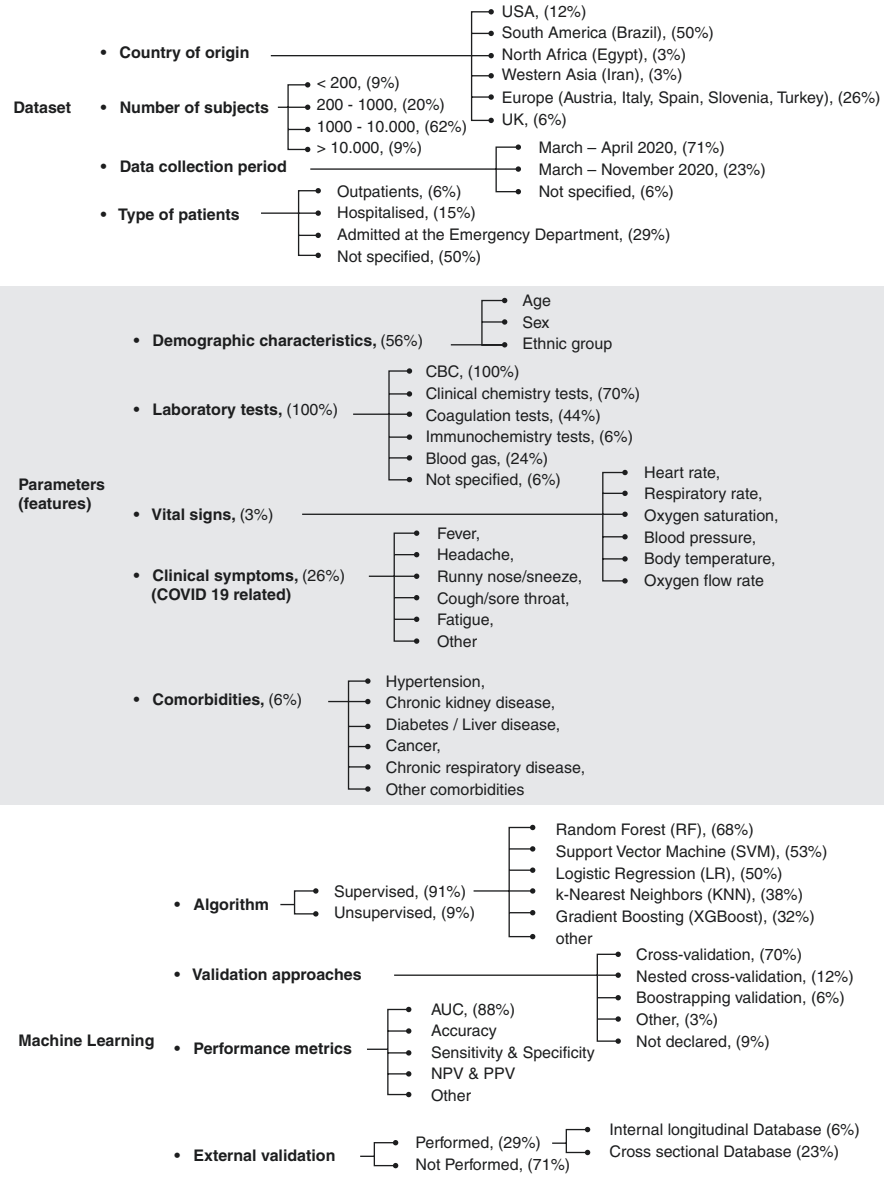


Fig. 5.2 Overview of the characteristics reported in the diagnostic studies evaluated. The variables included in the dataset (number and type of subjects and their origin, data collection period), in the features selected (demographic characteristics, laboratory tests, vital signs, clinical symptoms, comorbidities), and in the machine learning application (type of algorithm, mode of data validation, performance metric) are considered

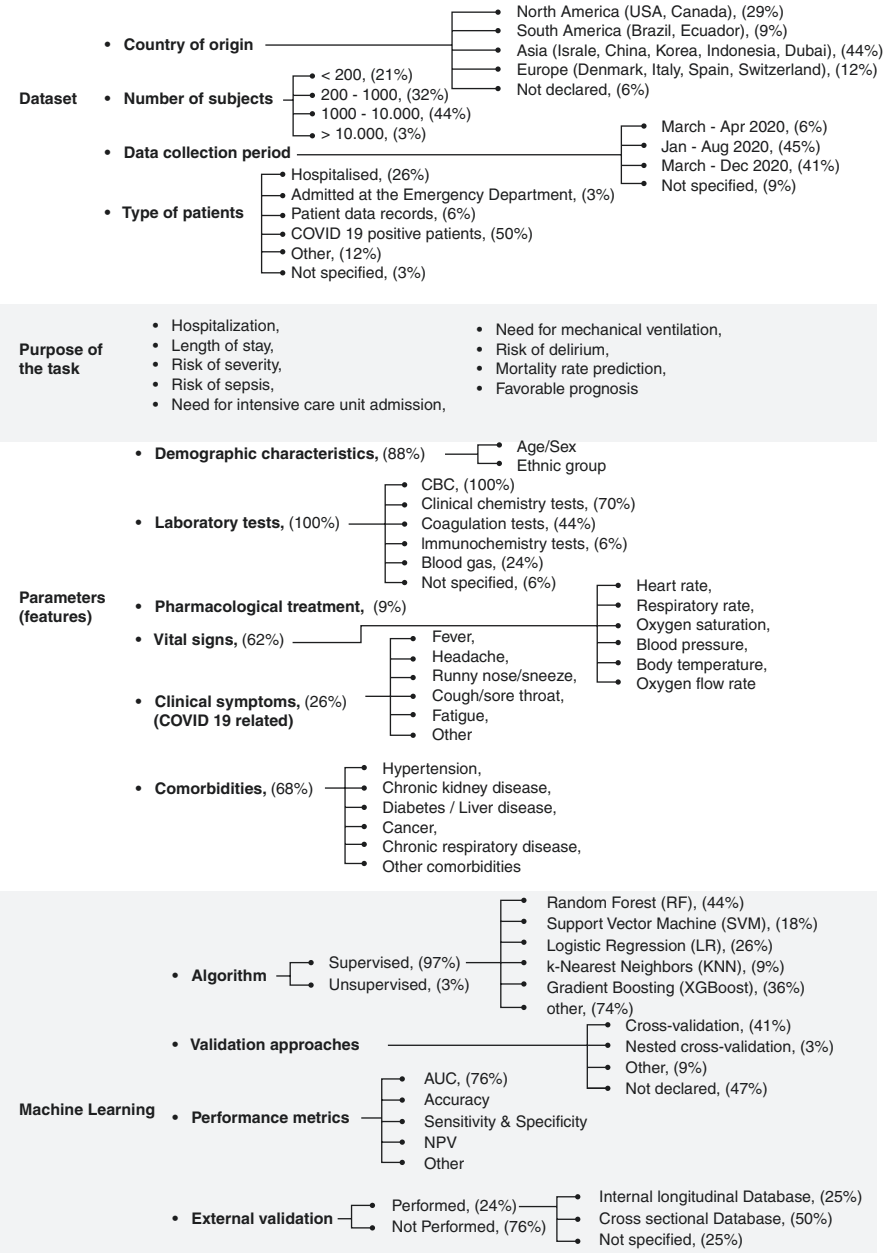


Fig. 5.3 Overview of the characteristics reported in the prognostic studies evaluated. The variables included in the dataset (number and type of subjects and their origin, data collection period), in the purpose of the study, in the features selected (demographic characteristics, laboratory tests, vital signs, clinical symptoms, comorbidities, pharmacological treatment), in the machine learning application (type of algorithm, mode of data validation, performance metric) are considered

Description of Parameters An essential element of variability concerns the number and type of parameters (features) considered to develop the different ML models for diagnostic purposes. Regarding biochemical values, all studies include CBC parameters, 24 also consider clinical chemistry values, 15 consider coagulation tests, and eight studies also include blood gas analysis parameters. Demographic data, such as ethnicity, sex and age, are considered only half of the diagnostic models. At the same time, very few studies include clinical parameters, comorbidities and COVID-19-specific symptomatology in the ML model (Fig. 5.2). Few studies report only the features (laboratory parameters) had in the model, without adding any related information.

Description of Models Twenty-six different supervised ML models are evaluated in the papers selected in this review. In most of these studies, 27 out of 34, the same dataset is analyzed with more than one model. The ML models primarily considered in the different studies are Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), k-Nearest Neighbors (KNN), and Gradient Boosting (XGBoost) (Fig. 5.2). Three studies (D16, D30, D33) have been performed using the following unsupervised ML models: Principal Component Analysis (PCA), Variational autoencoder (VAE), Generative adversarial networks (GAN), restricted Boltzmann machine (RBM), self-organizing maps (SOM).

Description of Results The results of the ML models are reported for most studies in terms of AUC (Fig. 5.2). However, three studies expressed the performances only in sensitivity and specificity [D1]. The performance of the studies described in terms of AUC covers a wide range: from a minimum of 74% [D12] to a maximum value of 99% [D8, D27, D30]. Only 10 out of 34 diagnostic studies reported external validations, two of them using internal longitudinal databases and eight cross-sectional databases. Of the reported data, the best accuracy value on an external validation was obtained using the support vector machine (SVM) model, which returned an AUC of 98% [D13].

Prognostic Studies

Description of the Population Generally, studies for prognostic purposes are based on cohorts of subjects of smaller numbers than those considered in the diagnostic studies. Like for diagnostics studies, the cohorts are highly variable (87 [P4] to 64,733 [P24]). As for the diagnostic studies, data for these studies were collected in the first months of the pandemic in different countries with different pandemic curve trends, even if three studies did not declare the period of the data collection (Fig. 5.3). Patients included have established positivity to COVID-19 by rRT-PCR; in one study [P6], salivary or molecular tests performed on pharyngeal swabs were also adopted.

Description of Parameters CBC and clinical chemistry parameters are primarily included among the features also in the prognostic models, but, unlike the diagnostic models, other variables are considered in most studies. Mainly, in 30/34 studies, demographic variables are included, while comorbidities and vital signs are considered in 23/34 and 21/34 studies, respectively (Fig. 5.3). Symptoms are included in 9 studies and pharmacological treatments in only three studies [P15, P19 and P28]. Four studies used only laboratory tests data to develop the prognostic models [P2, P18, P27 and P29]. Only in a few studies different sampling has been reported. P5 study, for example, reported four different time points for sampling: the time of diagnosis of positivity, the first 12 h after hospital admission, 12 h before ICU admission, and 12 h after ICU admission. Coagulation tests have been included in 62% of the study, while blood gas test data in 24% of the studies. Few studies included cardiac biomarkers, like troponins (TnI/TnT) and brain natriuretic peptide (BNP or proBNP), and inflammatory markers, like interleukins (ILs), procalcitonin (PCT) and tumour necrosis factor (TNF) (Fig. 5.3).

Description of Prognostic Purposes As shown in Fig. 5.3, prognostic models have been developed for various purposes, including the ability to predict: hospitalization, length of hospital stay, the risk of severe, the risk of sepsis, the need for intensive care unit admission, the need for mechanical ventilation, the prediction of delirium, mortality rate prediction, and favorable prognosis. Note that the definition of "risk of severity" is not unique in all studies: some studies have been defined, including intensive care unit admission, mechanical ventilation, and/or death, in others as ventilation or death or also as intensive care unit admission or death. P8 and P9 studies developed ML models to predict ICU admission and mortality for different purposes, P19 and P23 to predict the need for mechanical ventilation and separately the mortality, and P28 the length of stay and the mortality. P5 developed models for the prediction of the hospitalization, admission to intensive care unit, need for mechanical ventilation and mortality, as four different purposes.

Description of Patterns In 18 out of 34 studies, the purpose of the prognosis has been evaluated using one ML model, while in the others, the dataset has been analyzed by more than one model. Globally, 22 supervised ML models are examined to develop a prognostic tool; RF is the most used algorithm, but high accuracy values were obtained with other ML algorithms (Fig. 5.3). Models involving the combined use of multiple algorithms are used in three studies: P6, P7 and P20.

Description of Results Most of the models developed for prognostic purposes have reported the performance metrics in terms of AUC; however, 10/34 reported metrics expressed only as accuracy, sensitivity and specificity. In the three prognostic purposes listed in the previous paragraph (hospitalization, ICU admission, need for mechanical ventilation), accuracy values of around 80% were reported. In contrast, AUC values higher than 90% were reached in the case of models able to predict prognosis, mortality or increased risk of developing a severe form of COVID-19.

Considerations on the Literature Reviewed

In this review, among the 166 papers identified by search strings in Scopus and PubMed, 68 papers that used AI and ML techniques in the COVID-19 setting using only laboratory data have been selected. Most of the COVID-19 diagnostic studies developed ML models using imaging. These studies generally yield good results but, compared with models developed with only laboratory data, are associated with higher costs, longer timeframes, and more difficult patient management with increased risk of infection within the hospital and radiology department [38]. To obtain a helpful ML model in patients with COVID-19, several characteristics should be considered, including the purpose of the model, the patients to whom the model should be applied, the parameters, or features, selected, the ML algorithm, the performance metrics and the validation of the model. Therefore, it should be emphasized that the evaluation includes both clinical and laboratory aspects and mathematical–statistical elements. Considering only, or mainly, computational elements can lead to models that are not very useful and/or not very applicable. To this end, different professionals must cooperate to create, validate, and apply the ML model. Therefore, the further studies considered in this review will be commented on, given the elements mentioned above.

Heterogeneity in Patient Selection

It is important to emphasize that the purpose of an ML model for diagnostic purposes is not to replace the diagnosis by molecular swab, which remains the gold standard, but to assist the clinician in all those situations where the molecular swab has limitations (e.g. false negative, or in case of long waiting times for the result of the swab). The patient, identified as likely positive by the ML model, could be promptly and cautiously separated from the others pending the outcome of the molecular swab or its repetition [12].

The studies here evaluated present extreme heterogeneity in the patient selection. Most of the diagnostic models were obtained on emergency department (ED) patient populations, while in the remaining studies, inpatients or outpatients were considered. In some studies, the origin of the patients was not specified [D8, D11–12, D15, D20–25, D27, D30–D34]. In several the papers considered, the clinical description of the patients is not reported: in fact, in these papers ML model has been evaluated in cohorts of unselected patients, using as the only inclusion criteria the availability of the molecular swab result, other biochemical tests, and, in some cases, age sex or belonging to a specific ethnicity [D1, D3, D7–10]. This lack of accuracy in the patient's characterization is typical for the studies that used the same Brazilian dataset available online [37].

Sometimes, the authors state that they have included some clinical parameters in the analysis but do not report any [D5, 12]. In other cases, a partial description is

present (presence of symptoms suggestive of COVID-19) [D3, D8, D14–15, D17, D21–22, D25–26] and/or some exclusion criteria such as the presence of specific comorbidities are in addition reported [D2]. Finally, in some studies, the clinical presentation of subjects was considered and accurately described [D4] and/or the pattern was evaluated in different subgroups (in the whole sample or in the subset of asymptomatic versus symptomatic subjects) [D4].

It should be considered that even the characteristics of a patient with COVID-19, including clinical presentation, prevalence by age (which has changed over time), the introduction of new therapies that have significantly ensured improved prognosis (e.g. hyperimmune plasma), biochemical and instrumental tests, can vary very abruptly and present an extreme heterogeneity. From this point of view, a model developed using a dataset from a certain subgroup reached a high diagnostic accuracy, may perform much less in another clinical setting.

In a recent paper provocatively titled "The importance of being external. Methodological insights for the external validation of machine learning models in medicine", authors state that there is still a gap in the literature on how to interpret external validation results and hence assess the robustness of ML models [25]. Cabitza et al., by using eight different external validation CBC datasets collected across three different continents during different pandemic waves, report how the correlation between accuracy and similarity should serve as a warning sign that reproducing good performances across very heterogeneous settings can be overambitious and unrealistic [25].

It should be noted that focusing on accuracy alone is not always helpful. Its sensitivity to distortion in case of unbalanced datasets, which is datasets where one condition (e.g., death or ICU admission) is infrequent. Likewise, AUC and F1 could be misleading, as they do not distinguish between types of errors, e.g., false-negative errors in detecting potentially severe health problems (aggressive tumors, fast infections). Balanced accuracy and utility metrics could be more useful to assess the clinical utility of a classification model.

These considerations can also be made, and above all, in the case of predictive models: for these models, the patient's clinical course, the pharmacological therapy, the use of oxygen devices, the age, and the presence of possible comorbidities are fundamental elements. The predictive models considered in this review also present a wide variability both for the inclusion criteria of the selected population and for the prognostic endpoint of the study. In most cases, the model was calculated on inpatients or outpatients, and, in at least half of the studies, there is a detailed description of the clinical and anamnestic characteristics of the selected patients. Endpoints considered included mortality (at different time points) [P2–3, P5, P7], a severe form of COVID-19 characterized by respiratory failure with need for mechanical ventilation and/or admission to the intensive care unit [P1, 4–5], risk of sepsis [P30] or delirium [P13], the probability to be hospitalized and length of stay [P28], or a favorable prognosis [P6] (Fig. 5.3). The heterogeneity of the choice of endpoints makes it difficult to compare models.

Laboratory Parameters Used by Machine Learning Models

A second critical aspect related to applying the ML model to the patient with suspected (diagnostic models) or specific (predictive models) diagnosis of COVID-19 is the choice of features or laboratory tests. It should be kept in mind that rRT-PCR, although considered the best diagnostic system (gold standard), has limitations in sensitivity and specificity, and often for the correct classification of patients, some additional criteria are used (e.g., CT and/or clinical presentation) [27]. Therefore, the analytical method for diagnosing COVID-19 used in the different studies, with which the ML model is compared, indeed represents an additional factor of heterogeneity.

In selecting laboratory tests to be used in ML models, a key role should be played by the laboratory professional. There are, in fact, some critical elements that should be considered: the relation between variables and COVID-19 pathology; the availability of the variables (features) or diagnostic test in the context in which the ML model is to be applied; the characteristics of the diagnostic test.

Although from a theoretical point of view, the ML model can highlight a link, previously unknown, between a marker and the pathology, the features selection should interest, or at least include those variables that the scientific literature has already reported as associated or altered in the various stages of the COVID-19 disease.

A vast number of laboratory tests have been reported as altered in recent meta-analyses, with a possible role for monitoring, stratification, and prognosis: analytical, hematologic (leukocytosis, lymphopenia, neutrophilia, anemia, thrombocytopenia), biochemical (hypoalbuminemia, increased lactate dehydrogenase, aspartate aminotransferase, alanine aminotransferase total bilirubin, creatinine, troponin, C-reactive protein, infection markers (interleukins and procalcitonin) and coagulative (increased D-Dimer and prothrombin time) alterations [15–18, 39, 40]. Rapid changes in these parameters may occur during COVID-19, which may be particularly critical for the ML model and its application. For example, it is plausible to hypothesize that the clinical and laboratory picture of a patient who arrives late in an emergency department is very different from that observed in subsequent months. Therefore, the weight of the various laboratory tests could be different not only in the various phases of the disease but also in the different periods during the pandemic [12, 25].

As already observed for the selection of patients, the choice of parameters also presents a high heterogeneity. There are ML models developed using only CBC parameters [D1–2, 6–7, D13–15, D18, D20, D28, P18, P20, P25, P33], or CBC and clinical chemistry, or CBC, clinical chemistry, and coagulation, or also including blood gas tests (Figs. 5.2 and 5.3).

Moreover, other variables, like the clinical history, comorbidities, and pharmacology therapies, in addition to laboratory tests, were considered by predictive models. Relative to the number of predictors selected, there are parsimonious models, characterized by a small number of variables (less than 10 in D1, D29, P2 and P18),

and more frequently models with a vast number of predictors (more than 50 in D4, D10, D21–25 D27 and P3, P12, P14–15, P30–31).

It is then undoubtedly necessary to consider whether the selected diagnostic tests are available in the clinical setting where the ML model will be used. In this regard, to consider the possibility that the laboratory test may be requested, the timing and availability in the territory are necessary. The choice of parameters is, in fact, often debatable and seems not to be performed with an expert in laboratory medicine. For example, a diagnostic model includes a marker that cannot be requested urgently and of little or no diagnostic significance (urea or cholinesterase included in D8 and D12), with particularly long analytical times or excessively expensive, is not very applicable. Similarly, a model, diagnostic or predictive, that included a marker that is scarcely present in most hospital laboratories would be of little use.

In contrast, an ML model using only CBC data could be used anywhere, especially in developing countries that suffer from shortages of reagents and specialized laboratories [38]. The low variability of results in CBC parameters, both for the biological component [31–34] and the analytical component [35], could also make the use of a model based solely on these parameters more effective. Another essential aspect of being considered is, in fact, the analytical one (unit of measurement, reference interval, analytical method, instrumentation, traceability). Assuming, for example, the non-specificity of the measurement of serum creatinine, a parameter used by most of the studies considered, of Jaffe methods compared to enzymatic methods and the impact that the standardization process has had on the quality of its measurement [41, 42], it is evident that at least for some laboratory parameters, information related to analytical methods is fundamental. In this regard, it is certainly noteworthy a mini review dedicated to the results of D-Dimer in COVID 19 patients, in which a great deal of confusion is reported in the way data are registered, and thus, the considerable misinformation that results [43].

Except for a few [D4, D13, D20, D28, P20], most diagnostic and predictive models do not report analytic methods and/or instrumentation (Table 5.1). Considering what has been described, reviewing the cited papers, one gets the impression that, for most of them, the authors did not make a reasoned selection of laboratory parameters but instead have developed the various models simply by using all the data available in the management systems of individual health care facilities.

The evidence of a non-uniformity in the use of clinical data, due to the significant increase in recent years of studies with AI techniques, including ML, and the consequent poor reproducibility of the models developed [44], has made it necessary to draft guidelines that could improve the quality of AI studies in the medical field. The recently published MINIMAR (MINimum Information for Medical AI Reporting) guidelines were created with this intent [45]. MINIMAR considers a minimum list of information that should be reported in publications but does not indicate any information on how to report laboratory test results necessary information to allow reproducibility of the study, which paradoxically is precisely the purpose of the guidelines [45]. It is not a coincidence that guidelines dealing with studies using laboratory medicine results [46–49] always include a request for an adequate description and characterization of the procedures used to obtain them

(instrument used, analytical principle, unit of measurement, method optimization, generation, specificity). The analytical procedures must measure the same quantity to get comparable and reliable results. The description of the analytical methodology used to obtain the laboratory data is necessary to precisely identify the measurand [50].

This chapter is not certain to address the complex and multifaceted issues related to standardization and harmonization, and thus reliability and reproducibility of laboratory results. However, as laboratory professionals, we are equally aware that the risk of misuse of laboratory data is genuinely high without knowledge of these aspects.

Types of Models and Their Validation

As it is logical to expect, most of the algorithms used in the analyzed studies belong to supervised algorithms. Only three diagnostic studies and one prognostic [D16, D30, D33, P21] developed ML models based on unsupervised algorithms. In short, and simplifying a lot, with supervised algorithms, one tries to build a model from "labelled" training data to make predictions about future data. By "labelled", we mean that the dataset already contains the output or response information sought: the patient does or does not have SARS-CoV-2 infection (diagnostic models), the patient is alive or dead at 24–48 h, the patient has or has not been hospitalized, and the patient has or has not required mechanical ventilation (predictive models). By contrast, in the unsupervised case, only input data are given. The goal is to try to model the underlying structure or distribution of the data and find unknown patterns [51]. However, it is interesting to note the number of different algorithms and varying complexity (Figs. 5.2 and 5.3). In particular, the most frequently used supervised algorithms include RF (23 out of 34 diagnostic models, 15 of 34 predictive models) and LR (17/34 and 12/34, respectively) (Figs. 5.2 and 5.3).

A second aspect concerns the validation procedure used in the various studies.

The techniques most frequently used are split-sample validation: hold-out (by splitting the data into training and test), K-fold cross-validation (it's more robust than the former), K-fold nested/repeated CV, and bootstrapping techniques for estimating confidence intervals.

These methods present, obviously, different advantages and problems [52, 53] and constitute examples of internal validation, whose application is handy to verify the stability of the model. On the contrary, the external validation carried out by using an independent dataset, different from that used to train the model, allows estimating the reproducibility of a model and its generalizability to the timing of the data collection and the space (other hospital or geographic zone). Therefore, external validation is fundamental to evaluating the quality of the proposed model. A combination of the two procedures, the so-called internal–external validation, is also possible, particularly useful in those contexts where the amount of data

available is limited. For one-third of the diagnostic models considered, the authors reported external validation with one or more validation datasets (Fig. 5.2).

A further noteworthy element concerns the numerosity of the dataset used, which conditions the choice of the validation algorithm and procedure and, indeed, the model's accuracy. The studies considered present very different sample sizes, ranging from 170 [D2] to 115,394 [D10] for the diagnostic models and from 87 [P4] to 64,733 [P24] for the predictive models.

Model Implementation

Two ulterior aspects are fundamental for applying the model: the representativeness and the availability of the dataset and the explainability of the algorithm. It is also essential to make available the models by creating online tools to give the possibility to test the algorithms developed.

The availability of all the information used to generate the model allows us to evaluate the validity of the presented model definitively and verify its possible use. This aspect is particularly critical in the context of Big Data, ML algorithms and all those situations characterized by high computational complexity: it is the fundamental element for a true reproducible science [54]. To this end, it is necessary that the original data, the source code (for data cleaning, analysis, and presentation), is freely accessible. All analytical methods used are described in detail. Many authors do not agree to share their data for different reasons; however, such sharing, and data transparency in general, is the fundamental prerequisite for a free science to advance on a robust foundation. Such transparency would truly allow clinicians, in collaboration with their laboratory colleagues, to understand whether the proposed model can be used in their setting, through local validation of their data, before actual implementation. Although for only one study (D10) the source code is available (available at https://github.com/andrewsoltan/CURIAL-manuscript), for some of them the dataset is freely accessible [D4, 8, 14–15, 21, 23–24, 27, 30, 33–34; P10–12, 15, 20, 26, 29, 32].

A second important aspect is an availability to the end-user (the clinician) of IT tools that allow implementation of the model. These tools can be realized in the form of applications directly implemented in the management system of the individual operating units or the middleware of the laboratory; this would have the advantage of an automatic compilation of input fields, once the laboratory results are available, reducing both the response time and the rate of input errors. In this regard, it is interesting to note that some of the diagnostic and predictive models considered have been implemented in real applications available online [D 4, D14, D17, P10, P26].

For a detailed description of one of these applications, see Appendix 2.

In the light of what has been discussed above, some relevant conclusions can be drawn:

- The studies considered are characterized by extreme heterogeneity in terms of the type of patients, their numerosity, the laboratory parameters selected, the algorithms used, and their validation.
- The use of diagnostic and prognostic models requires an accurate description of the patients, of the statistical and analytical methods used (and if possible, the publication of the dataset and the source code); moreover, the availability of a web application, or possibly of a computer tool inserted in the hospital management systems, allows the implementation and the usability of the model.
- To obtain clinically usable models, effective collaboration between data scientists and clinicians is fundamental; however, laboratory medicine professionals must deepen concepts related to the world of Big Data and ML algorithms to understand their potentialities and correctly interpret their results. In "A Brief Guide to Medical Professionals in the Age of Artificial Intelligence" [55], the authors hope that clinicians will apply themselves to an understanding of the basics of these new technologies so that they will be able to evaluate opportunities for clinical studies. We can only share this wish. In this sense, the Italian Society of Clinical Chemistry and Clinical Molecular Biology (SIBioC) is already working to promote an interdisciplinary operative presence [6, 20].

The Role of Artificial Intelligence in the Vaccination Strategy Against SARS-COV-2 Through Laboratory Tests

Twelve months after the early COVID-19 cases were reported in Wuhan, China [56], the first COVID-19 vaccine was authorized by the US FDA for emergency use, and many Countries started national vaccine campaigns [57]. At the end of 2020, administrations began worldwide with RNA-based vaccines using the SARS-CoV-2 Spike protein (S-protein) epitopes to induce protective immunity [58]. The BNT162b2 mRNA vaccine, one of the FDA-approved vaccines, has shown up to 95% protection against the worst form of COVID-19 disease, providing significant public health advantages [59].

However, previous studies on the *Coronaviridae* family had shown both a short-lived antibody response and the possibility of reinfection in recovered individuals [60, 61]. Consequently, it has been considered essential to assess the efficacy of vaccination against SARS-CoV-2. For this purpose, the antibody titer against the SARS-CoV-2 S-protein and its longevity in sera of vaccinated subjects was evaluated. The first data were obtained in healthcare workers cohorts, as a prioritized category administered with the BNT162b2 mRNA vaccine; it was then observed that the S-protein antibody response was persistent at 6 months, even with a certain decline [62, 63]. This decrease was partly expected as all vaccine-induced short-lived plasmablasts do not necessarily differentiate into long-lived plasma cells [64].

The antibody titer analysis showed that, also for SARS-CoV-2, gender-based variables seem to play a significant role in determining the immune response [63].

From the literature, it is known that males and females show variability in innate and adaptive responses throughout life [65] as males are more susceptible to viral infections due to the inhibitory role of androgens in the production of IFN-γ. On the other hand, women show a more robust humoral response after vaccine administration [63]. In addition to these intra- and inter-individual variability factors limiting the routine use of serological tests as an indicator of the immune coverage, we must consider the difficulty in harmonizing the different commercial serological tests [66].

Despite these limitations, the potential of serological tests, in directing doses to the neediest subjects (for example, seronegative individuals) should not be underestimated. Such strategies could redefine inoculation priorities as a function of antibodies titer and consider targeted vaccination protocols, particularly for older people, whose humoral response to infections could be less efficient due to immunosenescence [65].

A valid vaccination strategy aiming to reduce severe morbidity, mortality, and harmful societal impact due to the transmission of SARS-CoV-2, allows optimization of public health interventions in facing the pandemic. The allocation criteria are focused on different risks: (1) acquiring infection, (2) severe morbidity and mortality, (3) negative societal impact, (4) transmission of the infection [67]. Applying allocation criteria to specific population groups could establish whether a category should be prioritized in receiving the immunization. In vaccine scarcity scenarios or contexts of organizational complexity, serological testing could be a valuable index for allocating doses to individuals based on their serostatus.

Indeed, low-income countries, where the probability of exposure and infection is higher due to extreme environmental conditions (e.g., the lack of personal protective equipment or inability to work from home [67]), the power of a serological triage strategy could be essential in controlling the pandemic. Moreover, less than 5% of the African population is fully vaccinated, causing a high risk of further waves, even if rich countries are engaging in the COVAX project, a global collaborative co-financing vaccine procurement strategy [68, 69].

Real-World Vaccination Strategies

A proper prioritization strategy should consider numerous variables such as age, gender, clinical history, and serostatus. In different viral infections, serological evaluation plays a crucial role in identifying unprotected individuals, distinguishing primary infection from reactivation, and testing the immune response elicited by vaccination. The antibody titer dosage could be an index of the immune response, informing about immunization coverage and durability against a specific pathogen (e.g., HBV and VZV).

To date, for the Hepatitis B virus, assessment of seropositivity against the S antigen is decisive in evaluating eventual revaccination. An anti-HbSAg concentration >10 mIU/mL measured 1–3 months after administering the last dose of the primary vaccination series is considered a reliable indicator of immunization [70]. Also, in varicella-zoster virus (VZV), antibody titer helps identify unprotected individuals and differentiate primary infection from reactivation. The humoral response begins to be detectable by the time of rash appearance, and testing timing is essential in exclusion of infection; secondly, due to high false-negative rates for IgM and IgG by validated commercial diagnostic systems, in cases of suspicion and presence of negative serological tests, the subject should be retested for antibodies within 2–3 weeks, guiding clinical decisions [71].

Other prioritization strategies could be based on age, gender, and comorbidities (e.g., HPV, Influenza Virus). For Human Papillomavirus vaccination when the first vaccination campaign started in Europe and Italy, the primary targets were preadolescent girls to obtain immunization before initiation of sexual activity and subsequent exposure to HPV [72].

Clinical history and comorbidities could guide the prioritization scheme for fragile subjects, as happens for the Influenza Virus. In many countries, vaccination to prevent influenza and its complications is recommended for "high risk" groups, such as the elderly, patients with chronic conditions, and institutionalized populations [73].

For SARS-CoV-2, in the early phases of the pandemic, the vaccine allocation framework emphasizes the prevention of severe morbidity and mortality to preserve essential health and emergency services [67]. So, to safeguard individuals aged 80 years and older, residents in long-term care facilities (LTCF), and healthcare workers, they were designated as a primary category to receive vaccine inoculation [74]. As vaccines become more available in later phases, the focus shifts towards immunization of non-essential categories at higher risk of acquiring infection, transmitting the infection to others, and at risk of negative societal impact (e.g., caregivers, teachers). Thus, this strategy could cause unequal access to vaccination between hospital-based healthcare personnel and informal caregivers or healthcare personnel outside of hospitals because the latter are not included in the prioritized category in the first phase [63].

Test&Vaccine strategies are conceived to address vaccination towards high-risk subgroups or select *post-serological testing* categories constituting a helpful shortcut in vaccine scarcity scenarios or developing a prioritization scheme. Despite these considerations, a *real-world* Test&Vaccine strategy takes more time to be enforced due to the elevate number of variables to consider. The most significant ones are sex, gender, age, serostatus, laboratory parameters, clinical features, and clinical scores.

Artificial Intelligence Potentialities

Machine learning algorithms constitute a helpful resource in analyzing substantial amounts of data in Laboratory Medicine. As the knowledge and application in this field will proceed, they could be applied to reduce costs, support clinical decision-making, and improve outcomes [11]. In the case of the SARS-CoV-2 pandemic, the significant aims considered were the reduction of COVID-19 incidence, mortality and years of life lost (YLL), severe illness, the pursuit of equality and the equity strategy in dose allocation. Although serological evaluation could be a marker of immunization, only Bubar et al. [75] considered serological status a relevant variable for their algorithm among the test and vaccine strategies briefly commented below. A SEIR (susceptible, exposed, infectious, recovered) model was developed to assess different prioritization strategies considering vaccine efficacy and its ability to reduce mortality and block transmission correlated with age variations in susceptibility, fatality rates, and immune decline. The population enrolled was classified into different subgroups: (1) children and teenagers, (2) adults aged 20–49 years, (3) adults aged 20 years or over, (4) adults aged 60 years or over, and (5) all individuals. It has been reported that a highly effective transmission-blocking vaccine prioritized to younger subgroups of the population reduces the cumulative incidence, assuming the fact that the latter are considered higher transmission groups with a minor infection fatality rate. Assuming a return to high contact rates and pre-pandemic behavior and considering virtually all plausible vaccine characteristics, the proposed vaccination strategy prioritizes individuals over 60 years and those with comorbidities, minimizing mortality and YLL. This recommendation is robust because of the vast differences in infection fatality rate by age. When vaccine efficacy is leaky or poorly effective in adults, the categories to prioritize are younger age groups. Moreover, they assess targeted vaccination based on serological status, finding that vaccinating seronegative individuals improves efficiency in settings where seroprevalence is high [76]. Potentially, serological testing could redirect doses to seronegative individuals, improving the marginal impact of each dose while potentially reducing existing inequities in the COVID-19 pandemic managing [75].

Further Test & Vaccine strategies have been based on epidemiological characteristics and clinical features both combined with age. In particular, age is a standard used variable because of its association with susceptibility, seroprevalence, severity, and mortality [75].

Ioannou et al. [77] developed a logistic regression model (COVIDVax) that considers sex, age, race, ethnicity, body mass index, diabetes, chronic kidney disease, congestive heart failure, and two clinical scores (Care Assessment Need score and Charlson Comorbidity Index). The enrolled population was well balanced in ethnicity, including White, Black, and Hispanic individuals. The model exhibited excellent stratification of risk of SARS-CoV-2—related deaths (that would occur by the time 50% of Veterans enrollees are vaccinated) with an area under the receiver operating characteristic curve (AUROC) of 85.3% (95% CI 84.6–86.1%), superior to the

AUROC of using age alone (72.6%; 95% CI 71.6–73.6%). As a result, prioritizing population using the COVIDVax model has proved to be more efficient (63.5%) than considering age alone (45.6%) or the US Centers for Disease Control and Prevention phases of vaccine allocation (41.1%). It highlights that many deaths expected to occur during vaccine roll-out could be prevented before sufficient herd immunity is achieved.

As previously reported, African nations are still waiting for vaccine stocks attesting the total number of people who received all doses prescribed by the initial vaccination protocol at 26.8% out of the total population [69]. In Mellado et al. [68], a deep neural network model (DNN) was constructed on fourteen characteristics of enrolled subjects living in Gauteng province in South Africa: age, eleven relevant comorbidities, gender and ethnicity, information about the type of hospitalization (general ward, ICU, high-level care, and isolation), and whether the subject was discharged alive or died in care. The data were clustered into two datasets: severe illness (ICU, high-level care, and mortality) and less severe illness (subject discharged from the general ward). The DNN model indicates that by vaccinating about 20% of the adult population, the probability of severe COVID-19 illness can be reduced by over 80%. Moreover, the DNN output provides a dimensionless number between 0 (highest risk of severe condition) and 1 (lowest risk of severe condition). AI-derived recommendations can be elaborated for classifying the risk of severe illness (1–greatest, 5–lowest risk) by demographic interval and by comorbidities. It results that a subgroup to be prioritized is constituted by adults >60 years old with cardiac disease, diabetes, and hypertension who have shown to be at higher risk of severe illness, reaching the same conclusion as Bubar et al. [75].

Combining age with crescent vaccine efficacy (VE: 10–100%), Matrajt et al. [78] obtain different desirable outcomes in reducing mortality. The population enrolled was divided into five vaccination groups: young aged between 0 and 19 years, adults between 20 and 49 years old, adults between 50 and 64 years old, adults between 65 and 74 years old, and those 75 and older. The model proves that vaccine effectiveness ≥50% would be enough to mitigate the ongoing pandemic, provided that a high percentage of the population is optimally covered. Thus, for low vaccine effectiveness, independently of vaccination coverage, it is optimal to allocate vaccines to high-risk (older) age groups first. In contrast, for higher vaccine effectiveness, there is a switch in the allocation of vaccines to high-transmission (younger) age groups first to obtain high vaccination coverage.

Vaccine prioritization strategies have opened a debate in Science and Ethics: the COVID-19 pandemic has highlighted social inequalities and has led to questions about the disparity existing in the healthcare system [67]. Pressman et al. [79] have developed a COVID-19 vaccine equity index (CVEI) to reach equitable outcomes in different subgroups (non-Hispanic White, non-Hispanic Black, non-Hispanic Asian, and Hispanic). The CVEI identifies and quantifies disparities in outcomes for each racial/ethnic group considered and illustrates the potential impact of vaccine distribution on advancing equality. Applying conditional probability and statistical theory, they measure equity in administering doses for unvaccinated individuals deriving an index to spotlight these existing inequities. The index measures equity

when the same patient outcomes are achieved despite socio-demographic factors (e.g., age, gender, race/ethnicity). By contrast, equality focuses on behaviours and actions, where all socio-demographic subgroups receive the same treatment. In this context, the desired outcome was to reduce severe illness, hospitalizations, and death. Equity is achieved when the likelihood of these adverse outcomes is comparable across all groups considered.

Conclusions

Vaccination strategies could consider serological tests as a companion diagnostic to allocate vaccines in various scenarios. The shortcut provided by ML models has the tangible effect of selecting categories to be prioritized in different settings and calculating how much the vaccination coverage obtained with a conventional prioritization strategy (mainly based on age comorbidities and social impact) deviates from the laboratory parameters based strategy. So, serological tests would potentially transform vaccination in a *tailored path* fitting to each immune response.

Appendix 1

References list of the diagnostic and prognostic papers included in the study.

Diagnostic Papers (D)

1. Joshi RP, Pejaver V, Hammarlund NE, et al. A predictive tool for identification of SARS-CoV-2 PCR-negative emergency department patients using routine test results. J Clin Virol. 2020;129:104502. https://doi.org/10.1016/j.jcv.2020.104502.
2. Formica V, Minieri M, Bernardini S, et al. Complete blood count might help to identify subjects with high probability of testing positive to SARS-CoV-2. Clin Med (Lond). 2020;20(4):e114–19. https://doi.org/10.7861/clinmed.2020-0373.
3. Yang HS, Hou Y, Vasovic LV, et al. Routine laboratory blood tests predict SARS-CoV-2 infection using machine learning. Clin Chem. 2020;66(11):1396–1404. https://doi.org/10.1093/clinchem/hvaa200.
4. Cabitza F, Campagner A, Ferrari D, et al. Development, evaluation, and validation of machine learning models for COVID-19 detection based on routine blood tests. Clin Chem Lab Med. 59(2):421–31.
5. Plante TB, Blau AM, Berg AN, et al. Development and external validation of a machine learning tool to rule out COVID-19 among adults in the emergency

department using routine blood tests: a large, multicenter, real-world study. J Med Internet Res. 2020;22(12):e24048. https://doi.org/10.2196/24048.
6. Avila E, Kahmann A, Alho C, et al. Hemogram data as a tool for decision-making in COVID-19 management: applications to resource scarcity scenarios. PeerJ. 2020;8:e9482. https://doi.org/10.7717/peerj.9482.
7. Banerjee A, Ray S, Vorselaars B, et al. Use of machine learning and artificial intelligence to predict SARS-CoV-2 infection from full blood counts in a population. Int Immunopharmacol. 2020;86:106705. https://doi.org/10.1016/j.intimp.2020.106705.
8. AlJame M, Ahmad I, Imtiaz A, et al. Ensemble learning model for diagnosing COVID-19 from routine blood tests. Inform Med Unlocked. 2020;21:100449. https://doi.org/10.1016/j.imu.2020.100449.
9. Bayat V, Phelps S, Ryono R, et al. A severe acute respiratory syndrome Coronavirus 2 (SARS-CoV-2) prediction model from standard laboratory tests. Clin Infect Dis. 2021;73(9):e2901–7. https://doi.org/10.1093/cid/ciaa1175.
10. Soltan AAS, Kouchaki S, Zhu T, et al. Rapid triage for COVID-19 using routine clinical data for patients attending hospital: development and prospective validation of an artificial intelligence screening test. Lancet Digit Health. 2021;3(2):e78–87.
11. de Freitas Barbosa VA, Gomes JC, de Santana MA. Heg.IA: an intelligent system to support diagnosis of Covid-19 based on blood tests. Res Biomed Eng. 2021.
12. Tschoellitsch T, Dünser M, Böck C, et al. Machine learning prediction of SARS-CoV-2 polymerase chain reaction results with routine blood tests. Lab Med. 2021;52(2):146–9. https://doi.org/10.1093/labmed/lmaa111.
13. Carobene A, Campagner A, Sulejmani A, et al. Identification of SARS-CoV-2 positivity using machine learning methods on complete blood count data: external validation of state-of-the-art models. Biochim Clin. 2021. https://doi.org/10.19186/BC_2021.033.
14. Brinati D, Campagner A, Ferrari D, et al. Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study. J Med Syst. 2020;44(8):135. https://doi.org/10.1007/s10916-020-01597-4.
15. Alves MA, Castro GZ, Oliveira BAS, et al. Explaining machine learning based diagnosis of COVID-19 from routine blood tests with decision trees and criteria graphs. Comput Biol Med. 2021;132:104335. https://doi.org/10.1016/j.compbiomed.2021.104335.
16. Planchuelo-Gómez Á, Trigo J, de Luis-García R, et al. Deep phenotyping of headache in hospitalized COVID-19 patients via principal component analysis. Front Neurol. 2020;11:583870. https://doi.org/10.3389/fneur.2020.583870.
17. Kukar M, Gunčar G, Vovko T, et al. COVID-19 diagnosis by routine blood tests using machine learning. Sci Rep. 2021;11(1):10738. https://doi.org/10.1038/s41598-021-90265-9.
18. Baktash V, Hosack T, Rule R, et al. Development, evaluation and validation of machine learning algorithms to detect atypical and asymptomatic presentations

of Covid-19 in hospital practice. QJM. 2021;114(7):496–501. https://doi.org/10.1093/qjmed/hcab172.
19. Babaei Rikan S, Sorayaie Azar A, Ghafari A, et al. COVID-19 diagnosis from routine blood tests using artificial intelligence techniques. Biomed Signal Process Control. 2021;103263. https://doi.org/10.1016/j.bspc.2021.103263.
20. Çubukçu HC, Topcu Dİ, Bayraktar N, et al. Detection of COVID-19 by machine learning using routine laboratory tests. Am J Clin Pathol. 2021;aqab187. https://doi.org/10.1093/ajcp/aqab187.
21. Cobre AF, Stremel DP, Noleto GR, et al. Diagnosis and prediction of COVID-19 severity: can biochemical tests and machine learning be used as prognostic indicators? Comput Biol Med. 2021;134:104531. https://doi.org/10.1016/j.compbiomed.2021.104531.
22. Darapaneni N, Singh A, Paduri A, et al. A machine learning approach to predicting Covid-19 cases amongst suspected cases and their category of admission. In: 2020 IEEE 15th International conference on industrial and information systems, ICIIS 2020—Proceedings (pp. 375–80).
23. Almansoor M, Hewahi NM. Exploring the relation between blood tests and Covid-19 using machine learning. In: 2020 International conference on data analytics for business and industry: way towards a sustainable economy, ICDABI 2020.
24. Doewes RI, Nair R, Sharma T. Diagnosis of COVID-19 through blood sample using ensemble genetic algorithms and machine learning classifier. World J Eng. 2021.
25. AlJame M, Imtiaz A, Ahmad I, et al. Deep forest model for diagnosing COVID-19 from routine blood tests. Sci Rep. 2021;11:16682. https://doi.org/10.1038/s41598-021-95957-w.
26. Marateb HR, Ziaie Nezhad F, Mohebian MR, et al. Automatic classification between COVID-19 and non-COVID-19 pneumonia using symptoms, comorbidities, and laboratory findings: the Khorshid COVID Cohort Study. Front Med. 2021;8.
27. Wu J, Shen J, Xu M, et al. A novel combined dynamic ensemble selection model for imbalanced data to detect COVID-19 from complete blood count. Comput Methods Programs Biomed. 2021;211:106444. https://doi.org/10.1016/j.cmpb.2021.106444.
28. Cabitza F, Campagner A, Soares F, et al. The importance of being external. Methodological insights for the external validation of machine learning models in medicine. Comput Methods Programs Biomed. 2021;208:106288. https://doi.org/10.1016/j.cmpb.2021.106288.
29. Hany N, Atef N, Mostafa N, et al. Detection COVID-19 using machine learning from blood tests. In: 2021 International mobile, intelligent, and ubiquitous computing conference, MIUCC 2021, pp. 229–34.
30. Dairi A, Harrou F, Sun Y. Deep generative learning-based 1-SVM detectors for unsupervised COVID-19 infection detection using blood tests. IEEE Trans Instrum Meas. 2021.

31. Barbosa VADF, Gomes JC, de Santana MA, et al. Covid-19 rapid test by combining a Random Forest-based web system and blood tests. J Biomol Struct Dyn. 2021.
32. Ahmad A, Safi O, Malebary S, et al. Decision tree ensembles to predict Coronavirus disease 2019 infection: a comparative study. Complexity. 2021.
33. Souza AA, Almeida DC, Barcelos TS, et al. Simple hemogram to support the decision-making of COVID-19 diagnosis using clusters analysis with self-organizing maps neural network. Soft Comput. 2021;1–12. https://doi.org/10.1007/s00500-021-05810-5.
34. Czako Z, Sebestyen G, Hangan A. COVID-19 preliminary patient filtering based on regular blood tests using auto-adaptive artificial intelligence platform. In: Proceedings—2020 IEEE 16th international conference on intelligent computer communication and processing, ICCP 2020, pp. 109–16.

Prognostic Papers (P)

1. Assaf D, Gutman Y, Neuman Y, et al. Utilization of machine-learning models to accurately predict the risk for critical COVID-19. Intern Emerg Med. 2020;15(8):1435–43. https://doi.org/10.1007/s11739-020-02475-0.
2. Booth AL, Abels E, McCaffrey P. Development of a prognostic model for mortality in COVID-19 infection using machine learning. Mod Pathol. 2021;34(3):522–31.
3. Zhu F, Li D, Zhang Q, et al. Prediction of COVID-19 severity from chest CT and laboratory measurements: evaluation of a machine learning approach. JMIR Med Inform. 2020;8:1–12.
4. Chow DS, Glavis-Bloom J, Soun JE, et al. Development and external validation of a prognostic tool for COVID-19 critical disease. PLoS One. 2020;15(12):e0242953. https://doi.org/10.1371/journal.pone.0242953.
5. Jimenez-Solem E, Petersen TS, Hansen C, et al. Developing and validating COVID-19 adverse outcome risk prediction models from a bi-national European cohort of 5594 patients. Sci Rep. 2021;11(1):3246. https://doi.org/10.1038/s41598-021-81844-x.
6. Razavian N, Major VJ, Sudarshan M, et al. A validated, real-time prediction model for favorable outcomes in hospitalized COVID-19 patients. NPJ Digit Med. 2020;3:130. https://doi.org/10.1038/s41746-020-00343-x.
7. Ko H, Chung H, Kang WS, et al. An artificial intelligence model to predict the mortality of COVID-19 patients at hospital admission time using routine blood samples: development and validation of an ensemble model. J Med Internet Res. 2020;22(12):e25442.
8. Li X, Ge P, Zhu J, et al. Deep learning prediction of likelihood of ICU admission and mortality in COVID-19 patients using clinical variables. PeerJ. 2020;8:e10337. https://doi.org/10.7717/peerj.10337.

9. Hou W, Zhao Z, Chen A, et al. Machining learning predicts the need for escalated care and mortality in COVID-19 patients from clinical variables. Int J Med Sci. 2021;18(8):1739–45. https://doi.org/10.7150/ijms.51235.
10. Schöning V, Liakoni E, Baumgartner C, et al. Development and validation of a prognostic COVID-19 severity assessment (COSA) score and machine learning models for patient triage at a tertiary hospital. J Transl Med. 2021;19:56. https://doi.org/10.1186/s12967-021-02720-w.
11. Aktar S, Ahamad MM, Rashed-Al-Mahfuz M, et al. Machine learning approach to predicting COVID-19 disease severity based on clinical blood test data: statistical analysis and model development. JMIR Med Inform. 2021;9(4):e25884. https://doi.org/10.2196/25884.
12. Karthikeyan A, Garg A, Vinod PK, et al. Machine learning based clinical decision support system for early COVID-19 mortality prediction. Front Public Health. 2021;9:626697. https://doi.org/10.3389/fpubh.2021.626697.
13. Castro VM, Sacks CA, Perlis RH, et al. Development and external validation of a delirium prediction model for hospitalized patients with Coronavirus Disease 2019. J Acad Consult Liaison Psychiatry. 2021;62(3):298–308. https://doi.org/10.1016/j.jaclp.2020.12.005.
14. Chen Y, Linli Z, Lei Y, et al. Risk factors for mortality in critically ill patients with COVID-19 in Huanggang, China: a single-center multivariate pattern analysis. J Med Virol. 2021;93(4):2046–55. https://doi.org/10.1002/jmv.26572.
15. Marcos M, Belhassen-García M, Sánchez-Puente A, et al. Development of a severity of disease score and classification model by machine learning for hospitalized COVID-19 patients. PLoS One. 2021;16(4):e0240200. https://doi.org/10.1371/journal.pone.0240200.
16. Statsenko Y, Al Zahmi F, Habuza T, et al. Prediction of COVID-19 severity using laboratory findings on admission: informative values, thresholds, ML model performance. BMJ Open. 2021;11(2):e044500. https://doi.org/10.1136/bmjopen-2020-044500.
17. Woo SH, Rios-Diaz AJ, Kubey AA, et al. Development and validation of a web-based severe COVID-19 risk prediction model. Am J Med Sci. 2021;362(4):355–62. https://doi.org/10.1016/j.amjms.2021.04.001.
18. Luo J, Zhou L, Feng Y, et al. The selection of indicators from initial blood routine test results to improve the accuracy of early prediction of COVID-19 severity. PLoS One. 2021;16(6):e0253329. https://doi.org/10.1371/journal.pone.0253329.
19. Yu L, Halalau A, Dalal B, et al. Machine learning methods to predict mechanical ventilation and mortality in patients with COVID-19. PLoS One. 2021;16(4):e0249285. https://doi.org/10.1371/journal.pone.0249285.
20. Famiglini L, Bini G, Carobene A, et al. Prediction of ICU admission for COVID-19 patients: a machine learning approach based on complete blood count data. In: 2021 Proceedings—IEEE symposium on computer-based medical systems, 2021-June, pp. 160–5.
21. Pulgar-Sánchez M, Chamorro K, Fors M, et al. Biomarkers of severe COVID-19 pneumonia on admission using data-mining powered by common laboratory

blood tests-datasets. Comput Biol Med. 2021;136:104738. https://doi.org/10.1016/j.compbiomed.2021.104738.
22. Chen Y, Ouyang L, Bao FS, et al. A multimodality machine learning approach to differentiate severe and nonsevere COVID-19: model development and validation. J Med Internet Res. 2021;23(4):e23948. https://doi.org/10.2196/23948.
23. Nguyen S, Chan R, Cadena J, et al. Budget constrained machine learning for early prediction of adverse outcomes for COVID-19 patients. Sci Rep. 2021;11(1):19543. https://doi.org/10.1038/s41598-021-98071-z.
24. Lee DS, Ma S, Chu A, et al. CORONA Collaboration. Predictors of mortality among long-term care residents with SARS-CoV-2 infection. J Am Geriatr Soc. 2021;69(12):3377–88. https://doi.org/10.1111/jgs.17425.
25. Chung H, Park C, Kang WS, et al. Gender bias in artificial intelligence: severity prediction at an early stage of COVID-19. Front Physiol. 2021;12.
26. Gabbay F, Bar-Lev S, Montano O, et al. A lime-based explainable machine learning model for predicting the severity level of covid-19 diagnosed patients. Appl Sci (Switzerland). 2021;11(21).
27. Qomariyah NN, Andi Purwita A, Atas Asri SD, et al. A Tree-based mortality prediction model of COVID-19 from routine blood samples. In: 2021 8th International conference on ICT for smart society: digital twin for smart society, ICISS 2021—Proceedings.
28. Mahboub B, Bataineh MTA, Alshraideh H, et al. Prediction of COVID-19 hospital length of stay and risk of death using artificial intelligence-based modeling. Front Med (Lausanne). 2021;8:592336. https://doi.org/10.3389/fmed.2021.592336.
29. Darapaneni N, Gupta M, Paduri AR. A novel machine learning based screening method for high-risk covid-19 patients based on simple blood exams. In: 2021 IEEE International IOT, electronics and mechatronics conference, IEMTRONICS 2021—Proceedings.
30. Tang G, Luo Y, Lu F, et al. Prediction of sepsis in COVID-19 using laboratory indicators. Front Cell Infect Microbiol. 2021;10:586054. https://doi.org/10.3389/fcimb.2020.586054.
31. Zhu JS, Ge P, Jiang C, et al. Deep-learning artificial intelligence analysis of clinical variables predicts mortality in COVID-19 patients. J Am Coll Emerg Physicians Open. 2020;1(6):1364–73. https://doi.org/10.1002/emp2.12205.
32. Lu JQ, Musheyev B, Peng Q, et al. Neural network analysis of clinical variables predicts escalated care in COVID-19 patients: a retrospective study. PeerJ. 2021;9:e11205. https://doi.org/10.7717/peerj.11205.
33. Chung H, Ko H, Kang WS, et al. Prediction and feature importance analysis for severity of COVID-19 in South Korea using artificial intelligence: model development and validation. J Med Internet Res. 2021;23(4):e27060. https://doi.org/10.2196/27060.
34. Ahmed F, Hossain MS, Islam RU, et al. An evolutionary belief rule-based clinical decision support system to predict COVID-19 severity under uncertainty. Appl Sci. 2021;11(13):5810. https://doi.org/10.3390/app11135810.

Appendix 2: Tool Online

It is an example of an IT tool, available online, for the application of one of the diagnostic models of ML evaluated in this review (D4). This model was obtained using the data of patients belonging to the emergency department (ED) of the San Raffaele Hospital and the Galeazzi Orthopedic Institute and was validated using the datasets of patients from the ED of Desio and Monza Hospitals, and from the Papa Giovanni XXIII Hospital of Bergamo (27, 38). The tool is accessible free of charge at: https://covid19-bloodtests-ml.herokuapp.com

The tool requires the insertion, through a mask, of the blood count data, including the leukocyte formula (as absolute counts), age (the only mandatory data), sex, the possible presence of symptoms related to COVID (data entered as presence/absence), and if known, the prevalence of the disease. RDW and MPV parameters are not included in the model. The characteristics of the variables used by the application (parameter, expected input, use in the model, and unit of measurement) are shown in Fig. 5.4.

The indication of the presence or absence of COVID-related symptoms (fever, cough, dyspnea, sore throat, loss of taste or smell, neurological symptoms, and others) is intended to differentiate patients who present in ED with suspected COVID from patients who present for different reasons (trauma, gynecological problems,

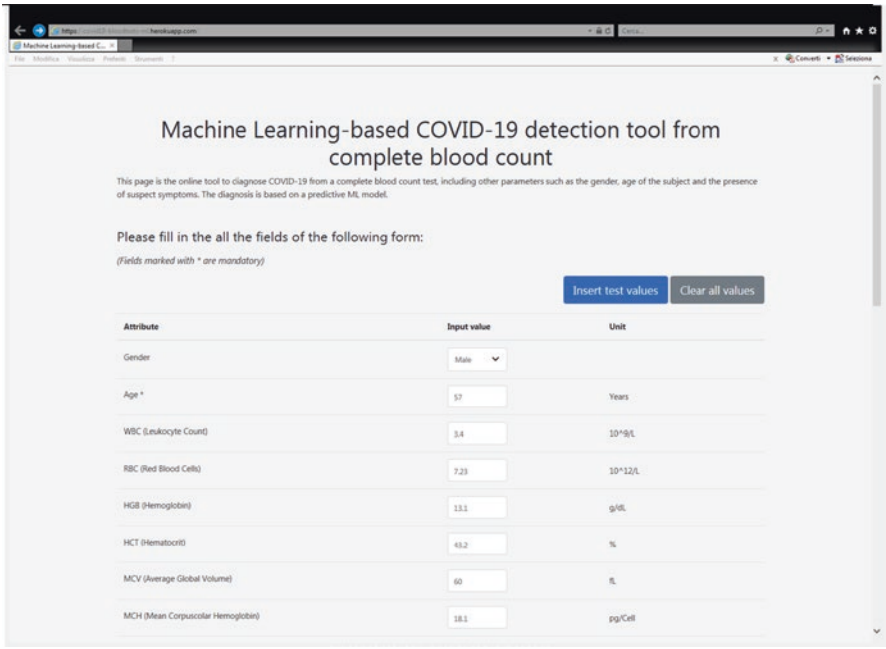


Fig. 5.4 CBC data entry mask for COVID-19 diagnosis based on a machine-learning model available at: https://covid19-bloodtests-ml.herokuapp.com

intestinal obstructions, and others) and to whom, in any case, a nasopharyngeal swab has been requested.

After entering the data in the form (Fig. 5.4), by pressing the "Submit" button the result of the model is obtained as "positive" or "negative", accompanied by a probability score expressed as a percentage. This result can be visualized in two different ways: stick (stick view) or circular (Fig. 5.5a, b respectively).

In the stick view (Fig. 5.5a), the score is displayed as color intensity: the more intense the color, the greater the probability of the result (the intensity of the band can be compared with that of the control, C, taken as a reference of 100%). In the circular view (Fig. 5.5b), the score is displayed as the position of the circle along the horizontal dimension of the figure (from the negative to the positive end), while the size of the circle represents the confidence interval of the score (the smaller the diameter, the greater the confidence).

By way of example, Figs. 5.4 and 5.5 show real data of a COVID positive patient with an estimated probability of 83.8% of actually being positive. The application also includes an "Insert test values" button with which you can simulate a test example.

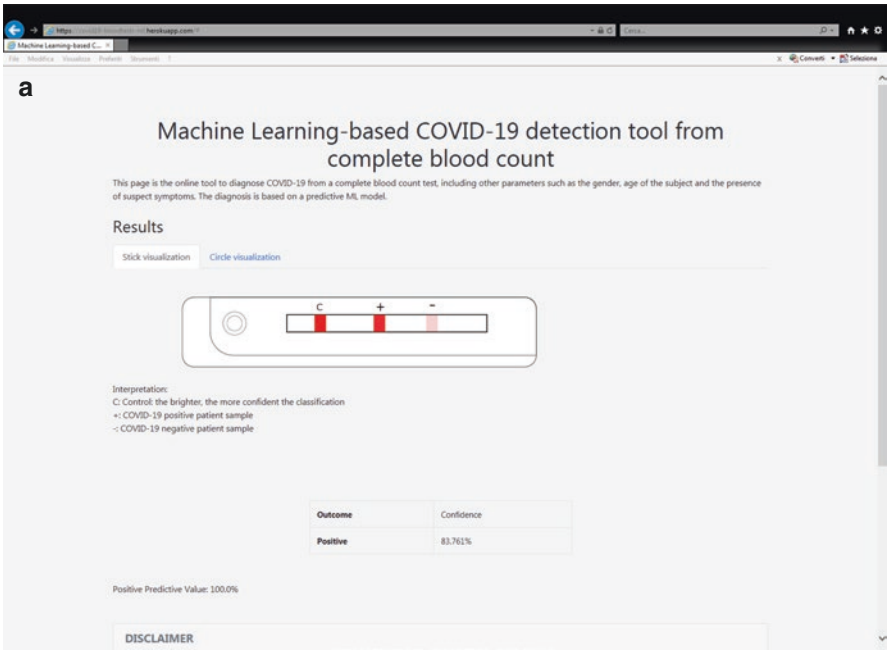


Fig. 5.5 Expression of the result of the machine learning model in terms of "positive" or "negative" accompanied by the confidence of the result in probabilistic terms, with stick (**a**) or circular (**b**) visualization

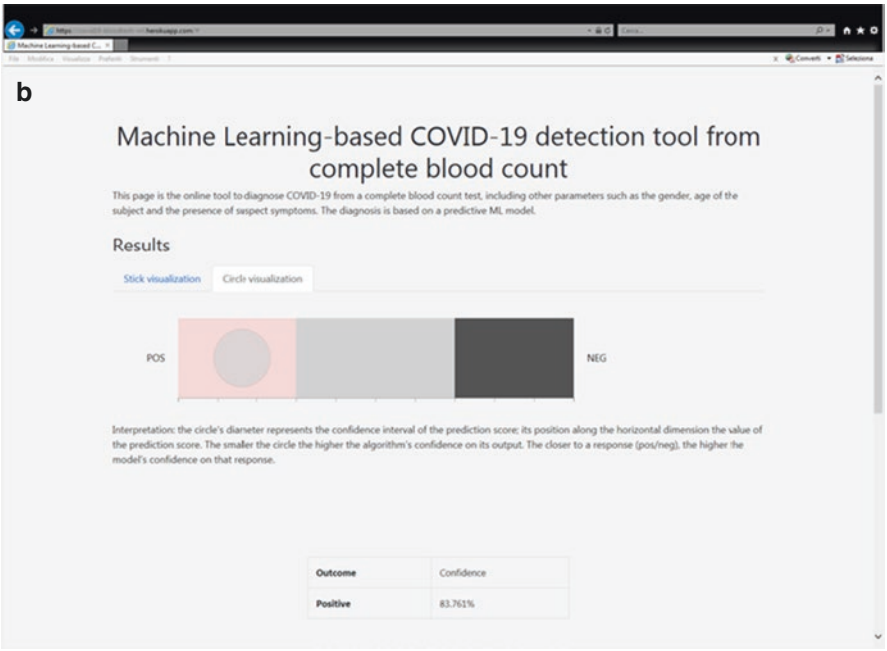


Fig. 5.5 (continued)

References

1. WHO. WHO Coronavirus (COVID-19) Dashboard. 2020. https://covid19.who.int/. Accessed Dec 2021.
2. Oran DP, Topol EJ. Prevalence of asymptomatic SARS-CoV-2 infection: a narrative review. Ann Intern Med. 2020;173:362–7. https://doi.org/10.7326/M20-3012.
3. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, et al. Euro Surveill. 2020;25:2000045.
4. Dinnes J, Deeks JJ, Adriano A, Berhane S, Davenport C, Dittrich S, Cochrane COVID-19 Diagnostic Test Accuracy Group, et al. Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. Cochrane Database Syst Rev. 2020;8(8):CD013705. https://doi.org/10.1002/14651858.CD013705.
5. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of Covid-19: systematic review and critical appraisal. BMJ. 2020;369:1328. https://doi.org/10.1136/bmj.m1328.
6. Guerranti R, Padoan A, Angeletti D, et al. Introduzione ai big data e all'Intelligenza artificiale in medicina di laboratorio. Biochim Clin. 2021;45:57–67.
7. Naugler C, Church DL. Automation and artificial intelligence in the clinical laboratory. Crit Rev Clin Lab Sci. 2019;56:98–110. https://doi.org/10.1080/10408363.2018.1561640.
8. Razavian N, Major VJ, Sudarshan M, Burk-Rafel J, Stella P, Randhawa H, et al. A validated, real-time prediction model for favorable outcomes in hospitalized COVID-19 patients. NPJ Digit Med. 2020;3:130. https://doi.org/10.1038/s41746-020-00343-x.

9. PubMed. National Library of Medicine. 2021. https://pubmed.ncbi.nlm.nih.gov/.
10. Ronzio L, Cabitza F, Barbaro A, Banfi G. Has the flood entered the basement? A systematic literature review about machine learning in laboratory medicine. Diagnostics (Basel). 2021;11:372. https://doi.org/10.3390/diagnostics11020372.
11. Cabitza F, Banfi G. Machine learning in laboratory medicine: waiting for the flood? Clin Chem Lab Med. 2018;56(4):516–24. https://doi.org/10.1515/cclm-2017-0287.
12. Carobene A, Sabetta E, Monteverde E, Locatelli M, Banfi G, Di Resta C, et al. Machine learning based on laboratory medicine test results in diagnosis and prognosis for COVID-19 patients: a systematic review. Biochim Clin. 2021;348:364. https://doi.org/10.19186/BC_2021.046.
13. Herman DS, Rhoads DD, Schulz WL, Durant TJS. Artificial intelligence and mapping a new direction in laboratory medicine: a review. Clin Chem. 2021;67:1466–82. https://doi.org/10.1093/clinchem/hvab165.
14. Mei X, Lee HC, Diao KY, Huang M, Lin B, Liu C, et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. Nat Med. 2020;26:1224–8. https://doi.org/10.1038/s41591-020-0931-3.
15. Ulhaq A, Born J, Khan A, Gomes DPS, Chakraborty S, Paul M. COVID-19 control by computer vision approaches: a survey. IEEE Access. 2020;8:179437–56. https://doi.org/10.1109/ACCESS.2020.3027685.
16. Fan BE, Chong VCL, Chan SSW, Lim GH, Lim KGE, Tan GB, et al. Hematologic parameters in patients with COVID-19 infection. Am J Hematol. 2020;95:131–4. https://doi.org/10.1002/ajh.25774.
17. Ferrari D, Seveso A, Sabetta E, Ceriotti D, Carobene A, Banfi G, et al. Role of time-normalized laboratory findings in predicting COVID-19 outcome. Diagnosis (Berl). 2020;7:387–94. https://doi.org/10.1515/dx-2020-0095.
18. Ferrari D, Cabitza F, Carobene A, Locatelli M. Routine blood tests as an active surveillance to monitor COVID-19 prevalence. A retrospective study. Acta Biomed. 2020;91(3):e2020009. https://doi.org/10.23750/abm.v91i3.10218.
19. Ferrari D, Carobene A, Campagner A, Cabitza F, Sabetta E, Ceriotti D, et al. Evidence of significant difference in key COVID-19 biomarkers during the Italian lockdown strategy. A retrospective study on patients admitted to a hospital emergency department in Northern Italy. Acta Biomed. 2020;91(4):e2020156. https://doi.org/10.23750/abm.v91i4.10371.
20. Vidali M. i big data e la medicina di laboratorio. Biochim Clin. 2021;45:13–4. https://doi.org/10.19186/BC_2021.00.
21. Deo RC. Machine learning in medicine. Circulation. 2015;132:1920–30. https://doi.org/10.1161/CIRCULATIONAHA.115.001593.
22. Luo Y, Szolovits P, Dighe AS, Baron JM. Using machine learning to predict laboratory test results. Am J Clin Pathol. 2016;145:778–88. https://doi.org/10.1093/ajcp/aqw064.
23. Rosado J, Pelleau S, Cockram C, Merkling SH, Nekkab N, Demeret C, et al. Multiplex assays for the identification of serological signatures of SARS-CoV-2 infection: an antibody-based diagnostic and machine learning study. Lancet Microbe. 2021;2:60–9. https://doi.org/10.1016/S2666-5247(20)30197-X.
24. Badrick T, Banfi G, Bietenbeck A, Cervinski MA, Loh TP, Sikaris K. Machine learning for clinical chemists. Clin Chem. 2019;65:1350–6. https://doi.org/10.1373/clinchem.2019.307512.
25. Cabitza F, Campagner A, Soares F, García de Guadiana-Romualdo L, Challa F, Sulejmani A, et al. The importance of being external. Methodological insights for the external validation of machine learning models in medicine. Comput Methods Programs Biomed. 2021;208:106288. https://doi.org/10.1016/j.cmpb.2021.
26. Formica V, Minieri M, Bernardini S, Ciotti M, D'Agostini C, Roselli M, et al. Complete blood count might help to identify subjects with high probability of testing positive to SARS-CoV-2. Clin Med (Lond). 2020;20:114–9. https://doi.org/10.7861/clinmed.2020-0373.
27. Cabitza F, Campagner A, Ferrari D, Di Resta C, Ceriotti D, Sabetta E, et al. Development, evaluation, and validation of machine learning models for COVID-19 detection based on routine blood tests. Clin Chem Lab Med. 2020;59:421–31. https://doi.org/10.1515/cclm-2020-1294.

28. Avila E, Kahmann A, Alho C, Dorn M. Hemogram data as a tool for decision-making in COVID-19 management: applications to resource scarcity scenarios. PeerJ. 2020;8:9482. https://doi.org/10.7717/peerj.9482.
29. Banerjee A, Ray S, Vorselaars B, Kitson J, Mamalakis M, Weeks S, et al. Use of machine learning and artificial intelligence to predict SARS-CoV-2 infection from full blood counts in a population. Int Immunopharmacol. 2020;86:106705. https://doi.org/10.1016/j.intimp.2020.106705.
30. Famiglini L, Bini G, Carobene A, Campagner A, Cabitza F. Prediction of ICU admission for COVID-19 patients: a machine learning approach based on complete blood count data. In: IEEE 34th International symposium on computer-based medical systems, CBMS. Cham: Springer; 2021. p. 160–5. https://doi.org/10.1109/CBMS52027.2021.00065.
31. Buoro S, Carobene A, Seghezzi M, Manenti B, Dominoni P, Pacioni A, et al. Short- and medium-term biological variation estimates of red blood cell and reticulocyte parameters in healthy subjects. Clin Chem Lab Med. 2018;56:954–63. https://doi.org/10.1515/cclm-2017-0902.
32. Buoro S, Carobene A, Seghezzi M, Manenti B, Pacioni A, Ceriotti F, et al. Short- and medium-term biological variation estimates of leukocytes extended to differential count and morphology-structural parameters (cell population data) in blood samples obtained from healthy people. Clin Chim Acta. 2017;473:147–56. https://doi.org/10.1016/j.cca.2017.07.009.
33. Buoro S, Seghezzi M, Manenti B, Pacioni A, Carobene A, Ceriotti F, et al. Biological variation of platelet parameters determined by the Sysmex XN hematology analyzer. Clin Chim Acta. 2017;470:125–32. https://doi.org/10.1016/j.cca.2017.05.004.
34. Coskun A, Braga F, Carobene A, Tejedor Ganduxe X, Aarsand AK, Fernández-Calle P, et al. Systematic review and meta-analysis of within-subject and between-subject biological variation estimates of 20 haematological parameters. Clin Chem Lab Med. 2019;58:25–32. https://doi.org/10.1515/cclm-2019-0658.
35. Vidali M, Carobene A, Apassiti Esposito S, Napolitano G, Caracciolo A, Seghezzi M, et al. Standardization and harmonization in hematology: instrument alignment, quality control materials, and commutability issues. Int J Lab Hematol. 2021;43:364–71. https://doi.org/10.1111/ijlh.13379.
36. Scopus. https://www.scopus.com/search/form.uri?display=basic#basic. Accessed Dec 2021.
37. Kaggle, Einstein Dataset. 2020. https://www.kaggle.com/einsteindata4u/covi. Accessed Dec 2021.
38. Carobene A, Campagner A, Sulejmani A, Leoni V, Seghezzi M, Buoro S, et al. Identification of Sars-CoV-2 positivity using machine learning methods on complete blood count data: external validation of state-of-the-art models. Biochim Clin. 2021;45:281–9.
39. Lippi G, Plebani M. Laboratory abnormalities in patients with COVID-2019 infection. Clin Chem Lab Med. 2020;58:1131–4. https://doi.org/10.1515/cclm-2020-0198.
40. Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, Villamizar-Peña R, Holguin-Rivera Y, Escalera-Antezana JP, et al. Clinical, laboratory and imaging features of COVID-19: a systematic review and meta-analysis. Travel Med Infect Dis. 2020;34:101623. https://doi.org/10.1016/j.tmaid.2020.101623.
41. Carobene A, Ceriotti F, Infusino I, Frusciante E, Panteghini M. Evaluation of the impact of standardization process on the quality of serum creatinine determination in Italian laboratories. Clin Chim Acta. 2014;427:100–6. https://doi.org/10.1016/j.cca.2013.10.001.
42. Paroni R, Fermo I, Cighetti G, Ferrero CA, Carobene A, Ceriotti F. Creatinine determination in serum by capillary electrophoresis. Electrophoresis. 2004;25:463–8. https://doi.org/10.1002/elps.200305687.
43. Favaloro EJ, Thachil J. Reporting of D-dimer data in COVID-19: some confusion and potential for misinformation. Clin Chem Lab Med. 2020;58:1191–9. https://doi.org/10.1515/cclm-2020-0573.
44. National Academies of Sciences, Engineering, and Medicine, Policy and Global Affairs, Committee on Science, Engineering, Medicine, and Public Policy. Reproducibility and replicability in science. Washington, DC: National Academies Press; 2019.

45. Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (MINimum Information for Medical AI Reporting): developing reporting standards for artificial intelligence in health care. J Am Med Inform Assoc. 2020;27:2011–5. https://doi.org/10.1093/jamia/ocaa088.
46. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Standards for Reporting of Diagnostic Accuracy Group, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Croat Med J. 2003;44:639–50.
47. Bartlett WA, Braga F, Carobene A, Coşkun A, Prusa R, Fernandez-Calle P, et al. A checklist for critical appraisal of studies of biological variation. Clin Chem Lab Med. 2015;53:879–85. https://doi.org/10.1515/cclm-2014-1127.
48. Zhang GM, Guo XX, Zhu BL, Zhang GM, Bai SM, Wang HJ, et al. Establishing reference intervals of aspartate aminotransferase-to-platelet ratio index for apparently healthy elderly. Clin Lab. 2016;62:135–40. https://doi.org/10.7754/clin.lab.2015.150609.
49. Aarsand AK, Røraas T, Fernandez-Calle P, Ricos C, Díaz-Garzón J, Jonker N, et al. The biological variation data critical appraisal checklist: a standard for evaluating studies on biological variation. Clin Chem. 2018;64:501–14. https://doi.org/10.1373/clinchem.2017.281808.
50. Vesper HW, Myers GL, Miller WG. Current practices and challenges in the standardization and harmonization of clinical laboratory tests. Am J Clin Nutr. 2016;104:907–12. https://doi.org/10.3945/ajcn.115.110387.
51. Carobene A, Campagner A, Uccheddu C, Banfi G, Vidali M, Cabitza F. The multicenter European Biological Variation Study (EuBIVAS): a new glance provided by the principal component analysis (PCA), a machine learning unsupervised algorithms, based on the basic metabolic panel linked measurands. Clin Chem Lab Med. 2021;60(4):556–68. https://doi.org/10.1515/cclm-2021-0599.
52. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal–external, and external validation. J Clin Epidemiol. 2016;69:245–7. https://doi.org/10.1016/j.jclinepi.2015.04.005.
53. Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;54:774–81. https://doi.org/10.1016/s0895-4356(01)00341-9.
54. Vidali M. La scienza riproducibile. Biochim Clin. 2020;44:386–96.
55. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. NPJ Digit Med. 2020;3:126. https://doi.org/10.1038/s41746-020-00333-z.
56. Gao Q, Hu Y, Dai Z, Xiao F, Wang J, Wu J. The epidemiological characteristics of 2019 novel coronavirus diseases (COVID-19) in Jingmen, Hubei, China. Medicine (Baltimore). 2020;99(23):e20605.
57. Dooling K, McClung N, Chamberland M, Marin M, Wallace M, Bell BP, et al. The advisory committee on immunization practices' interim recommendation for allocating initial supplies of COVID-19 vaccine—United States, 2020. MMWR Morb Mortal Wkly Rep. 2020;69(49):1857–9.
58. Pouquet M, Decarreaux D, Prévot-Monsacré P, Hervé C, Werner A, Grosgogeat B, et al. Nationwide seroprevalence of SARS-CoV-2 IgG antibodies among four groups of primary healthcare workers and their household contacts 6 months after the initiation of the COVID-19 vaccination campaign in france: seroPRIM study protocol. Pathogens. 2021;10(1):911.
59. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. N Engl J Med. 2020;383(27):2603–15. https://doi.org/10.1056/nejmoa2034577.
60. Chia WN, Zhu F, Ong SWX, Young BE, Fong SW, Le Bert N, et al. Dynamics of SARS-CoV-2 neutralizing antibody responses and duration of immunity: a longitudinal study. Lancet Microbe. 2021;2(6):e240–9.
61. Huang AT, Garcia-Carreras B, Hitchings MDT, Yang B, Katzelnick LC, Rattigan SM, et al. A systematic review of antibody mediated immunity to coronaviruses: kinetics, correlates of protection, and association with severity. Nat Commun. 2020;11(1):4704.

62. Campo F, Venuti A, Pimpinelli F, Abril E, Blandino G, Conti L, et al. Antibody persistence 6 months post-vaccination with BNT162b2 among health care workers. Vaccine. 2021;9(10):1125.
63. Di Resta C, Ferrari D, Viganò M, Moro M, Sabetta E, Minerva M, et al. The gender impact assessment among healthcare workers in the SARS-CoV-2 vaccination—an analysis of serological response and side effects. Vaccine. 2021;9(5):522.
64. Naaber P, Tserel L, Kangro K, Sepp E, Jürjenson V, Adamson A, et al. Dynamics of antibody response to BNT162b2 vaccine after 6 months: a longitudinal prospective study. Lancet Reg Health Eur. 2021;1:10.
65. Ray D, Yung R. Immune senescence, epigenetics and autoimmunity. Clin Immunol. 2018;196:59–63. https://doi.org/10.1016/j.clim.2018.04.002.
66. Ferrari D, Clementi N, Spanò SM, Albitar-Nehme S, Ranno S, Colombini A, et al. Harmonization of six quantitative SARS-CoV-2 serological assays using sera of vaccinated subjects. Clin Chim Acta. 2021;1(522):144–51.
67. Gayle H, Foege W, Brown L, Kahn B. Framework for equitable allocation of COVID-19 vaccine. Washington, DC: National Academy of Sciences; 2020. p. 1–252.
68. Mellado B, Wu J, Kong JD, Bragazzi NL, Asgary A, Kawonga M, et al. Leveraging artificial intelligence and big data to optimize COVID-19 clinical public health and vaccination roll-out strategies in Africa. Int J Environ Res Public Health. 2021;18:7890.
69. https://ourworldindata.org/grapher/share-people-fully-vaccinated-covid. Accessed 20 Jan 2022.
70. Coates T, Wilson R, Patrick G, André F, Watson V. Hepatitis B vaccines: Assessment of the seroprotective efficacy of two recombinant DNA vaccines. Clin. Ther. 2001;23:392–403.
71. Freer G, Pistello M. Varicella-Zoster virus infection: natural history, clinical manifestations, immunity and current and future vaccination strategies. New Microbiol. 2018;41(2):95–105.
72. Drolet M, Laprise JF, Martin D, Jit M, Bénard É, Gingras G, et al. Optimal human papillomavirus vaccination strategies to prevent cervical cancer in low-income and middle-income countries in the context of limited resources: a mathematical modelling analysis. Lancet Infect Dis. 2021;21(11):1598–610.
73. Mannino S, Villa M, Apolone G, Weiss NS, Groth N, Aquino I, et al. Effectiveness of adjuvanted influenza vaccination in elderly subjects in northern Italy. Am J Epidemiol. 2012;176(6):527–33.
74. Russell FM, Greenwood B. Who should be prioritized for COVID-19 vaccination? Hum Vaccin Immunother. 2021;17(5):1317–21. https://doi.org/10.1080/21645515.2020.1827882.
75. Bubar KM, Reinholt K, Kissler SM, Lipsitch M, Cobey S, Grad YH, et al. Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. Science (80-). 2021;371(6532):916–21.
76. Fitzpatrick BMC, Galvani AP. Optimizing age-specific vaccination. Science. 2021;37(6532):890–2.
77. Ioannou GN, Green P, Fan VS, Dominitz JA, O'Hare AM, Backus LI, et al. Development of COVIDVax model to estimate the risk of SARS-CoV-2-related death among 7.6 million US veterans for use in vaccination prioritization. JAMA Netw Open. 2021;4(4):e214347.
78. Matrajt L, Eaton J, Leung T, Brown ER. Vaccine optimization for COVID-19: who to vaccinate first? Sci Adv. 2021;7:1374. https://www.science.org.
79. Pressman AR, Lockhart SH, Shen Z, Azar KMJ. Measuring and promoting SARS-CoV-2 vaccine equity: development of a COVID-19 vaccine Equity Index. Health Equity. 2021;5(1):476–83.

Chapter 6
AI and the Infectious Medicine of COVID-19

Vardan Andriasyan, Anthony Petkidis, and Artur Yakimovich

Abbreviations

AAK1	AP2-associated protein kinase 1
AI	Artificial Intelligence
AST	Aminotransferase
CNN	Convolutional neural network
COVID-19	Coronavirus disease 2019
CRP	C-reactive protein
CT	Computed tomography
CXR	Chest X-Ray
DL	Deep learning
ELISA	Enzyme-linked immunosorbent assay
GISAID	Global initiative on sharing all influenza data
LDH	Lactate dehydrogenase
LSTM	Long short-term memory
ML	Machine learning
PCR	Polymerase chain reaction
QSAR	Quantitative structure-activity relationship
RT-LAMP	Reverse transcription loop-mediated isothermal amplification
RT-qPCR	Reverse transcription quantitative polymerase chain reaction

V. Andriasyan · A. Petkidis
Department of Molecular Life Sciences, University of Zurich, Zurich, Switzerland

Artificial Intelligence for Life Sciences CIC, Dorset, UK
e-mail: vardan.andriasyan@uzh.ch; anthony.petkidis@uzh.ch

A. Yakimovich (✉)
Artificial Intelligence for Life Sciences CIC, Dorset, UK

Bladder Infection and Immunity Group (BIIG), Department of Renal Medicine, Division of Medicine, University College London, Royal Free Hospital Campus, London, UK

Center for Advanced Systems Understanding (CASUS), Helmholtz-Zentrum Dresden-Rossendorf e.V. (HZDR), Görlitz, Germany
e-mail: a.yakimovich@hzdr.de

N. Lidströmer, Y. C. Eldar (eds.), *Artificial Intelligence in Covid-19*,
https://doi.org/10.1007/978-3-031-08506-2_6

RWD	Real world data
RWE	Real world evidence
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2

Introduction

The public health crisis caused by the COVID-19 pandemic has among other things exposed the need for novel data-driven techniques in Infectious Medicine. Big data collection and analysis are surrounding the COVID-19 patient journey already today as a result of digital transformation of healthcare providers (Fig. 6.1). These capabilities are growing rapidly, aimed at improving the speed and precision of biomedical research and development, real-world data (RWD) and evidence (RWE) generation, clinical decision support or diagnostics. However, the opportunities that artificial intelligence (AI) and machine learning (ML) bring to the field of Infectious Medicine come with their own caveats. Heterogeneous, small or proprietary datasets make the development of novel machine learning techniques challenging in this

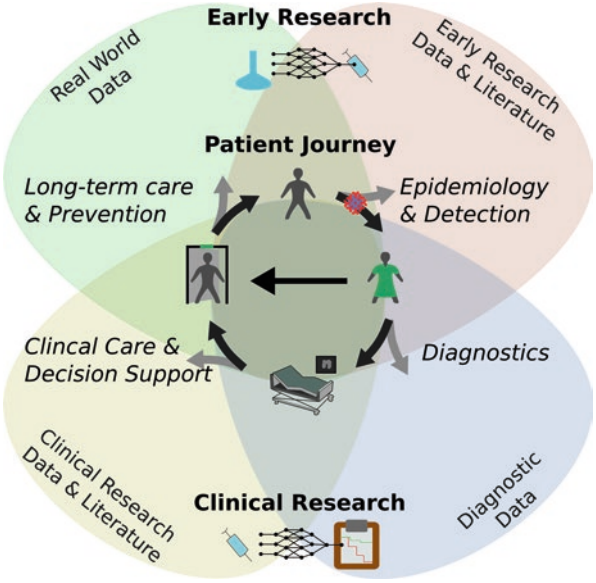


Fig. 6.1 Schematic overview of the role data and artificial intelligence may play in COVID-19 patient journey. Diagram shows in a clockwise circular fashion the patient journey from infection to diagnosis, to disease, to recovery, and finally to long-term sequelae and potential reinfection. Tangential to the patient journey, the diagram depicts intersection of real world data, early research data and literature, diagnostic data, clinical research data and literature contributing to epidemiology and detection of SARS-CoV-2, diagnostics of COVID-19, clinical care and decision support, and finally long-term care and prevention. These data sources contribute to the early research and clinical research as depicted on the diagram

domain. Conversely, simply applying techniques from other domains to biomedical data without due validation and safety evaluation may prove catastrophic.

The pandemic has drawn the minds of the global research community to the problems surrounding Infectious Medicine resulting in an avalanche of research works on preprint servers and in the peer-reviewed journals. To make the works published on AI application to Infectious Medicine available at a glance, in this chapter we have structured them into the following topics of AI and ML for: SARS-CoV-2 early research using pathogen sequence data, research of SARS-CoV-2 antivirals, COVID-19 Infectious Medicine early research using language data, Real World Data Analysis of COVID-19, molecular diagnostics of COVID-19, image-based diagnostics of COVID-19 and clinical decision support, and COVID-19 medical care. While all connected to COVID-19, these topics touch upon various aspects of the COVID-19 patient journey (described in Fig. 6.1). Hence, the research undertaken in these topics often has rather different terminology and tools. While the works reviewed in this chapter are far from a comprehensive list, we believe this list provides a perspective on the breadth of AI application to the Infectious Medicine of COVID-19.

AI and ML for SARS-CoV-2 Early Research Using Pathogen Sequence Data

From the moment of identification of the novel coronavirus species, which later became widely known as SARS-CoV-2 [1], to the most recent research on SARS-CoV-2 variants [2, 3], sequencing data played a pivotal role in guiding humanity through the current pandemic. In fact, it would be fair to say that the growing capacity in pathogen sequencing is what makes the currently ongoing SARS-CoV-2 pandemic different from the pandemics of the past. With large sequence datasets available through projects like the global initiative on sharing all influenza data (GISAID) [4], the Bioinformatics research community is looking to ML techniques to help make sense of the developing epidemiological situation. For example, Park et al. have developed an ML-powered scoring approach capable of linking pathogen genome sequence to case fatality rate [5]. To achieve this authors integrated a number of well-established approaches into a statistical metamodel allowing them to highlight putative parts of the variant genome. Another example by Saha and co-authors employs a deep learning (DL) approach based on Long Short-Term Memory (LSTM) architecture [6] to attempt a retrospective SARS-CoV-2 prediction from a number of pathogen sequences [7]. Accompanied by the k-mer approach to generate Bag-of-Unique-Descriptors set of features from pathogen sequence, authors show remarkable performance of their methodology. Taking the complexity a step further, Mateos and colleagues proposed a DL-based detector of SARS-CoV-2 co-infection with another RNA-virus [8]. Employing raw RNA-seq data their proposed algorithm detected the presence and relative concentrations of viruses with >99%

precision and recall. While without any doubt these examples show the field at its infancy, it is very likely that more successful applications of the novel ML techniques will follow suit.

AI and ML for Research of SARS-CoV-2 Antivirals

The SARS-CoV-2 pandemic exposed not only the dangers of the emerging pathogens, but also the dire lack of broad range antivirals available off-the-shelf. In the face of an excruciatingly lengthy drug development process and a healthcare emergency buildup early in the pandemic, unsurprisingly, drug repurposing—i.e. an approach of finding well-tested drugs for off-label use (reviewed in [9]) became a popular line of research. Aiming at developing the second line of defense to support the vaccination campaign with SARS-CoV-2 antivirals, several researchers proposed a number of computational approaches to sift through the databases of existing drugs with well-demonstrated safety profiles. For example de Siqueira Santos and colleagues proposed a tandem of two approaches [10]. First approach employed a matrix factorization algorithm for broad-spectrum antivirals ranking. Second, utilized graph kernels to rank drugs according to the perturbation they cause on a human interactome subnetwork. Both approaches that de Siqueira Santos et al. proposed were capable of focusing on host and pathogen factors equally.

Richardson and colleagues employed the proprietary BenevolentAI's knowledge graph they built from structured medical information, including numerous connections extracted from scientific literature using ML [11]. Using this approach they focused on AP2-associated protein kinase 1 (AAK1). Their analysis yielded six high-affinity leads binding AAK1, of which they selected one—the janus kinase inhibitor baricitinib. Given that the pharmacokinetics and pharmacodynamics of baricitinib are well established Richardson and colleagues proposed a clinical trial for this compound. In another graph-based approach [12], authors proposed a COVID-19 related biological network aiming mostly at virus targets or their associated biological processes (i.e. host factors). Authors proposed a set of features based on protein-protein interactions, as well as drug-target interactions. Using these features and two clustering algorithms (spectral partitioning and betweenness value) Aghdam and co-authors identify five clusters containing drugs putatively applicable for repurposing as SARS-CoV-2 antivirals. Rather than looking at the interactome or drugs as indivisible entities, a study performed by Gawriljuk and colleagues focused on using ML for extracting the quantitative structure–activity relationship (QSAR) of small molecules against viral host factors [13]. Specifically, authors employed Bernoulli Naïve Bayes and other Bayesian models, AdaBoost Decision trees, Random Forest, support vector machine classifier, k-nearest neighbors, and DL combined with five-fold cross validation. Their approach predicted seven candidates for repurposing with varying confidence. Remarkably, authors have also performed the validation of identified leads using *in vitro* tissue culture.

AI and ML for COVID-19 Infectious Medicine Early Research Using Language Data

Coinciding with the beginning of the SARS-CoV-2 pandemic, recent advances in DL for natural language processing (NLP) including transformer architectures [14] provided data scientists with tools to analyze large text datasets at scale. To harness the ability of these methods to analyze research literature related to SARS-CoV-2 in an automated fashion, Wang and colleagues proposed collating relevant research articles into a large public dataset they named COVID-19 Open Research Dataset (CORD-19) [15]. This in turn, sparked the attention of NLP researchers to infection biology and emerging pathogens. For example, Köksal and colleagues proposed a framework employing CORD-19 to search for protein-compound pairs in the research literature [16]. While admittedly somewhat prone to biases [17], CORD-19 has demonstrated the applicability of NLP techniques and AI to research literature in infection biology. Automated search for protein-compound interactions may well contribute to accelerating antiviral drug discovery. Similar efforts include more narrow application datasets aimed to address more advanced NLP tasks, e.g. question answering. Möller et al. for instance, proposed using a pre-trained transformer model, which they fine-tuned on their own question answering dataset [18].

AI and ML in Real World Data Analysis of COVID-19

With sizes of medical data repositories growing and with real world data (RWD) and real world evidence (RWE) gaining incredible traction, it is clear that ML may be used beyond clinical and pre-clinical research [19]. Needless to say, with a global scale event like the SARS-CoV-2 pandemic, RWD and RWE are playing a crucial role in understanding the rapid development of the situation (see Fig. 6.1). This, for example, can be seen in the global effort to collect SARS-CoV-2 variants sequencing data [4], as well as accumulation of large epidemiological data [20]. Such datasets allow to perform remarkable retrospective *in silico* studies, infeasible previously due to sheer scale requirements. For example, Park et al. have used the global sequencing data to model disease severity of SARS-CoV-2 variants in a metaviromic fashion [5]. In another example, de Figueiredo and colleagues have used such an approach to build a model aimed at mapping global trends in vaccination uptake, as well as understanding global barriers to uptake [21]. However, it is important to understand that RWD does not equal RWE, and many datasets come with an inherent bias, which may be exacerbated through modern day ML algorithms [22].

AI and ML in Molecular Diagnostics of COVID-19

A rapid and reliable identification of infected individuals and their subsequent isolation has been a key measure in the fight against the spread of the pandemic. Common methods for detection of acute or past COVID-19 infection include testing for the presence of viral genetic material, viral proteins, or virus-targeting antibodies in the patient's serum [23]. The current gold standard for confirmation of suspected COVID-19 cases is reverse transcription quantitative polymerase chain reaction (RT-qPCR). In this method, viral RNA is first reverse transcribed into DNA and later exponentially amplified using sequence-specific primers, which leads to high sensitivity and allows the quantification of the starting material. Similarly, reverse transcription loop-mediated isothermal amplification (RT-LAMP) allows the detection of viral genome fragments, but compared to RT-qPCR alleviates the need for temperature changes during the amplification procedure, enabling a portable detection system. Rapid antigen tests are implemented as lateral flow tests, in which chromatography is combined with immunodetection of a viral protein of interest, typically the nucleocapsid protein. Detection of virus-targeting antibodies in blood samples, which may indicate a past infection, is possible by enzyme-linked immunosorbent assay (ELISA).

While detection of viral genetic material in nasopharyngeal or oropharyngeal swabs by RT-PCR is the recommended procedure for confirming SARS-CoV-2 infection by WHO [24], shortages of reagents, certified labs, and trained personnel can impede a timely diagnosis. To aid rapid COVID-19 detection, several ML methods have been developed to draw information from routine blood samples. Studies have shown that increased blood levels of aspartate aminotransferase (AST), lactate dehydrogenase (LDH), C-reactive protein (CRP) and ferritin, which arise as consequences of tissue damage and inflammation, can serve as diagnostic markers for COVID-19 [25, 26]. Diagnostic indicators also include low lymphocyte counts, and low leukocyte counts in general. Such laboratory results can also be combined with other ML methods for COVID-19 diagnosis to improve the classification accuracy of chest computed tomography (CT) imaging [27], which is further discussed below.

AI and ML in Image-Based Diagnostics of COVID-19 and Clinical Decision Support

Methods like chest X-Ray (CXR) and CT scans are common medical imaging techniques, which are widely used to find the infection spread within the patient's lungs. Unlike RT-qPCR, medical imaging techniques are non-invasive and cost-effective techniques that are ubiquitous worldwide. CT scans are capable of visualizing organs in a 3D representation including the thin tissues. These techniques are widely used to detect abnormalities in lungs such as tumors, excess fluids, and pneumonia. COVID-19 being a similar disease can be well diagnosed using CT scans. CXRs

detect and represent the dense tissues in the organs in 2D view, which require much less sophisticated hardware. This, in turn, makes it easily affordable for hospitals and laboratories all over the world. Numerous experiments showed that CT scans detect significant differences between various respiratory diseases including COVID-19 [28]. In addition, during the COVID-19 pandemic in Italy, using CXR in conjunction with RT-qPCR reduced the fluctuating sensitivity of the RT-qPCR test findings [29]. CXRs were also useful for severe pneumonia and critically sick COVID-19 patients in terms of diagnosis and therapy [30]. With pathology information, CXRs assist radiologists in confirming the existence of COVID-19 infection. An epidemic condition requires not just early diagnosis but also the most accurate and sensitive test findings to assist therapy. Clinical investigations of COVID-19 diagnosis and other comparable respiratory illnesses reveal that CT scans and CXRs can be utilized for early patient categorization and infection severity evaluation. CXRs are less expensive than CT scans, yet most countries have an adequate supply of CT equipment. Because both CT scans and CXRs are equally useful in identifying COVID-19 infection, using both imaging modalities can aid in the screening of COVID-19 patients.

There has been a substantial amount of research publications on the application of DL and convolutional neural network (CNN) techniques in the identification of COVID-19 patients utilizing CXR and CT images. For example, Xu et al. [31] developed a prediction model to distinguish COVID-19 pneumonia from influenza. The maximum degree of precision achieved by the prediction model was 86 percent. Apostolopoulos and Mpesiana [32] evaluated the performance of the CNN architectures developed in recent years (VGGNet-19, MobileNet-v2, Inception, Xception, and Inception ResNet-v2) as tools for automatic detection of coronavirus disease using medical imaging data. The researchers employed a process known as Transfer Learning which utilizes already learned features from common image datasets (e.g. CIFAR). This approach boosts the overall performance on poorly defined and irregular small datasets, which is typical for medical images. Their dataset consisted of 1442 X-ray images of patients, including 714 with bacterial and viral pneumonia, 224 with proven COVID-19 illness, and 504 with healthy lungs. According to the results, VGGNet-19 and MobileNet-v2 have the highest classification accuracy. Ai et al. [33] created a CNN based on the InceptionV2, Inception-ResNetV3, and ResNet50 models to classify COVID-19 chest radiographs into COVID-19 positive and negative classes. They identified a high correlation between CT imaging and PCR results. In [34] the authors utilized InceptionV3 to identify COVID-19 abnormalities in lung CT scans. The InceptionNet model was tested on 1065 CT scans, and 325 infected patients were detected with an accuracy of 85 percent. In [35] a more systematic approach was taken. According to the results of their numerical experiments, the VGGNet-19 model outperforms the ResNet-50, InceptionV3, Xception on the CT image dataset, achieving 88.5 percent precision, 86 percent recall, 86.5 percent F1-score, and 87 percent accuracy. Meanwhile the tuned Xception version gives the highest precision, recall, F1-score, and accuracy values of 98 percent respectively on the CXR image dataset. Furthermore, by

averaging the two modalities, CXR and CT, VGG-19 achieved the best score, which is 90.5 percent for accuracy and the F1-score, 90.3 percent for recall, and 91.5 percent for precision.

Noteworthy, despite the enormous research efforts to construct ML models for COVID-19 diagnosis and prognosis, there are methodological problems and several biases throughout the literature, leading to overly optimistic reported results. review article demonstrates that the majority of published ML models for image-based diagnostics and clinical decision support [36] are not plausible candidates for clinical translation in their currently reported form, as they lack high-quality datasets, are hard to reproduce and cross validate. To this end, in [37] authors implemented the EfficientNetB0 model. The model was tested on several independent datasets of CXR images and validated by RT-qPCR. The model reached over 95% diagnostic accuracy and was cross validated against the performance of trained radiologists, particularly in detecting early disease. Additionally, the authors provide an automated pipeline for dynamical image retrieval and segmentation from hospital chest radiography equipment. Such an approach can vastly facilitate and standardize COVID-19 patient care.

AI and ML in COVID-19 Medical Care

Application of ML methods can be particularly valuable for predicting epidemiological developments, infection risk factors and disease prognosis [19]. Such forecasts can inform political decision-making in public health and guide dynamic adjustments of epidemic measures as well as allocation of medical support. While substantial efforts have been taken to create reliable forecasting models, the urgency of the pandemic has imposed severe time constraints for the development of robust AI tools, leaving most in an experimental state [38].

Prevention, Infection Risk and Epidemiology

A logistic regression algorithm for risk prediction of future SARS-CoV-2 infection in a cohort of nearly 12,000 patients found age, ethnicity and sex as factors associated with infection risk [39]. Conversely, a study performed by Roy and Gosh found little to no impact of ethnicity and sex, while highlighting population density, testing rate and airport traffic as important contributing factors for disease spread [40]. It is now acknowledged that epidemiological development and transmission risk factors can be highly dependent on local circumstances, even within the same country [41].

To predict epidemiological development globally and in individual countries, logistic and LSTM models have been employed with some success [42, 43]. The

susceptible-infected-removed (SIR) model has also been used to simulate the spread of COVID-19 [44]. Despite potentially providing useful heuristics that can aid containment measures, accurate forecasting was shown to be computationally intractable [45].

Treatment and Prognosis

ML has provided insight into factors associated with a severe course of disease, identifying patients that may require additional medical attention. Using logistic regression analysis, studies have found a positive correlation between increased mortality and high levels of the pro-inflammatory cytokine IL-6 as well as low CD8+ T cell counts [46, 47]. Metaanalyses found that shortness of breath (dyspnoea), smoking, old age (>75 years), male sex, obesity, and various other comorbidities increase the mortality of hospitalized COVID-19 patients [48, 49]. Combining a CNN with autoencoders for the analysis of CT images was a feasible method for survival chance prediction in [50].

Conclusions

The call to fight the global pandemic has been well received by the global AI and ML researchers. The sheer volume of research contributions on AI and ML in Infectious Medicine of the COVID-19 has led to the development of multiple new subfields of biomedical AI. However, in many cases it would be premature to expect an immediate application of this research in fighting the pandemic. Critically, the clinical application requires all software to meet strict safety standards, as well as obtain meticulous regulatory approval. The necessity for such hurdles lies in painful lessons of the past, where software malfunctions caused loss of health and life (e.g. the infamous Therac-25 incidents). Furthermore, as the biomedical data analysis becomes more and more dependent on balckbox models these lessons of the past are ever so poignant.

At the same time, the increased influx of data science practitioners to Infectious Medicine shows great promise for the field. Once the necessary quality and reproducibility standards are established for ML and DL application to COVID-19 and other infectious disease data, direct clinical application may become a reality. Furthermore, it is hard to overestimate the importance of AI for early research and development. For example, drug repurposing is already profiting significantly from algorithmic improvements resulting in improvement in lead generation. Perhaps a pattern where AI techniques are applied complementary to other more traditional approaches may become prevalent in the future.

Cross-References

Symptom Based Detection Models of COVID-19 Infraction Using AI, *AI and Point Of Care Image Analysis For COVID-19*, *AI Techniques for Forecasting Epidemic Dynamics: Theory and Practice*, *AI and Biochemical Testing and Laboratory Values of COVID-19*, *Usage of AI for Anti-Viral Therapies in Present and Future Pandemics*.

References

1. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579:265–9.
2. Cele S, Jackson L, Khoury DS, Khan K, Moyo-Gwete T, Tegally H, et al. Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. Nature. 2021. https://doi.org/10.1038/s41586-021-04387-1.
3. Wang L, Cheng G. Sequence analysis of the emerging SARS-CoV-2 variant Omicron in South Africa. J Med Virol. 2021. https://doi.org/10.1002/jmv.27516.
4. Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data – from vision to reality. Euro Surveill 2017;22. https://doi.org/10.2807/1560-7917.es.2017.22.13.30494.
5. Park JJ, Chen S. Metaviromic identification of discriminative genomic features in SARS-CoV-2 using machine learning. Patterns (N Y). 2022;3(2):100407.
6. Hochreiter S, Schmidhuber J. LSTM can solve hard long time lag problems. Adv Neural Inf Process Syst. 1997;9:473–9.
7. Saha I, Ghosh N, Maity D, Seal A, Plewczynski D. COVID-DeepPredictor: recurrent neural network to predict SARS-CoV-2 and other pathogenic viruses. Front Genet. 2021;12:569120.
8. Acera Mateos P, Balboa RF, Easteal S, Eyras E, Patel HR. PACIFIC: a lightweight deep-learning classifier of SARS-CoV-2 and co-infecting RNA viruses. Sci Rep. 2021;11:3209.
9. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov. 2019;18:41–58.
10. de Siqueira Santos S, Torres M, Galeano D, Sánchez MDM, Cernuzzi L, Paccanaro A. Machine learning and network medicine approaches for drug repositioning for COVID-19. Patterns (N Y). 2022;3:100396.
11. Richardson P, Griffin I, Tucker C, Smith D, Oechsle O, Phelan A, et al. Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. Lancet. 2020;395:e30–1.
12. Aghdam R, Habibi M, Taheri G. Using informative features in machine learning based method for COVID-19 drug repurposing. J Cheminform. 2021;13:70.
13. Gawriljuk VO, Zin PPK, Puhl AC, Zorn KM, Foil DH, Lane TR, et al. Machine Learning Models Identify Inhibitors of SARS-CoV-2. J Chem Inf Model. 2021;61:4224–35.
14. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv [cs.CL]. 2018. Available: http://arxiv.org/abs/1810.04805.
15. Lu Wang L, Lo K, Chandrasekhar Y, Reas R, Yang J, Eide D, et al. CORD-19: the covid-19 open research dataset. ArXiv. 2020. Available: https://www.ncbi.nlm.nih.gov/pubmed/32510522.
16. Köksal A, Dönmez H, Özçelik R, Ozkirimli E, Özgür A. Vapur: a search engine to find related protein-compound pairs in COVID-19 literature. arXiv [cs.IR]. 2020. Available: http://arxiv.org/abs/2009.02526.
17. Kanakia A, Wang K, Dong Y, Xie B, Lo K, Shen Z, et al. Mitigating biases in CORD-19 for analyzing COVID-19 literature. Front Res Metr Anal. 2020;5:596624.
18. Möller T, Reina A, Jayakumar R, Pietsch M. COVID-QA: a question answering dataset for COVID-19. 2020. Available: https://openreview.net/pdf?id=JENSKEEzsoU.
19. Bachtiger P, Peters NS, Walsh SL. Machine learning for COVID-19-asking the right questions. Lancet Digit Health. 2020;2:e391–2.

20. Xu B, Gutierrez B, Mekaru S, Sewalk K, Goodwin L, Loskill A, et al. Epidemiological data from the COVID-19 outbreak, real-time case information. Sci Data. 2020;7:106.
21. de Figueiredo A, Simas C, Karafillakis E, Paterson P, Larson HJ. Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modelling study. Lancet. 2020;396:898–908.
22. Sáez C, Romero N, Conejero JA, García-Gómez JM. Potential limitations in COVID-19 machine learning due to data source variability: a case study in the nCov2019 dataset. J Am Med Inform Assoc. 2021;28:360–4.
23. Vandenberg O, Martiny D, Rochas O, van Belkum A, Kozlakidis Z. Considerations for diagnostic COVID-19 tests. Nat Rev Microbiol. 2021;19:171–83.
24. World Health Organization, Others. Laboratory testing for 2019 novel coronavirus (2019-nCoV) in suspected human cases: interim guidance, 14 January 2020. World Health Organization; 2020. Available: https://apps.who.int/iris/bitstream/handle/10665/332300/WHO-2019-nCoV-laboratory-2020.2-eng.pdf.
25. Goodman-Meza D, Rudas A, Chiang JN, Adamson PC, Ebinger J, Sun N, et al. A machine learning algorithm to increase COVID-19 inpatient diagnostic capacity. PLoS One. 2020;15:e0239474.
26. Brinati D, Campagner A, Ferrari D, Locatelli M, Banfi G, Cabitza F. Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study. J Med Syst. 2020;44:135.
27. Mei X, Lee H-C, Diao K-Y, Huang M, Lin B, Liu C, et al. Artificial intelligence–enabled rapid diagnosis of patients with COVID-19. Nat Med. 2020;26:1224–8.
28. Song F, Shi N, Shan F, Zhang Z, Shen J, Lu H, et al. Emerging 2019 novel coronavirus (2019-nCoV) pneumonia. Radiology. 2020;295:210–7.
29. Schiaffino S, Tritella S, Cozzi A, Carriero S, Blandi L, Ferraris L, et al. Diagnostic performance of chest X-Ray for COVID-19 pneumonia during the SARS-CoV-2 pandemic in Lombardy, Italy. J Thorac Imaging. 2020;35:W105–6.
30. Wu G, Li X. Mobile X-rays are highly valuable for critically ill COVID patients. Eur Radiol. 2020;30(9):5217–9.
31. Xu X, Jiang X, Ma C, Du P, Li X, Lv S, et al. A deep learning system to screen novel coronavirus disease 2019 pneumonia. Engineering. 2020;6:1122–9.
32. Apostolopoulos ID, Mpesiana TA. Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. Australas Phys Eng Sci Med. 2020;43:635–40.
33. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology. 2020;296:E32–40.
34. Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, et al. A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19). Eur Radiol. 2021;31:6096–104.
35. Chouat I, Echtioui A, Khemakhem R, Zouch W, Ghorbel M, Hamida AB. COVID-19 detection in CT and CXR images using deep learning models. Biogerontology. 2022. https://doi.org/10.1007/s10522-021-09946-7.
36. Roberts M, Driggs D, Thorpe M, Gilbey J, Yeung M, Ursprung S, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nat Mach Intell. 2021;3:199–217.
37. Kim CK, Choi JW, Jiao Z, Wang D, Wu J, Yi TY, et al. An automated COVID-19 triage pipeline using artificial intelligence based on chest radiographs and clinical data. NPJ Digit Med. 2022;5:5.
38. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ. 2020;369:m1328.
39. Jehi L, Ji X, Milinovich A, Erzurum S, Rubin BP, Gordon S, et al. Individualizing risk prediction for positive coronavirus disease 2019 testing: results from 11,672 patients. Chest. 2020;158:1364–75.

40. Roy S, Ghosh P. Factors affecting COVID-19 infected and death rates inform lockdown-related policymaking. PLoS One. 2020;15:e0241165.
41. Hou X, Gao S, Li Q, Kang Y, Chen N, Chen K, et al. Intracounty modeling of COVID-19 infection with human mobility: assessing spatial heterogeneity with business traffic, age, and race. Proc Natl Acad Sci U S A. 2021;118:e2020524118. https://doi.org/10.1073/pnas.2020524118.
42. Wang P, Zheng X, Li J, Zhu B. Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. Chaos Solit Fract. 2020;139:110058.
43. Wang P, Zheng X, Ai G, Liu D, Zhu B. Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: case studies in Russia, Peru and Iran. Chaos Solit Fract. 2020;140:110214.
44. Cooper I, Mondal A, Antonopoulos CG. A SIR model assumption for the spread of COVID-19 in different communities. Chaos Solit Fract. 2020;139:110057.
45. Rosenkrantz DJ, Vullikanti A, Ravi SS, Stearns RE, Levin S, Poor HV, et al. Fundamental limitations on efficiently forecasting certain epidemic measures in network models. Proc Natl Acad Sci U S A. 2022;119:e2109228119. https://doi.org/10.1073/pnas.2109228119.
46. Luo M, Liu J, Jiang W, Yue S, Liu H, Wei S. IL-6 and CD8+ T cell counts combined are an early predictor of in-hospital mortality of patients with COVID-19. JCI Insight. 2020;5:e139024. https://doi.org/10.1172/jci.insight.139024.
47. Belaid B, Lamara Mahammad L, Mihi B, Rahali SY, Djidjeli A, Larab Z, et al. T cell counts and IL-6 concentration in blood of North African COVID-19 patients are two independent prognostic factors for severe disease and death. J Leukoc Biol. 2022;111:269–81.
48. Mesas AE, Cavero-Redondo I, Álvarez-Bueno C, Sarriá Cabrera MA, Maffei de Andrade S, Sequí-Dominguez I, et al. Predictors of in-hospital COVID-19 mortality: a comprehensive systematic review and meta-analysis exploring differences by age, sex and health conditions. PLoS One. 2020;15:e0241742.
49. Booth A, Reed AB, Ponzo S, Yassaee A, Aral M, Plans D, et al. Population risk factors for severe disease and mortality in COVID-19: a global systematic review and meta-analysis. PLoS One. 2021;16:e0247461.
50. Khozeimeh F, Sharifrazi D, Izadi NH, Joloudari JH, Shoeibi A, Alizadehsani R, et al. Combining a convolutional neural network with autoencoders to predict the survival chance of COVID-19 patients. Sci Rep. 2021;11:15343.

Chapter 7
AI and ICU Monitoring on the Wake of the COVID-19 Pandemic

Alfredo Vellido and Vicent Ribas

Introduction

Clinical medicine has a unique place in society as a hub between science and practice, between the advancement of knowledge and its direct application as a service to society. Such sensitive role requires both the consideration of ethics and a strong governance and regulation: from the Hippocratic Oath to the European Medical Device Regulation, from clinical guidelines to medical ethics committees.

The modern concept of EBM talks about the use of current best evidence in clinical decision making, opening the door to new forms of non-traditional evidence, such as those made available by the digitalization and networking of the medical environment. Among those, none with higher impact and potential than data-centred approaches making the best of the synergies between the advances in computing and in data-acquisition devices at the point of care. In this context, and mediated by the push of large IT companies trying to muscle their way into the health and medical markets, AI, most usually in the form of ML, with its Deep Learning (DL) variant, is likely to play a central role [1].

A. Vellido (✉)
Computer Science Department, Universitat Politècnica de Catalunya (UPC BarcelonaTech), Barcelona, Spain

Intelligent Data Science and Artificial Intelligence (IDEAI-UPC) Research Center, Barcelona, Spain

Centro de Investigación Biomédica En Red, Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Madrid, Spain
e-mail: avellido@cs.upc.edu

V. Ribas
Eurecat Centre Tecnològic de Catalunya, eHealth Unit, Barcelona, Spain
e-mail: vicent.ribas@eurecat.org

N. Lidströmer, Y. C. Eldar (eds.), *Artificial Intelligence in Covid-19*,
https://doi.org/10.1007/978-3-031-08506-2_7

The modern ICU (probably more than most other clinical settings) is a major data generator and, therefore, a natural challenge for AI data-centred analysis. In this brief chapter, we summarily review the ascending impact of AI in the pre-pandemic ICU and the critical care domain. This followed by an assessment of the ways in which AI-based methods are addressing the massive challenges for the ICU posed by the COVID-19 pandemic that started in 2020.

ICU Monitoring Through AI

Please re-name this heading and add your own subheadings.

Cross-References
Include a list of related entries from the handbook here that may be of further interest to the readers.

ICU Monitoring and AI in Pre-pandemic Times

Critical care caters for critically ill patients who depend on life support and invasive monitoring techniques. They are therefore technologically dependent on life-sustaining medical devices.

Intensive care specialists oversee the "chain of survival" and, thus, apply cardiopulmonary resuscitation, defibrillation and, if necessary, advanced life support. This requires assessing whether a patient will suffer a sudden deterioration that may require resuscitation.

As a result, monitoring techniques based on advanced medical devices and the need of critical patient status prediction coalesce to make data-based decision-making very relevant in critical care. This decision making can be based on data analysed using AI and ML methods, an area that is making quick and substantial advances.

One of the main problems in the ICU is the management of sepsis and its associated organ dysfunction. The integration of multi-level (i.e. socio-demographics, epidemiological, genetic, systemic and cellular-level data) and multi-modal data (such as patient records, monitoring data, lab tests, biomarkers and so on) through ML-based solutions provides a valuable opportunity to further understand the mechanisms of organ dysfunction, its phenotyping, and assessing the clinical response to interventions at the onset (or prior to) of organ dysfunction [2].

In this context, the current ML trend in critical care is the integration of multi-modal data with point of care devices for assessing the specific protein and metabolic biomarkers required to develop robust clinical decision support systems for diagnosis and early-management of organ dysfunction. These solutions will enable the adoption of more effective management and personalized care by taking a systems-level approach for fine-tuning treatment and vital support.

Such approaches should ultimately lead to a reduction in time to diagnose, days on vasopressor support, days on mechanical ventilation and, ultimately, of mortality. Data-centred integrated solutions based on AI and ML aim at ensuring:

1. the efficient prediction and assessment of organ dysfunction;
2. the phenotyping and assessment of patients; and
3. the timely delivery of treatment recommendations based on the assessment of the clinical trajectory of each patient.

Further and comprehensive surveys of the of AI and ML applications on critical care are beyond the scope of this chapter and the reader can find in, for instance, [3, 4].

The Impact of the COVID-19 Pandemic on the ICU and the Role of AI

The modern management of ICUs guarantees their smooth operation in normal times, but catastrophic or unpredictable events, such as the COVID-19 pandemic that started in 2020, can quickly overwhelm and throw critical care out of balance, due to lack of material and human resources and even physical space at the hospital. For this reason, data-based prediction of ICU requirements through the development of risk scoring systems for general critically ill patients, may become an extremely useful tool to alleviate that burden through preventive planning. This prediction can naturally be based on ML models [5, 6], although this would require identifying biomarkers that are specific for the COVID-19 patient, as in [7]. A more advanced ML application that, beyond the prediction of critical illness and in-hospital mortality, also helps to predict bed utilization and valuates the impact of patient influx scenarios on utilization, was recently presented in [8]. It used a multistate survival model to analyze a nationwide registry following the day-by-day clinical status of all hospitalized COVID-19 patients in Israel for two months.

COVID-19 deaths are the result of a viral sepsis, which is defined as "life-threatening organ dysfunction caused by a dysregulated host response to infection" [9]. In this context, scores for assessing organ dysfunction have become fundamental for managing COVID-19 patients in the ICU. The new consensus definitions (Sepsis-3) also establish the qSOFA (for quick Sequential Organ Failure Assessment) "[which incorporates], altered mentation, systolic blood pressure of 100 mmHg or less, and respiratory rate of 22/min or greater, [providing] simple bedside criteria to identify adult patients with suspected infection who are likely to have poor outcomes". However, several studies have reported poor performance of this indicator for predicting risk and mortality [10, 11].

Different studies have tried to address this lack of specificity through the definition of patient phenotypes for assessing risk and patient trajectories in the ICU. These phenotypes have provided health practitioners with a detailed and interpretable health status of septic patients (accurate diagnosis, and disease trajectory), whilst setting the basis for the adoption of effective therapy.

For sepsis in general, Seymour et al. [12] defined four different phenotypes of sepsis through probabilistic clustering: low vasopressor titration, patients with chronic conditions and lower renal function, patients with more inflammation and pulmonary dysfunction and patients with liver dysfunction and septic shock. An older study by Ribas et al. [13] also defines 11 phenotypes for sepsis with logistic regression over latent factors. In this study, five of the eleven phenotypes (patients requiring ventilator support, high organ dysfunction, high severity, or impaired hepatic function) were found to be relevant for predicting ICU outcomes.

Besides the agreement between these two independent data-driven approaches for patient phenotyping, they clearly establish a phenotype for critical-care patients with lung dysfunction and requiring ventilator support. In particular, COVID-19 patients present acute hypoxemic respiratory failure (AHRF) as the most common cause for admission in the ICU. The management of AHRF in the ICU is supportive while treating the underlying cause of this lung failure. Traditional ventilatory support includes invasive ventilation (i.e. intubation and coupling with a mechanical ventilator), as well as less invasive methods such as high-flow nasal oxygen (HFNO) and non-invasive ventilation (NIV) for the less critical cases. Regardless of the ventilation approach, [12, 13] clearly show the association of mechanical ventilation with poor ICU and hospital outcomes. The mechanisms underlying hypoxemia in COVID-19 patients is not yet fully understood, so that new models have been proposed for assessing the extension of the pulmonary infection and pulmonary fibrosis.

A recent work by Pezzano et al. [14] proposes model based on DL called the CoLe-CNN+ for assessing ground-glass opacities and lesion segmentation in the lung. Lesion segmentation is fundamental for assessing lung function, the status of the illness, its severity and treatment planning. In fact, the area covered by the lesion approximately corresponds to the area where the pulmonary alveoli are not working normally.

Despite the increase in accuracy detecting COVID-19 lesions through CT images, the reading time necessary to interpret 3D CT volumes and extract the morphological properties of the lesion have had a negative impact in the already overwhelmed radiology services around the world. Indeed, manual segmentation is a very time-consuming task, which takes approximately seven hours to accurately delineate the lesion on a CT scan with 250 slices. In this context, DL strategies can reduce this time significantly and, thus, expedite the lung-function assessment, treatment planning and administration.

A feasible convolutional neural network (CNN) workflow for this task includes the following steps:

1. Lung segmentation,
2. COVID-19 detection,
3. Lesion segmentation.

The lungs are first segmented from the input CT image to reduce the searching area. Afterwards, the detection algorithm is used to analyze the lungs' area in order to detect the presence of COVID-19. In the case of a positive finding, the CT image is processed by the last network (COVID-19 lesion segmentation) in order to identify the areas affected by the disease.

In this study, the authors managed to segment fibrotic lesions caused by COVID-19 (ground-glass opacities) with an impressive 99% accuracy.

This valuable information about lung injury and functionality, can be used for enhancing the phenotypes outlined above through multiscale integration of data. This integration of data at different system levels can lead to more accurate patient trajectory definition and ultimately dynamically assessing bundled responsiveness to therapy.

Conclusions

The digitalization of healthcare systems is quickly advancing data-centred approaches to EBM. Beyond the hype, AI and ML are beginning to be become part of medical devices for decision support. The ongoing COVID-19 pandemic has generated a flurry of activity in the AI community and the resulting research can be of particular use in critical care: for preventive patient risk assessment, critical ward management, organ dysfunction assessment associated to viral sepsis, or lung function monitoring, to name just a few of the potential representative problems that could be tackled using this type of data-based approaches and which we have sketched in this brief chapter.

Sepsis associated COVID-19 has become a prevalent pathology in the ICUs worldwide, and one with relatively high mortality rates associated. Its medical management is therefore both a sensitive issue and a serious challenge to healthcare systems.

The clinical indicators of sepsis currently in use are known to be of limited relevance as outcome predictors. In this context, several phenotyping models have been proposed to mitigate this issue by leveraging upon different data-driven approaches. Even though these phenotyping methods have been developed for assessing the risk associated to patients with organ dysfunction in general, the very definition of these phenotypes clearly shows the importance of lung dysfunction in patient outcomes. Ventilator support being the main feature in this phenotype, which subsumes the COVID-19 patient population in the ICU. This phenotyping can be significantly improved by adding functional and lesion information from CT scans, which has been proven to be a valuable source of information for assessing patient prognosis. The redefinition of patient phenotypes through the integration of data at different system levels with machine-learning techniques can lead to more accurate patient trajectory definition and ultimately dynamically assessing bundled responsiveness to therapy.

Acknowledgements A.V. acknowledges partial funding from Spanish grant PID2019-104551RB-I00.

References

1. Bacciu D, Lisboa PJG, Vellido A, editors. Deep learning in biology and medicine. World Scientific; 2022.
2. Vellido A, Ribas V. Artificial intelligence in critical care: the path from promise to practice. In: Lidstromer N, editor. Artificial intelligence in medicine. Springer; 2022.
3. Vellido A, Ribas V, Morales C, Ruiz-Sanmartín A, Ruiz-Rodríguez JC. Machine Learning for critical care: state-of-the-art and a sepsis case study. Biomed Eng Online. 2018;17(S1):135.
4. Nguyen D, Ngo B, van Sonnenberg E. AI in the intensive care unit: up-to-date review. J Intensive Care Med. 2021;36(10):1115–23.
5. Liang W, Yao J, Chen A, et al. Early triage of critically ill COVID-19 patients using deep learning. Nat Commun. 2020;11:1–7.
6. McRae MP, Simmons GW, Christodoulides AJ, et al. Clinical decision support tool and rapid point-of-care platform for determining disease severity in patients with COVID-19. Lab Chip. 2020;20(12):2075–85.
7. Banoei MM, Dinparastisaleh R, Zadeh AV, Mirsaeidi M. Machine-learning-based COVID-19 mortality prediction model and identification of patients at low and high risk of dying. Crit Care. 2021;25(1):1–14.
8. Roimi M, Gutman R, Somer J, et al. Development and validation of a machine learning model predicting illness trajectory and hospital utilization of COVID-19 patients: a nationwide study. J Am Med Inform Assoc. 2021;28(6):1188–96.
9. Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA. 2016;315(8):801–10.
10. Giamarellos-Bourboulis EJ, Tsaganos T, Tsangaris I, et al. Validation of the new Sepsis-3 definitions: proposal for improvement in early risk identification. Clin Microbiol Infect. 2017;23(2):104–9.
11. Askim Å, Moser F, Gustad LT, et al. Poor performance of quick-SOFA score in predicting severe sepsis and mortality – a prospective study of patients admitted with infection to the emergency department. Scand J Trauma Resusc Emerg Med. 2017;25(1):56.
12. Seymour CW, Kennedy JN, Wang S, et al. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. JAMA. 2019;321(20):2003–17.
13. Ribas VJ, Vellido A, Ruiz-Rodríguez JC, et al. Severe sepsis mortality prediction with logistic regression over latent factors. Expert Syst Appl. 2012;39(2):1937–43.
14. Pezzano G, Díaz O, Ribas-Ripoll V, Radeva P. CoLe-CNN+: context learning - convolutional neural network for COVID-19-ground-glass-opacities detection and segmentation. Comput Biol Med. 2021;136:104689.

Chapter 8
Symptom Based Models of COVID-19 Infection Using AI

Songqiao Liu, Yuan Hong Sun, Alex A. Waese-Perlman, Nathan Yee Lee, Haibo Zhang, and Kang Lee

Introduction

Coronavirus Disease 2019 (COVID-19) caused by the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV2) has spread around the world in a global pandemic [1–4]. Early and daily detection of suspected COVID-19 patients is the most important approach not only for tracing close contacts to prevent further spread [5], but also providing crucial information for healthcare providers and officials to make resource allocation and policy decisions [6].

Most patients with COVID-19 experience a variety of characteristic symptoms. Such symptoms can serve as early indicators of infection as well as severity such that appropriate medical decisions can be made. However, not all COVID-19 symptoms

S. Liu
Zhongda Hospital, Southeast University, Nanjing, China

Y. H. Sun · A. A. Waese-Perlman
Department of Applied Psychology & Human Development, University of Toronto, Toronto, Canada
e-mail: billyuanhong.sun@mail.utoronto.ca

N. Y. Lee
Western University, London, ON, Canada
e-mail: nlee293@uwo.ca

H. Zhang
St. Michael's Hospital, Unity Health Toronto, University of Toronto, Toronto, ON, Canada
e-mail: haibo.zhang@unityhealth.to

K. Lee (✉)
St. Michael's Hospital, Unity Health Toronto, University of Toronto, Toronto, ON, Canada

Department of Applied Psychology & Human Development, University of Toronto, Toronto, Canada
e-mail: kang.lee@utoronto.ca

N. Lidströmer, Y. C. Eldar (eds.), *Artificial Intelligence in Covid-19*,
https://doi.org/10.1007/978-3-031-08506-2_8

are of equal importance. Some symptoms may be indicators of serious infection whereas others may suggest that infection is mild. Therefore, patients with different presentations of symptoms may have different outcomes. Treating patients the same way regardless of their severity and outcomes is clearly not a viable strategy. This is not only because it wastes valuable medical resources but also because some treatments are highly invasive (e.g., ventilator) and should be used only when absolutely necessary. Thus, it is critical that different presentations of COVID-19 symptoms are used to assist medical personnel in making the appropriate treatment decisions for different patients infected with COVID-19.

In addition, COVID-19 symptoms can sometimes be confused with symptoms from other illnesses such as the common cold, influenza, or pneumonia [7, 8]. Such similarities lead to several problems: (1) some patients with other illnesses may seek confirmation by using RT-PCR or serologic assays in COVID-19 clinics, which is unnecessary and wastes precious testing resources; (2) some patients with COVID-19 may not seek clinical confirmation because they misjudge their symptoms as other illnesses, potentially leading them to expose others to the disease unwittingly; and 3) patients with COVID-19 and those with other illnesses may seek medical attention at the same outpatient clinic due to bearing similar symptoms, which may result in cross infection and possibly lead to cluster outbreaks. These problems highlight the urgent need to use different presentations of symptoms in differentiating patients infected with COVID-19 versus others diseases. Such differentiation will assist patients to make appropriate decisions about whether they should seek medical attention in a COVID-19 clinic.

To more effectively identify the symptoms of COVID-19 and predict its risks, machine learning techniques have been successfully applied to large datasets collected from COVID-19 patients and hospitals worldwide since the onset of the pandemic. Unlike the traditional statistical methods such as Generalized Linear Modeling that is only able to detect linear relations between variables, machine learning algorithms are able to detect both linear and nonlinear relations. Machine learning models are therefore more capable than traditional methods for establishing accurate predictive models between the predictor and predicted variables. Utilizing patient demographics and health data, machine learning algorithms have indeed been successfully implemented to predict variable clinical conditions such as mental health disorders [9]. Given this evidence, it follows that machine learning should be capable of classifying patients with COVID-19 infection, as it is believed that COVID-19 presents a unique combination of symptoms.

This chapter reviews the three existing studies that used large scale patient COVID-19 symptom data along with various machine learning techniques. The three use cases are (1) to predict mortality of patients with COVID-19 infection; (2) to determine whether an individual has COVID-19 infection or not, and (3) to differentiate COVID-19 infection from such other diseases as flu and common cold that share similar symptoms.

Using Machine Learning Methods to Determine Mortality of Patient with COVID-19

Quiroz-Juárez et al. [10] investigated whether the machine learning approach can predict mortality of a patient infected with COVID-19 based on a series of predictors including presenting symptoms. To do so, they obtained a dataset with 4.7 million patients from Mexico who were suspected or confirmed to have been infected with COVID-19. Among them, 200,000 were deceased. The dataset was extracted from the database published by the Mexican Federal Government regarding COVID-19 patients from all states in Mexico. The dataset contains a detailed record of each patient's demographics, medical history (including pre-existing conditions), symptoms, hospitalization status (e.g., intubation, ICU) and whether the patient survived or became deceased (Table 8.1).

The patients in the dataset were further divided according to the progression of their treatment at the hospitals or clinics. Stage 1 patients only received initial medical assessment and care. Stage 2 patients had their COVID-19 status confirmed and presented COVID-19 symptoms. Stage 3 patients were admitted into a hospital or sent back home. Stage 4 patients were hospitalized and were either intubated or admitted into an intensive care unit.

Table 8.1 Included and excluded characteristics for the COVID-19 risk assessment model

Category	Included characteristics	Excluded characteristics
Medical History	Diabetes	Pregnancy
	COPD	Smoking
	Immunosuppressive drugs	Asthma
	Hypertension	
	Chronic renal failure	
	Cardiovascular diseases	
Demographics	Gender	Indigenous
	Birth state	Indigenous language
	State of residence	Migrant
	Age	Foreigner
Category 3a recent medical information	USMER designation	Exposure to positive patients
	Medical facility sector	
	State (treatment)	
	Days symptom-treatment	
Category 3b recent medical information	COVID-19 status	
	COVID-19-related pneumonia	
	Hospitalization status	
	Intubation	
	ICU	

The researchers then trained a multilayer perceptron (MLP) neural network with one hidden layer with two sigmoid neurons. The predicted variable was whether the patients survived or became deceased. The dataset was then partitioned, with 70% for training, 15% for testing, and 15% as holdouts. The holdout dataset was never used in training and testing of the model, allowing it to be used as a validation dataset for assessing the generalizability of the trained and tested models. Model performance was evaluated in terms of accuracy, sensitivity, and specificity. The researchers also used additional machine learning techniques, and analyzed the same data in the same fashion to compare whether MLP was a superior technique to use for their purpose of predicting patient mortality. These techniques include logistic regression (LR), support vector machine (SVM), and k-nearest neighbors (kNN). Results showed that MLP overperformed these techniques by a few percentage points in terms of accuracy, sensitivity, and specificity. In addition, the MLP model performance increased when patients progressed from Stage 1 to Stage 4 (see Table 8.2). For example, the mortality prediction accuracy increased from 83.5% at Stage 1, to 93.5% at Stage 4, sensitivity increased from 86.3% to 96.1%, and specificity increased from 82.4% to 90.9%.

In an attempt to explain the MLP model's excellent performance to predict mortality, the researchers further extracted the normalized absolute values of the synaptic weights for the two sigmoid neurons in the hidden layer for each clinical stage. They found that the most important predictors varied from Stage 1 to Stage 4 (Fig. 8.1).

For example, for the first neuron, the top two predictors were days with symptoms and treatment and patient age at Stage 1, COVID-19 related pneumonia and age at Stage 2, whether the patient is taking immune-suppressive drugs and gender at Stage 3, and patient age and days with symptoms and treatment at Stage 4. In contrast, for the second neuron, the top two predictors were patient age and medical facility at Stage 1, patient age and days with symptoms and treatment at Stage 2, patient age and hospitalization status at Stage 3, and days with symptoms and treatment and patient age at Stage 4. Because the MLP model took into consideration all predictors in nonlinear combinations for the two neurons in the hidden layer, Fig. 8.1 only provides a glimpse of how the predictors worked together in the MLP model to achieve high performance in predicting mortality.

The researchers have implemented the model into a web application form where physicians can quickly fill out and identify which COVID-19 patients are high risk (https://www.mathworks.com/matlabcentral/fileexchange/87202-idecovid19).

Table 8.2 Neural Network accuracy, specificity, and sensitivity for clinical stages 1–4

Stage	Accuracy (%)	Specificity (%)	Sensitivity (%)
1	84.3	82.4	86.3
2	90.5	89.1	91.9
3	93.1	90.8	95.5
4	93.5	90.9	96.1

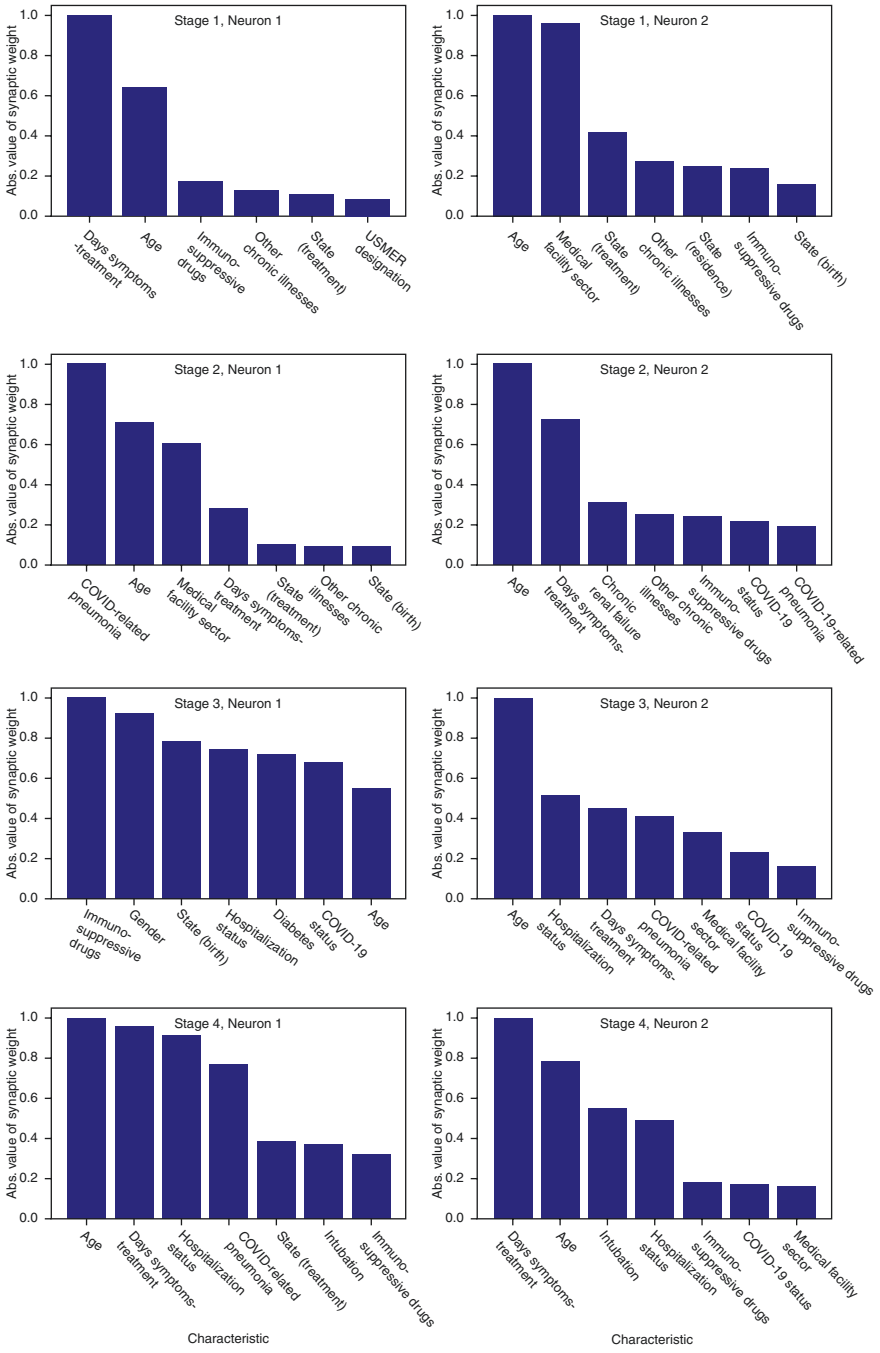


Fig. 8.1 Top seven dominant characteristics for both neurons in the hidden layer ranked by absolute value of synaptic weight

Using Machine Learning Methods to Detect the Presence of COVID-19 Infection

Instead of predicting the severity and mortality of patients infected with COVID-19, Zoabi et al. [11] was among the first to use public health data in determining the likelihood of COVID-19 infection presence based on presenting symptoms. The research team obtained a dataset of 51,831 tested individuals (of whom 4769 were confirmed to have COVID-19) from Israel collected between March 22th, 2020 and March 31st, 2020. This nationwide dataset was released by the Israeli Ministry of Health publicly and contained patients who were tested for SARS-CoV-2 via RT-PCR assay of a nasopharyngeal swab11. The dataset consisted of the following eight predictors: sex (male/female), age $\geq$ 60 years (true or false), cough, fever (true or false), sore throat (true or false), shortness of breath (true or false), headache (true or false), and contact with an individual confirmed to have COVID-19 (true or false).

The researchers used this dataset to train and internally test machine learning models by separating the dataset into a training set (80%) and a testing set (20%). Using a gradient boosting tree-based technique with the LightGBM25 Python package, researchers trained various computational models that could predict whether a patient would be diagnosed with COVID-19 or not. To ascertain the external validity of the computational models, the team validated the models against a prospective dataset from a new dataset released by the Israeli Ministry of Health from April first through April 7th (47,401 tested, of which 3624 were confirmed to have COVID-19).

Out of the models produced by the training-testing process, the best model achieved about 90% AUC (area under the curve) of the ROC (receiver operating characteristic) curve, with excellent sensitivity and specificity (Fig. 8.2). This suggests that the best model provides a good sensitivity and specificity of classifying patients into ones with confirmed COVID-19 infection vs. without such infection. One potential problem of the model's high accuracy could be due to the unbalanced nature of the data. That is, there were far more uninfected participants than infected ones, which could potentially lead to an inflated model accuracy. To address this concern, researchers also tested on a balanced dataset. Although the AUC based on the balanced dataset was 86%, they still successfully replicated the model initial results based on the unbalanced holdout dataset.

The researchers further used SHapley Additive exPlanations (SHAP), based on Shapley's importance values, to determine the most important predictors to predict the model's high performance. Shapley importance value is a solution

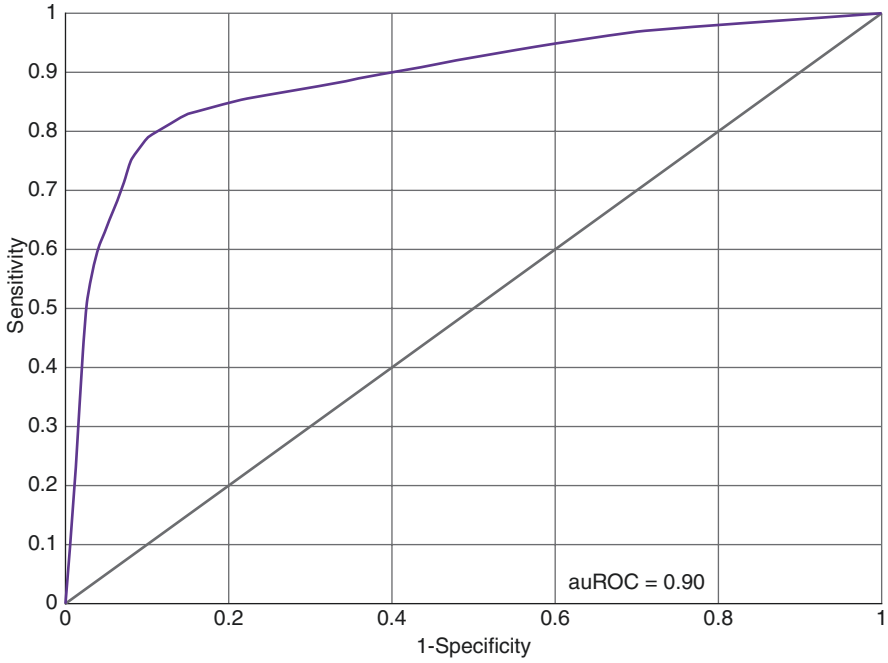


Fig. 8.2 Receiver Operating Characteristic curves of the predictive model of the prospective test set. The light band around the curve represents pointwise 95% confidence intervals derived by bootstrapping. The grey line represents an area under curve (auROC) score of 0.50, or a random predictor [11]

concept based on game theory, first proposed by Nobel Prize winner Lloyd Shapley [12]. Historic use of the Shapley concept is predicated around explaining how to distribute benefits in complex cooperative relations, with the underlying notion that any benefits obtained by individuals should be proportional to the value of their contributions. The concept has seen a repurposing in recent years, being used to provide a novel explanation for why a machine learning model makes certain predictions. This is achieved by determining the relative contribution of each feature to said prediction, also known as the 'feature importance' [13, 14]. The researchers found that among the top predictors, the presence of coughing, fever, and exposure to someone with COVID-19 are the most important for predicting the presence of COVID-19 in a patient (Fig. 8.3).

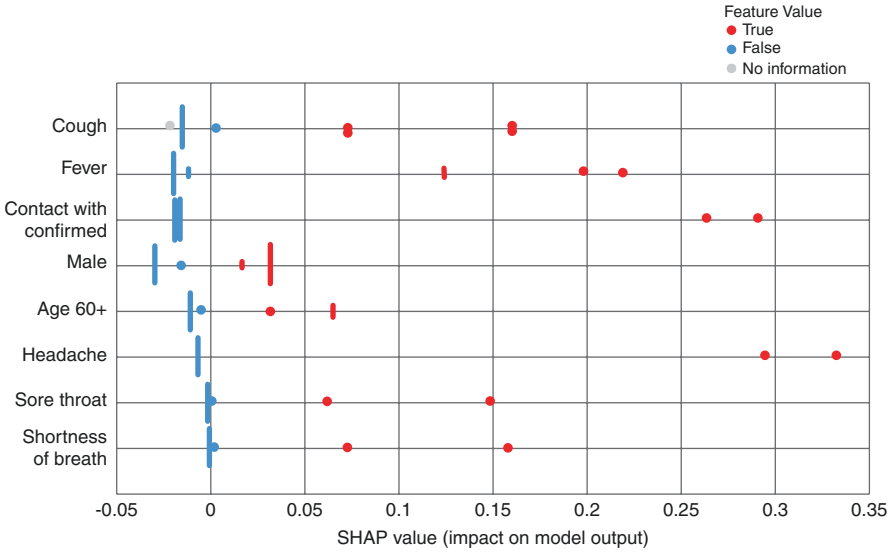


Fig. 8.3 SHapley Additive exPlanations (SHAP) beeswarm plot for the COVID-19 prediction model, showing SHAP values for the most important features of the model. Features in the summary plots (y-axis) are organized by their mean absolute SHAP values. Each point corresponds to an individual person in the study. The position of each point on the x-axis shows the impact that feature has on the classifier's prediction for a given individual. Values of those features (i.e., fever) are represented by their color [11]

Using Machine Learning Methods to Differentiate COVID-19 and Influenza/Common Cold Infections

The algorithm presented by Zoabi et al. [11] was validated by a largely healthy comparison group. As a result, it was unclear whether the machine learning approach could be used to differentiate individuals who are suffering from different kinds of respiratory diseases. To address this question, we conducted a study to investigate whether patients with COVID-19 and common flu or cold could be differentiated using the machine learning approach. We obtained two datasets of patients from 24 hospitals across Jiangsu Province, China. Inclusion criteria were patients with confirmed COVID-19 between January 1 and March 15, 2020 in one dataset (N = 599), and patients with seasonal flu or common cold (January 1 to October 30, 2019) in a large hospital in China in another dataset (N = 832).

To conduct machine learning, the dataset of COVID-19 and that of seasonal flu and common cold patients were randomly divided into three subsets: training (70%), testing (15%), and holdout (15%) (see Fig. 8.4).

Extreme gradient boosting decision trees (XGBoost) (https://github.com/dmlc/xgboost) were used to train the first machine learning model to classify the two cohorts of patients with the training set. XGBoost uses the principle of gradient boosting to construct decision trees from the data for binary classification [15].

Gradient boosting involves an ensemble of predictions from individual decision trees (i.e., weak learners) using boosting techniques. Specifically, prediction errors are minimized by using gradient descent. Additionally, XGBoost also decreases the likelihood of overfitting using regularization techniques. During training, XGBoost decreases the weights of all the samples that the tree classifies well and increases the weights of all samples that are difficult to classify [16]. It repeats the process until a set number of trees are constructed to obtain a best fit model [17, 18] (Fig. 8.1).

After we trained the first model, the training and testing sets were combined into a single dataset and then randomly divided into a second training and testing set (again 70% vs. 15% split for each cohort). We then trained a second model using the new training set and tested it with the second testing set (Fig. 8.4). We repeated this recombining-dividing-training-testing process, 100, 200, and 500 times to obtain

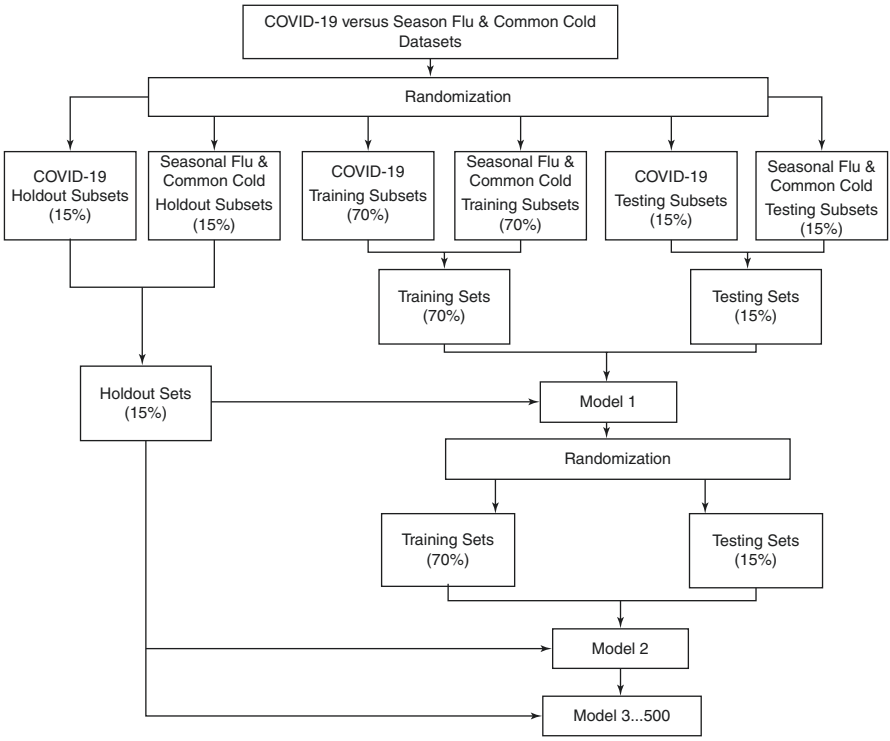


Fig. 8.4 A schematic description of the machine learning process. Patients with COVID-19 and patients with seasonal flu or common cold were randomly divided into three subsets respectively: training subset (70%), testing subset (15%), and holdout subset (15%). The XGBoost trained model was then tested to classify the two cohorts of patients in the testing subset. The training and testing subsets were recombined and then randomly divided into another training and testing subset. This recombining-randomly dividing training-testing process were repeated 500 times to obtain 500 models. The 500 trained and tested models were validated against the holdout subset to ensure that all models' performance was generalizable to classify a new group of patients. This holdout subset was never used in training and testing and therefore pristine

100, 200, and 500 models, respectively (Fig. 8.3). Training multiple models allowed us to prevent the potential undue influence of any particular training-testing dataset partitioning on our assessment of machine learning performance.

After we trained and tested the models, we further validated the models against the holdout set to ensure that all models' performance we observed during training and testing was generalizable to classify a new dataset (Fig. 8.5). Because the holdout set was never used during training and testing and hence pristine, it provided an accurate assessment of our models' generalizability.

To measure model performance, mean squared error (MSE in %) was used to assess a model's classification performance. The percent of variance accounted for MSE (100% - MSE) was reported in our study. We henceforth refer to the percent of variance accounted for as R2. The greater R2 is, the better the model is at classification. We also used the receiver operating curve (ROC) and the area under curve (AUC) to evaluate the model's performance and its sensitivity and specificity at different probability thresholds.

In addition, we obtained a measure of relative importance for each predictor in a XGBoost trained model. Predictor importance reflects how much a predictor variable contributes to the performance of a model. It is based on the frequency of a

Table 8.3 Age, sex, body temperature and self-reported symptoms at admission

	COVID-19 (N = 599)	Seasonal flu or common cold (N = 832)	P value
Age (years)	46.5 ± 15.4	30.1 ± 15.6	<0.001
Sex (male)	287 (47.9%)	376 (45.2%)	0.309
Temperature (°C)	37.1 ± 0.7	38.1 ± 0.9	<0.001
Fever	61.4	91.2	<0.001
Cough	52.6	79.1	<0.001
Sputum	35.2	65.0	<0.001
Fatigue	24.0	47.4	<0.001
Chest pain	19.2	3.1	<0.001
White sputum	18.2	25.7	<0.001
Malaise	14.7	5.9	<0.001
Chill	13.4	77.3	<0.001
Myalgia	9.3	2.9	<0.001
Diarrhea	8.2	4.2	<0.001
Shortness of breath	7.0	2.0	<0.001
Headache	7.0	53.1	<0.001
Nausea	5.5	6.0	<0.001
Sinus congestion	5.1	32.3	<0.001
Yellow sputum	2.5	0	<0.001
Vomiting	2.3	5.5	0.003
Dyspnea	1.8	0	<0.001
Sore throat	1.3	2.8	0.067
Runny nose	0.8	11.7	<0.001

Note: Results are reported in % or otherwise separately indicated. Statistical tests were performed using either t-test or Chi-square analysis whichever was appropriate

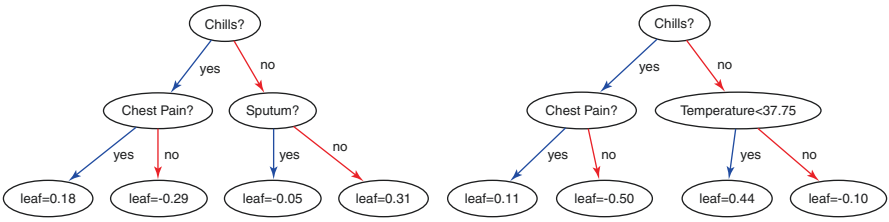


Fig. 8.5 Examples of decision trees in one of the 100 XGBoost models. Each model contains 40 trees. Each tree gives a score that the model uses to predict the overall probability. Each node asks a question regarding a predictor variable. This process eventually leads to a "leaf", or a score based on log odds ratios. The leaf value represents the probability of the data point belonging to the COVID-19 cohort. The leaf value can be either positive or negative, with the value 0 representing 50% probability. The final probability prediction is obtained by taking the sum of all leaf values (raw scores) in all the trees and then transforming it between 0 and 1 using a sigmoid function. A higher sum score means a higher probability of COVID-19 class. The goal was to produce a model with the best parameters to achieve the least mean squared errors. We found the best fitting model of 40 decision trees with a maximal depth of 2 layers

predictor being selected for tree branching, weighted by the squared improvement to the model as a result of each tree branching, and then averaged over all trees.

The demographic characteristics, temperature, and self-reported symptoms at admission from the two cohorts of patients are shown in Table 8.3. The two cohorts shared many similar symptoms, and no symptoms were unique to either patient cohort. However, the frequencies of some of the symptoms were significantly different between the two cohorts, suggesting that using these symptoms, machine learning may indeed be able to classify the two cohorts correctly.

The trained XGBoost models performed exceedingly well in classifying patients with or without COVID-19 during training and testing as showed in Table 8.4. This superior performance was reflected in terms of the near perfect R2 ($p < 0.0001$) and the high AUC ($p < 0.0001$). If AUC = 50%, it suggests a model can only classify two cohorts at the chance level. Models excluding age and sex performed slightly worse than those with the two elements ($p < 0.05$). When age, sex and temperature were removed from the models, the performance decreased significantly in terms of both R2 and AUC ($p < 0.01$).

We found that model performance (R2 and AUC) was highly similar for between all models, be it the 100, 200 or 500 models. Thus, all results reported were based on the 100 models (Table 8.4).

Table 8.5 confirmed that the machine learning models with age, sex, temperature, and self-reported symptoms as predictors generalized very well to the holdout patients, whose data were never seen during training and testing. This excellent performance was supported by high R^2 and AUC.

The models without age and sex performed slightly worse than with such information ($p < 0.05$), suggesting that the demographics may be not critical for classifying the two cohorts of patients. However, the models with age, sex, temperature, and symptoms performed much better than the models with symptoms only, suggesting temperature is an important predictor for classifying whether or not patient have COVID-19 (Fig. 8.6).

Table 8.4 Model performance in the training and testing sets

	R^2 (100 – BCE)	AUC
Training set		
Models with age, sex, temperature & 19 symptoms		
Mean ± SD	95.7 ± 0.2	98.7 ± 0.1
95% confidence interval	95.7–95.7	98.7–98.7
Models with temperature & 19 symptoms		
Mean ± SD	95.2 ± 0.2	98.5 ± 0.1
95% confidence interval	95.1–95.3	98.4–98.5
Models with 19 symptoms		
Mean ± SD	94.2 ± 0.2	97.72 ± 1.1
95% confidence interval	94.2–94.3	97.4–97.8
Testing set		
Models with age, sex, temperature & 19 symptoms		
Mean ± SD	94.4 ± 0.9	97.6 ± 0.8
95% confidence interval	94.3–94.6	97.4–97.8
Models with temperature & 19 symptoms		
Mean ± SD	93.9 ± 1.0	97.3 ± 0.9
95% confidence interval	93.7–94.1	97.1–97.4
Models with 19 symptoms		
Mean ± SD	93.0 ± 1.1	96.5 ± 1.1
95% confidence interval	92.8–93.3	96.3–96.8

[a]Note: MSE: mean squared errors; AUC: area under curve. All measures are in %. All AUC means were significantly greater than the chance level (50%) with $p < 0.0001$

Table 8.5 Model performance in the holdout (validation) set

	R^2 (100-MSE)	AUC
Models with age, sex, temperature and 19 symptoms		
Mean ± SD	94.7 ± 0.3	98.2 ± 0.2
95% confidence interval *Models with temperature & 19 symptoms*	94.9–94.8	95.1–95.4
Mean ± SD	93.9 ± 0.3	97.4 ± 0.2
95% confidence interval: *Models with 19 symptoms*	93.9–94.0	97.4–97.4
Mean ± SD	91.4 ± 0.3	95.40 ± 0.3
95% confidence interval	91.4–91.5	95.4–95.5

[a]Note: MSE: mean squared errors; AUC: area under curve. All measures are in %. All AUC means were significantly greater than the chance level (50%) with $p < 0.0001$

Figure 8.7 shows the relative importance of all predictors in terms of their contributions to the performance of the 100 models. In addition to temperature, several symptoms including chills, sputum, headache, sinus congestion, and chest pain were among the most important predictors in classifying the two patient cohorts (Fig. 8.7). However, simply having one or more of these top ranked symptoms did

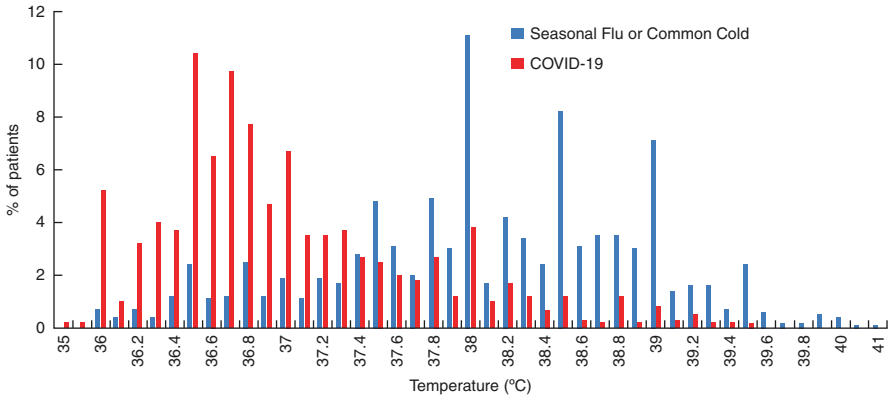


Fig. 8.6 Body temperature distributions of patients with COVID-19 versus seasonal flu or common cold

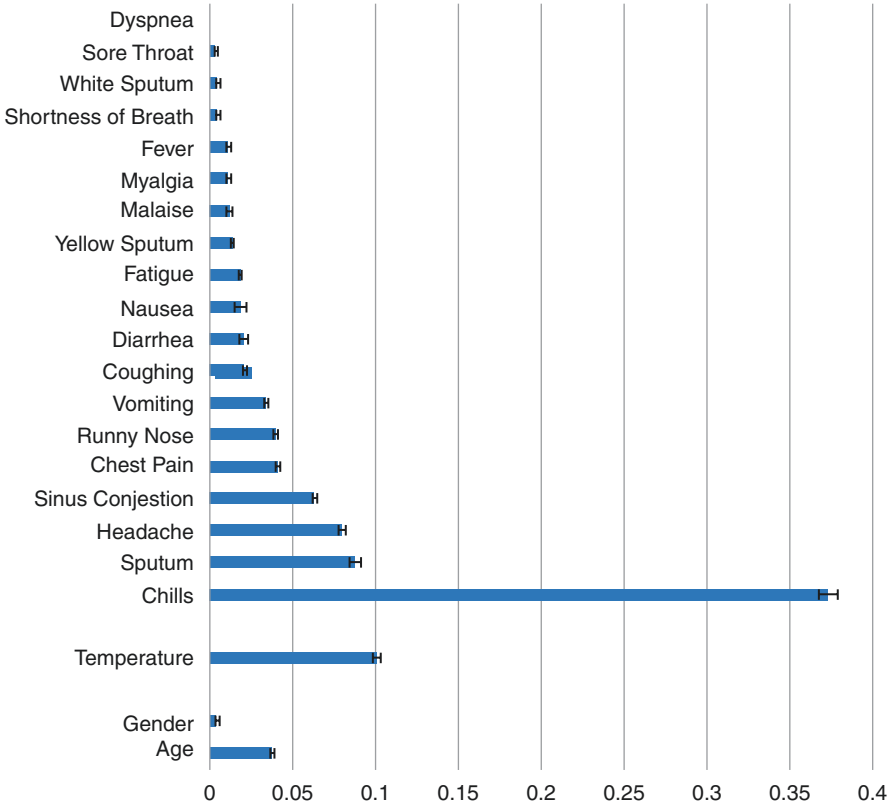


Fig. 8.7 The mean relative importance measures (90% confidence intervals) of all predictors in the 100 models using age, sex, temperature and 19 symptoms as predictors. The greater the measure, the greater the contribution of a predictor to the classification of COVID-19 versus seasonal flu or common cold patients

not necessarily lead to an accurate classification of patients. For example, among the patients, 30% had both chills and sputum, 11.8% had chills, sputum, and headache, 4.6% had chills, sputum, headache, and sinus congestion, and 2.1% had all top five symptoms. A reliable and accurate performance to classify the two cohorts of patients was achieved by considering the unique patterns of the presence and absence of all symptoms along with age, sex, and body temperature.

We created a website that implemented the models for easy access by physicians and patients online (https://survey-demo.deepaffex.ai/). The website included a form that physicians or patients can fill out, which asks for information about the patient's demographics and symptoms. After the form is submitted, the model computes the percent likelihood of the patient having COVID-19.

Summary, Limitations, Challenges, and Future Applications

The present chapter showcases three studies that use the machine learning approach along with patient data to predict COVID-19 patient mortality, determine the presence of COVID-19, and differentiate between patients with COVID-19 and patients with flu and common cold. All three studies produced computational models that are highly accurate, with high sensitivity and specificity. These findings suggest that the machine learning approach is a viable methodology to address present and future public health issues. In addition, our hypothesis that the COVID-19 cohort presents a unique non-linear combination of not only patient demographics, medical history but also presenting symptoms was confirmed, validating their importance for making appropriate medical decisions.

Further, these studies provide useful insights about COVID-19. For example, Zoabi et al. [11] found that cough and fever are the top predictors of COVID-19 infection. Our study showed that a patient with low but not high fever was significantly more likely to be infected with COVID-19. Quiroz-Juárez et al. [10] revealed that patient age, days with symptoms and treatment, and whether patients present symptoms of pneumonia are highly important in presenting mortality at all stages of disease progression. Finally, all three studies indicate that the accuracy of classification is not predicated around the absence or presence of predictors. Rather, it is the combinatory patterns of their presence and absence that allow computational models to accurately classify different cohorts of patients. These insights are highly useful in assisting physicians with making patient specific medical decisions, and also provide insight for public health officials to devise public health measures. For example, current screening practices at various public locations (e.g., airports, shopping malls) involve the use of contactless thermal sensors to detect individuals with high temperature as an indication of COVID-19 infection. Our work revealed that COVID-19 patients who have a fever tend to have lower body temperatures than those with the flu or common cold. This

finding suggests that these modern COVID-19 detection practices are not only insufficient but possibly misleading.

The success of the machine learning approach showcased here can be attributed to a number of factors. First, the success of the three studies benefited from the availability of the large-scale datasets that are publicly accessible. It illustrates the importance of the open science movement that is underway in many scientific disciplines, including medicine and public health. Second, the advance of AI in general and machine learning techniques in the last two decades has made it possible for researchers to train, test, and validate accurate and robust computational models quickly and efficiently. This has been made possible by increased processing capacities and decreased computational costs due to Moore's Law as well as the existence of publicly shared and free programming codes (e.g., those on GitHub). Third, the three studies not only produced accurate computational models but also use such methods as the python based SHAP to identify the key predictors that explain why the computational models are accurate. This information is highly crucial for medical decision making and public health policy implementation because physicians, public health officials, and the general public not only wish to know a model to be accurate but also what factors contribute to its accuracy.

The advantage of the machine learning approach lies not only in its ability to produce highly accurate models, but also its ease to be implemented in an online platform to be used by everyone. For example, Quiroz-Juárez et al. [10] implemented their model on a website such that physicians can implement the required patient data to predict the likelihood of mortality of the patients under their care so as to make appropriate medical decisions. Our models were also implemented on a website that any individual in the world with internet access to self-assess the likelihood of COVID-19. Such self-assessment can be performed by people without having to leave their homes for a swab test. Although our machine learning models cannot be used to make or rule out the diagnosis of COVID-19, they can be used to identify suspected cases in need of medical attention [19]. Further, as the training dataset is expanded with more patients, the accuracy of model prediction will only continue to improve, which may make it possible to use such models as a reliable and inexpensive COVID-19 screen test.

Several limitations of the current machine learning approach must be acknowledged. First, the patient data used in the three studies, though large, were collected at the early stages of the pandemic. Since then, COVID-19 virus has evolved, which have led to different presenting symptoms depending on the nature of the variant. To ensure their accuracy at the present, the models need to be retrained with updated datasets. Second, although all models' performance generalized well to the pristine holdout sample, the studies are in general retrospective in nature. Prospective clinical studies are needed to validate the models' real-world utility. Third, all models were developed with data from COVID-19 patients who presented some symptoms. They are not suitable to identify asymptomatic COVID-19 patients [20–22], a challenge that calls for a specifically designed study with possibly different machine

learning approaches. Fourth, although the models used a sufficiently large sample to perform machine learning, the patients were all geographically and racially homogeneous, the generalizability of these models to classify patients of different races and from different geographic locations is yet to be tested. Nevertheless, these machine learning models can easily be expanded with global data for future international application.

The chapter showcases three studies that confirmed the hypothesis that patients with COVID-19 present a unique combinatory pattern of symptoms along with personal information, which in turn can be used to predict patient mortality, detect the presence of COVID-19, and differentiate between patients with COVID-19 and patients with common cold or flu. They demonstrate that the machine learning approach is a powerful tool to solve specific problems which arose during the COVID-19 pandemic. The lessons learned from these studies can be applied to not only the ongoing pandemic, but also future public health crises. In the future, through international collaboration, large and diverse patient datasets should be quickly be made available. International teams should collaborate to develop machine learning models to address pressing questions concerning a similar public health crisis in a timely and convenient fashion. Such models should then be implemented on websites and mobile devices for wide use so as to assist medical and public health decision making, which in turn should lead to strategical allocation of scarce healthcare resources to where they are truly needed.

References

1. Gao Y, Yan L, Huang Y, et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science (New York, NY). 2020. https://doi.org/10.1126/science.abb7498.
2. Guan WJ, Ni ZY, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med. 2020;382(18):1708–20. https://doi.org/10.1056/NEJMoa2002032.
3. Grasselli G, Zangrillo A, Zanella A, et al. Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy Region, Italy. JAMA. 2020. https://doi.org/10.1001/jama.2020.5394.
4. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City Area. JAMA. 2020. https://doi.org/10.1001/jama.2020.6775.
5. Pung R, Chiew CJ, Young BE, et al. Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures. Lancet. 2020;395(10229):1039–46. https://doi.org/10.1016/S0140-6736(20)30528-6.
6. Li LQ, Huang T, Wang YQ, et al. COVID-19 patients' clinical characteristics, discharge rate, and fatality rate of meta-analysis. J Med Virol. 2020. https://doi.org/10.1002/jmv.25757.
7. CDC. Cold versus flu. National center for immunization and respiratory diseases (NCIRD). 2019. https://www.cdc.gov/flu/symptoms/coldflu.htm.
8. Eccles R. Understanding the symptoms of the common cold and influenza. Lancet Infect Dis. 2005;5(11):718–25. https://doi.org/10.1016/S1473-3099(05)70270-X.
9. Sun Y H, Luo H, Lee K. A novel approach for developing efficient and convenient short assessments to approximate a long assessment. Behav Res Methods. 2022. https://doi.org/10.3758/s13428-021-01771-7.

10. Quiroz-Juárez MA, Torres-Gómez A, Hoyo-Ulloa I, León-Montiel RJ, U'Ren AB. Identification of high-risk COVID-19 patients using machine learning. PLoS One. 2021;16(9):e0257234. https://doi.org/10.1371/journal.pone.0257234.
11. Zoabi Y, Deri-Rozov S, Shomron N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. NPJ Digit Med. 2021;4:3. https://doi.org/10.1038/s41746-020-00372-6.
12. Shapley LS. A value for n-person games. Ann Math Stud. 1953;28:307–17.
13. Ghorbani A., Zou J. Data shapley: equitable valuation of data for machine learning. arXiv:1904.02868v2. 2019.
14. Smith M, Alvarez F. Identifying mortality factors from machine learning using shapley values - a case of COVID19. Expert Syst Appl. 2021;176:114832. https://doi.org/10.1016/j.eswa.2021.114832.
15. Chen Y, Wang X, Jung Y, et al. Classification of short single-lead electrocardiograms (ECGs) for atrial fibrillation detection using piecewise linear spline and XGBoost. Physiol Meas. 2018;39(10):104006. https://doi.org/10.1088/1361-6579/aadf0f.
16. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. J Anim Ecol. 2008;77(4):802–13. https://doi.org/10.1111/j.1365-2656.2008.01390.x.
17. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001;29(5):1189–232. https://doi.org/10.1214/aos/1013203451.
18. Torlay L, Perrone-Bertolotti M, Thomas E, et al. Machine learning-XGBoost analysis of language networks to classify patients with epilepsy. Brain Inform. 2017;4(3):159–69. https://doi.org/10.1007/s40708-017-0065-7.
19. Alimadadi A, Aryal S, Manandhar I, et al. Artificial intelligence and machine learning to fight COVID-19. Physiol Genomics. 2020;52(4):200–2. https://doi.org/10.1152/physiolgenomics.00029.2020.
20. Arons MM, Hatfield KM, Reddy SC, et al. Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. N Engl J Med. 2020. https://doi.org/10.1056/NEJMoa2008457.
21. Kimball A, Hatfield KM, Arons M, et al. Asymptomatic and presymptomatic SARS-CoV-2 infections in residents of a long-term care skilled nursing facility - King County, Washington, March 2020. MMWR Morb Mortal Wkly Rep. 2020;69(13):377–81. https://doi.org/10.15585/mmwr.mm6913e1.
22. Qian G, Yang N, Ma AHY, et al. A COVID-19 Transmission within a family cluster by presymptomatic infectors in China. Clin Infect Dis. 2020. https://doi.org/10.1093/cid/ciaa316.

Chapter 9
AI Techniques for Forecasting Epidemic Dynamics: Theory and Practice

Aniruddha Adiga, Bryan Lewis, Simon Levin, Madhav V. Marathe, H. Vincent Poor, S. S. Ravi, Daniel J. Rosenkrantz, Richard E. Stearns, Srinivasan Venkatramanan, Anil Vullikanti, and Lijing Wang

Introduction

COVID-19 is the largest pandemic the world has seen since the 1918 Influenza pandemic. As of February 2022, worldwide, we had approximately 433 M confirmed cases, 33 M active cases, and 5.95 M confirmed deaths; see https://nssac.bii.virginia.edu/covid-19/dashboard/. The pandemic and its unprecedented social,

A. Adiga · B. Lewis · S. Venkatramanan
Biocomplexity Institute and Initiative, University of Virginia, Charlottesville, VA, USA
e-mail: aniruddha@virginia.edu; brylew@virginia.edu; srini@virginia.edu

S. Levin
Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA
e-mail: slevin@princeton.edu

M. V. Marathe · A. Vullikanti
Biocomplexity Institute and Initiative, University of Virginia, Charlottesville, VA, USA

Department of Computer Science, University of Virginia, Charlottesville, VA, USA
e-mail: marathe@virginia.edu; vsakumar@virginia.edu

H. V. Poor (✉)
Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ, USA
e-mail: poor@princeton.edu

S. S. Ravi · D. J. Rosenkrantz · R. E. Stearns
Biocomplexity Institute and Initiative, University of Virginia, Charlottesville, VA, USA

Department of Computer Science, University at Albany—State University of New York, Albany, NY, USA
e-mail: ssravi@virginia.edu, ssr6nh@virginia.edu; dr3xx@virginia.edu; rs7vs@virginia.edu

L. Wang
Boston Children's Hospital and Harvard Medical School, Boston, MA, USA
e-mail: lw8bn@virginia.edu

N. Lidströmer, Y. C. Eldar (eds.), *Artificial Intelligence in Covid-19*,
https://doi.org/10.1007/978-3-031-08506-2_9

economic, and political impact have been discussed in many publications (e.g., [1–7]). Computational and mathematical models have been used extensively during the COVID-19 pandemic to support policy makers and citizens' response efforts (e.g., [8–15]). An essential use of such models is in forecasting infectious disease dynamics. Over the past several years, there has been an increasing interest in developing systems that provide early warnings regarding epidemics. This is borne out by a large number of epidemic forecasting challenges such as the FluSight by the Centers for Disease Control and Prevention (CDC) [16–19], "CHIKV Challenge" by DARPA [20], and the IARPA Flu challenge [21] that rely on collaborative efforts to forecast various disease parameters. COVID-19 strengthened these initiatives; in April 2020, CDC, in collaboration with academic partners, initiated the COVID-19 Forecast Hub (henceforth referred to as *The Hub*), a consortium of modeling teams to coordinate the forecasting efforts (viz.covid19forecasthub.org). In this chapter, we summarize various mathematical and computational approaches to forecasting infectious disease dynamics, and pay particular attention to forecasting dynamics of infectious diseases such as COVID-19 and seasonal influenza-like illness (ILI). In section "Discussion", we briefly discuss the applicability of these methods in studying other infectious diseases and questions related to forecasting viral dynamics.

Challenges Several challenges present themselves when forecasting infectious disease dynamics. We now highlight some of the critical challenges.

Data Availability of timely, well-informed, and fine-grained data is crucial for generating reliable epidemic predictions. Challenges in this context include: (*i*) frequency of updates (timing and quality); (*ii*) access to surrogate data such as real-time mobility data, digital health device data, social media feeds, weather data, etc., that is often useful in improving forecast quality; (*iii*) spatial resolution of the measurements; (*iv*) noise and other inaccuracies in the reported data and (*v*) heterogeneity in reporting structures of data across states and local health districts; see [8, 19, 22, 23] for further discussion on this topic.

Targets A target corresponds to the particular epidemic signal of interest. Previous efforts like FluSight[1] focused on out-patient visits for ILI and have included long-term targets such as season onset, peak week, peak intensity and short-term targets of 1–4 week ahead forecasts. In COVID-19 forecasting efforts, the scientific community has considered a number of different targets, including the following: multiple disease indicators such as cumulative or daily/weekly confirmed case counts, deaths, and hospitalization counts. Death counts, although reliable, have proved problematic in policy planning as they are lagging indicators of the disease prevalence. Case counts are typically near-real-time indicators of the state of the pandemic but are severely limited by the testing capacity. Other targets include: peak week, cumulative incidence during the peak week. However, they are dependent on multiple factors such as the epidemic-specific pathogen variant, human behavior,

[1] https://www.cdc.gov/flu/weekly/flusight/index.html.

non-pharmaceutical interventions (NPIs), etc., and have been hard to forecast. See [24, 25] for additional examples of targets studied in the literature.

Physical, Spatial, Temporal, Socio-economic, and Political Dependencies Epidemic dynamics in a given region co-evolve with other neighboring regions. They also depend on a number of exogeneous factors such as weather, socio-economic and political dynamics. For instance, epidemic dynamics in a county depend on the dynamics in neighboring counties and local, state and federal policies, socio-economic considerations such as compliance, and political leanings. A natural method to capture these inter-dependencies is to use network models, but obtaining the necessary data to synthesize the networks and drive the dynamics remains a challenge. Calibrating these coupled models is another challenge. For example, considering the United States network with 3000+ counties and W weeks of data, there are technically over $3000 \times W$ entries in the spatio-temporal transmissibility matrix to be calibrated, making traditional Bayesian techniques computationally intensive and susceptible to overfitting due to the limited training data size. Lack of data and improper use of machine learning methods can also lead to overfitting and thus poor generalizability.

Model Evaluation Evaluation of forecasting performance is crucial for model improvement. Traditionally, epidemic forecasting involves *point* and *probabilistic* forecasts. However, there are multiple metrics available and the performance of a model can vary based on the choice of a metric employed for evaluation. Hence choosing the appropriate metrics for evaluating model performance for forecasting a specific target is crucial. In section "Metrics for Forecast Evaluations", we discuss a variety of metrics that have been employed in the context of epidemiological signal forecasting.

Limits to Forecasting In various domains, including epidemics, ecology, weather, and evolutionary biology, limits to forecasting have been discussed. For example, Drake [26, 27] provides several important reasons (such as the nature of stochastic disease propagation models and high sensitivity of disease parameters to changes in the environment) to observe that while systems may provide good forecasts of some epidemic measures (e.g., timing), there are limits on the effectiveness of such systems in forecasting other measures (e.g., final epidemic size). May [28] emphasizes the need for caution in concluding the dynamics of ecological systems when there is uncertainty in the structure and parameter values of the underlying network model. A similar observation was made by Jacob [29] in evolutionary biology. In section "Theoretical Foundations for Forecasting in Network Models", we will briefly discuss limitations on the efficient computations of epidemic forecasts from the perspective of computational complexity; for a more comprehensive discussion on this topic, we refer the reader to [30].

Additional Challenges While Forecasting During an Ongoing Pandemic We make four additional points as they pertain to forecasting during an ongoing pandemic (cf. Fig. 9.1) (*i*) the causal reasons for each wave were different; (*ii*) the state of the underlying social systems was different; (*iii*) different individuals were

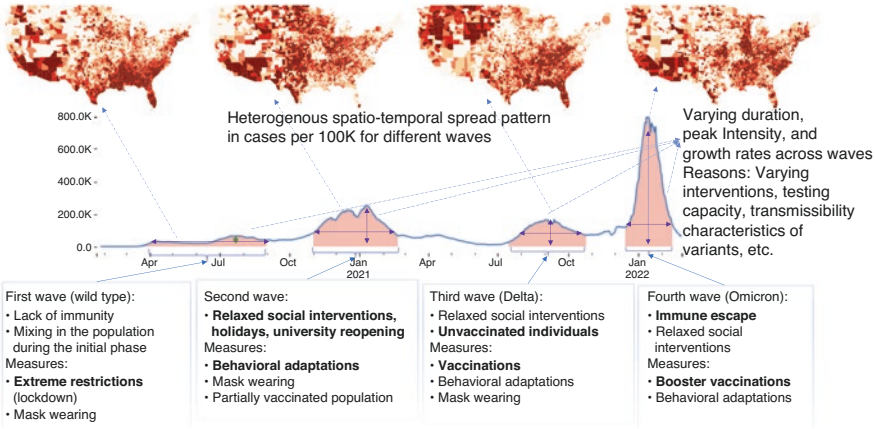


Fig. 9.1 Multiple *waves* of the COVID-19 pandemic in the US and their characteristics. Similar waves were observed in other countries. The waves observed in all the states/counties in the US were similar, although the intensity and timing differed quite a bit as shown in the choropleth maps (where darker shades indicate higher average number of cases per 100 K reported during the corresponding wave). The underlying causal reasons for each wave were different, making forecasting extremely challenging. (The concept of a *wave* is subjective. We considered periods of substantially high COVID-19 activity as a *wave*.) In the choropleth maps, the scales are different across the maps. The data used to generate each map was analyzed independently and divided into seven equal-count intervals (quantiles)

affected physically, socially and economically, furthermore different social, behavioral and political considerations were at play during each phase, (*iv*) the shape of the curves, the intensity, peak and the growth was different for each phase and (*v*) the advent of social media, international trade and travel and prevailing political polarization introduced new challenges in the form of misinformation, breakdown of global supply chains, reduced compliance and higher levels of hesitancy than one would have anticipated. These factors along with the challenges stated earlier made forecasting COVID-19 pandemic challenging—it is thus not surprising that methods have often struggled to produce accurate and timely forecasts.

Organization This chapter is organized as follows. In section "A Review of Model Types and Limits to Forecasting", we briefly review several models used in the literature for forecasting. In section "Preliminaries" we mention some fundamental problems in forecasting under several models. Section "AI-driven Engineering", presents an ensemble model and a hybrid forecasting model as examples of state-of-the-art forecasting systems. Section "Theoretical Foundations for Forecasting in Network Models", discusses the computational intractability of short-term forecasting for network-based epidemic models. Section "Discussion" concludes the chapter with a discussion on some general issues and possible research directions in infectious disease forecasting.

A Review of Model Types and Limits to Forecasting

As remarked earlier, forecasting epidemic dynamics has become a very active area of research over the last two decades. The class of models used varies depending on the application, data availability, computational requirements and the choice of theoretical or data driven constructs. We briefly summarize the state-of-the art below; see [31] for additional discussion and references.

Statistical Methods This class of forecasting methods employs statistical models on time-series data, determines model parameters from observed data and leverages the model to make future predictions. These methods assume that the observed data is the outcome of a parametric random process whose parameters are unknown. A popular class of models are the autoregressive models (such as AR, ARMA, and ARIMA) which assume that the observed time series at the current time step can be expressed as a linear combination of past observations. In the context of ILI forecasting, in addition to the ARIMA terms of the ILI time series, other exogenous variables such as search trends, social media data, weather data, etc. have been employed as exogenous regressors to enhance nowcast performance. Yang et al. [32], Rangarajan et al. [33], Kandula et al. [34], Soebiyanto et al. [35], and Paul et al. [36] assume a Gaussian distribution on the data when modeling ILI rates and activity level. AR models for count data are modeled using Poisson [37] and negative binomial distributions [38, 39], and these result in a class of generalized linear models. Owing to a large number of explanatory variables, techniques such as LASSO [40], log-likelihood ratio test [33], and block coordinate descent methods [41] are employed to select a sparse subset of most relevant variables. In the presence of sufficient seasonal data, a Bayesian weighted average of trajectories from past seasons to model the current season as a mixture of Gaussian models is shown to perform reasonably well in the case of influenza in [42, 43] and dengue in [44]. Though technically not a statistical method, method of analogues is a non-parametric model which attempts to find the most relevant historical segments of data or nearest neighbors with respect to the observed data and uses a weighted average of the nearest neighbors to produce forecasts [43]. Exponential smoothing is another class of non-parametric regression models which employ exponentially decaying weights on historical samples (e.g., Petropoulos et al. [45]). Given that the observed data is governed by several underlying parameters, a natural attempt is to consider complex models such as compartmental models that encapsulate the interaction between these parameters. These underlying parameters, such as number of actual infections, susceptibles, transmissibility, etc., are typically unobserved states in the model and their estimation has been addressed by a popular class of statistical techniques called *Filtering methods*. The objective of filtering methods is to estimate the probability distribution of the model's states given the observations. Ensemble Kalman filter [46], Ensemble adjusted Kalman filter [47], particle filtering [48] and their variants are some of the popular algorithms used for recursive estimation of model parameters. These methods have been extensively explored in modeling influenza epidemics, and a detailed comparison of the filtering techniques is pro-

vided in [49]. The filter-based modeling framework, in conjunction with ILI metrics, testing data, and web-based ILI estimates, has been used for both retrospective analysis and real-time forecasting [50–52]. The filter-based methods rely on accurate modeling of the relationships between underlying parameters to be effective. Since obtaining accurate models of such relationships has been particularly challenging for the novel SARS-CoV-2 virus, the use cases for filtering methods in that context have been limited.

Deep Learning Unlike statistical methods, deep learning (DL) does not make any assumption on the underlying data (statistical or domain knowledge) but typically attempts to find a non-linear mapping between the input and output data. In the presence of sufficiently large datasets and model parameters, DL strategies have the ability to approximate complex mapping functions [53] and are hence gaining increasing prominence in epidemic forecasting. The mapping is realized through a deep neural network (DNN) whose structure varies depending on the data being analyzed. Some common structures used in the literature are feedforward neural networks (FNNs), recurrent neural networks (RNNs), convolutional neural networks (CNNs), and graph neural networks (GNNs). An FNN refers to a traditional artificial neural network wherein the flow of information is unidirectional, and has been used for forecasting prevalence (e.g., for dengue [54, 55] and influenza [56]). In contrast to FNN, the RNNs and their variants (e.g., gated recurrent unit (GRU) [57] and long-short term memory (LSTM) [58]) use feedback connections and are designed to process sequential data. LSTM networks have been used in influenza-like illness forecasting along with auxiliary data streams such as Twitter data [59], climatic factors, and geographical proximity [60]. Attention-based LSTM models are employed in [61] and [62]. Initially, FNNs were attempted for forecasting COVID-19 deaths counts [63], but subsequent efforts [22, 64] have employed LSTM networks to predict COVID-19 transmission. Other networks like CNNs are generally used for capturing spatial patterns in image data through shift-invariant learnable filters. A combination of CNN and RNN called the CNNRNN-Res [65] model, and Deepcovidnet [66] have been used in processing multi-variate epidemic time series corresponding to various geographical regions as images. GNNs are a generalization of CNNs designed to process graph data and capture spatial patterns. The basic idea is to generate node embeddings based on local network neighborhoods through message passing. Cola-GNN is a cross-location attention-based graph neural network for forecasting ILI [67]. During COVID-19, inter-region connections have greatly varied, and [68] and [69] have accounted for it by considering mobility-informed adjacency of nodes for COVID-19 daily case prediction.

Mechanistic Models These are also referred to as causal models since they attempt to incorporate the causal relationships between inputs and outputs. They have been the workhorses of the infectious disease modeling community, and they come in a variety of flavors. Broadly, they can be classified into ordinary-differential-equation-based compartmental models and agent-based models (ABM). The compartmental models typically assume homogeneous mixing within a population and are

used to study the large-scale dynamics of a disease in the population. ABMs are detailed models that represent individuals in a population as agents with heterogeneous properties, and are typically useful in studying the effects of targeted interventions in a population. The literature on mechanistic models is vast; for a more detailed description, we refer the reader to [8, 70] and references cited therein.

Hybrid Approaches and Human Judgment in Forecasting An emerging trend in epidemic forecasting is combining theory-based mechanistic models and deep learning models. Such methods attempt to reap the benefits of both models and obtain better forecasting performance. In the context of influenza forecasting, Zhao et al. [71] and Hua et al. [72] employed social media mining techniques to enhance theory-based mechanistic models, while Wang et al. [73] trained an LSTM-based model with theory-generated training data for forecasting at a high geographical resolution. Recently for COVID-19 forecasting, Dandekar and Barbastathis [74] calibrated causal model parameters by combining first-principles epidemiological equations and a data-driven neural network model. Gao et al. [75] attempt to capture the spatio-temporal dynamics using a GNN and employ a transmission dynamics loss term to regularize model predictions in GNN training. Despite the use of complex models and auxiliary data, models have failed on several occasions to forecast accurately [76]. Human judgment can often be leveraged to make reliable forecasts and reduce outliers in forecasts. Multiple human-judgment-based models have been used in the context of Influenza, and involves either crowd-based voting of various model forecasts [77, 78] or predictions provided by humans [79, 80] through user-friendly web-based applications. In all these experiments, it was observed that human-judgment-based forecasts were more accurate than individual model forecasts.

Forecasting Dynamics of Complex Systems Many researchers have studied various aspects of reliably forecasting the dynamics of complex systems. We first briefly discuss their observations regarding complex systems in general and then outline some challenges that have been identified in the context of epidemic dynamics. Cheng et al. [81] address the question of predicting whether a cascade will continue to grow in a social network. They remark that "a robust way to formulate the problem of cascade prediction remains an open problem." Martin et al. [82] observe that even a small degree of uncertainty can limit the predictability of complex social systems. Hofman et al. [83] mention that "theoretical limits to the predictive accuracy of complex systems must be better characterized." Using Google Flu Trends as an example, Lazer et al. [84] caution against the use of social media data and search information as a substitute for traditional data collection and analysis methods to predict epidemic measures. They highlight the importance of systematically "studying the evolution of the socio-technical systems that are embedded in our society." Shaman et al. [51] use new data analysis techniques to predict the timing of the peak number of infections for influenza outbreaks and suggest that their methods can be made more robust when additional data are available. Using results from CDC's "Predict the 2013–2014 Influenza Season Challenge", Biggerstaff et al. [16] con-

clude that "further efforts are needed to improve the accuracy so that policy makers can reliably use those predictions." Pinto et al. [85] study the problem of locating the sources of infection in networks. Karrer and Newman [86] use a message-passing model to compute epidemic probabilities in networks. Other researchers have used approaches based on message-passing and belief propagation to obtain practical solutions to the problem of identifying the source of infection in general and special classes of networks [87, 88]. Althouse et al. [89] observe that superspreader events have played a significant role in the spread of COVID-19, and that prediction of such events is important in controlling disease outbreaks.

Predictability In terms of various parameters of an outbreak (incident cases, importation probability, growth rate, etc.), several models have been shown to provide reasonable forecasts, thus demonstrating the feasibility of predicting the data [17]. Using permutation entropy on the signal-of-interest as a model independent measure of predictability, it is shown in [90], that across different diseases, although a fundamental entropy barrier exists, it is often well beyond the time scale of a single outbreak, implying short-term predictions are likely to succeed.

Preliminaries

Forecasting Problem Consider the epidemiological time series $\{y_t\}_{\{t=1\}}^{T}$, where y_t can correspond to case counts, number of hospitalizations, or individual disease states. The forecasting problem in this work involves predicting the h-step ahead point forecasts $y_{\{T+h\}}$ or the forecast probabilistic distribution $P\left(y_{\{T+h\}} \mid \{y_t\}_{t=1}^{T}\right)$. These forecast values can be realized through the model f_Θ where Θ denotes the model parameters that can correspond to either the coefficients of the AR terms, weights in a DNN, or the transition rates in an ordinary-differential equation based compartmental model.

Model Details

Compartmental Model These include the SIR class of models and their variants. The simplest form of these models [91–93] assume that the population is completely mixed, and compartmentalize (i.e., partition) a population of size N into three sets, each corresponding to a disease state, which is one of susceptible (S), infective (I) and removed or recovered (R). (Some complex models also include additional states such as Exposed (E), Vaccinated (V), Hospitalized (H), Death (D), etc.) The model then specifies the transition (governed by the transition rates) from susceptible individuals to infectious, and then to recovered. Let $S(t)$, $I(t)$ and $R(t)$

denote the number of people who are in susceptible, infected, and recovered states at time t. Let $s(t) = S(t)/N$, $i(t) = I(t)/N$ and $r(t) = R(t)/N$; then, $s(t) + i(t) + r(t) = 1$. Then, the SIR model can be described by the following system of ordinary differential equations

$$\frac{ds}{dt} = -\beta si, \qquad \frac{di}{dt} = \beta si - \gamma i, \qquad \frac{dr}{dt} = \gamma i, \tag{9.1}$$

where β is referred to as the **transmission rate** and γ as the **recovery rate**. A key parameter in such a model is the **reproductive number** R_0, which is defined by $R_0 = \beta/\gamma$. At the start of an epidemic, much of the public health effort focuses on estimating R_0 from observed infections [94]. Practical implementations approximate Eq. (9.1) using difference equations where t is discretized into days (or weeks) based on the speed of evolution of the epidemic. In addition, owing to multiple reasons such as varying pathogen strains, social distancing norms, etc. the transition rates can vary across time. The difference equations with Δ representing the first-order difference can be expressed as follows:

$$\Delta s_{t+1} = -\beta_t s_t i_t, \quad \Delta i_{t+1} = \beta_t s_t i_t - \gamma_t i_t, \quad \Delta r_{t+1} = \gamma_t i_t. \tag{9.2}$$

Note the parameter t included in the transition rates. More complex models such as the structured metapopulation models further disaggregate the population to account for their inherent heterogeneity [8].

SIR Epidemic Model for Networks For simplicity, we discuss an SIR model on a network with unit infectious duration. The contagion is assumed to spread on an undirected network $G(V, E)$, where V and E represent the set of nodes and edges, respectively. At any time[2] instant, each node $v \in V$ is in one of the states from the domain $\mathcal{D} = \{S, I, R\}$, where, S, I and R represent susceptible, infected (or infectious) and recovered states, respectively. A node stays in state I for exactly one time step. For any node $v \in V$, the **neighborhood** of v, denoted by N_v, contains each node u such that the edge $\{u, v\}$ is in E. Each edge $e = \{u, v\} \in E$ is associated with a **transmission probability**[3] p_e with the following interpretation. At time t, suppose node v is in state S and $X(v, t)$ denotes the subset of N_v such that each node in $X(v, t)$ is in state I. Then, at time t, each node $w \in X(v, t)$ tries to infect node v with probability $p_{\{v,w\}}$ *independently* of the other nodes in $X(v, t)$. If any of these attempts are successful, then the state of v changes to I at time $t + 1$.

A **configuration** of an SIR process on a network at time t is an n-vector $\left(b_1^t, b_2^t, \ldots, b_n^t\right)$, where $b_i^t \in \mathcal{D}$ is the state of node v_i at time t, $1 \le i \le n = |V|$. A single transition of an SIR system from one configuration to another is obtained by

[2] The unit of time (which may be a day, a week, etc.) depends on the epidemic that is being modeled.

[3] This is analogous to the transmission rate β introduced under the compartmental model above.

updating the state of each node v *synchronously* in the following manner, depending on the state of v at time t.

1. The state of v at time t is R: In this case, the state of v at time $t + 1$ is also R. (Thus, once a node reaches the state R, it remains in that state forever.)
2. The state of v at time t is I: In this case, the state of v at time $t + 1$ is R. This is because of the assumption that each node remains in state I for exactly one[4] time unit.
3. The state of v at time t is S: In this case, the state of v at time $t + 1$ is determined by the following stochastic process. Recall that $X(v,t) \subseteq N_v$ denote the set of neighbors of v whose state is I at time t. Let $\pi(v,t)$ be defined as follows.

$$\begin{aligned} \pi\left(v,t\right) &= 0 && \text{if } X\left(v,t\right) = \varnothing \\ &= 1 - \prod_{u \in X\left(v,t\right)} \left(1 - p_{\{u,v\}}\right) && \text{otherwise.} \end{aligned}$$

The above expression for $\pi(v,t)$ can be seen to be a simple consequence of the assumption that each node in $X(v,t)$ tries to infect v independently. The state of v at time $t + 1$ is I with probability $\pi(v,t)$ and S with probability $1 - \pi(v,t)$.

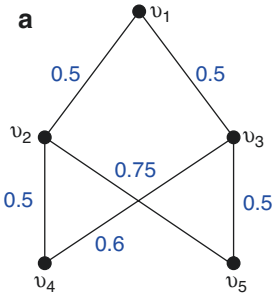
Initially (i.e., at $t = 0$), one or more nodes are in state I and the other nodes are in state S. Starting from the given initial configuration $\mathcal{C}_0$, the system goes through a sequence of configurations over time. We now present an example below to illustrate this process.

Example 1 The graph of an SIR system consisting of five nodes is shown in Fig. 9.2. The nodes are labeled v_1 through v_5 and the transmission probability of each edge is shown in blue. Suppose at $t = 0$, node v_1 is in state I and all other nodes are in state S. Starting from this initial configuration, one possible sequence of configurations that the system may go through is shown in the table in the right panel of Fig. 9.2. The stochastic process through which this sequence of configurations is generated is explained below.

Time $t = 1$*:* Since node v_1 is in state I at $t = 0$, each of the nodes v_2 and v_3 can get infected at $t = 1$ with a probability of 0.5. Assume that the SIR process causes both nodes v_2 and v_3 to be infected by v_1 at $t = 1$. Thus, at $t = 1$, node v_0 changes to state R, nodes v_2 and v_3 change to state I, and nodes v_4 and v_5 remain in state S.

Time $t = 2$*:* We can compute the probabilities of nodes v_4 and v_5 getting infected at $t = 2$ as follows. First, consider node v_4. The neighborhood of v_4 is $\{v_2, v_3\}$ and both of these nodes are in state I at $t = 1$. Since the transmission probabilities of edges $\{v_2, v_4\}$ and $\{v_3, v_4\}$ are 0.5 and 0.6 respectively, the probability of v_4 getting infected is given by $1 - [(1 - 0.5) \times (1 - 0.6)] = 0.8$. Similarly, it can be seen that the probability of node v_5 getting infected is 0.875. For this example, let us assume that the stochastic nature of the process causes v_4 to get infected, but v_5 does *not* get

[4] With minor modifications, our theoretical results discussed in section "Theoretical Foundations for Forecasting in Network Models" can be shown to hold even when the infectious period for each node is any constant number of time units.



b

Time	Configuration
0	(*I, S, S, S, S*)
1	(*R, I, I, S, S*)
2	(*R, R, R, I, S*)
3	(*R, R, R, R, S*)

A sequence of configurations of the SIR system

Fig. 9.2 The underlying graph of an SIR system and one possible sequence of its configurations, leading to a fixed point at time $t = 3$. The transmission probability of each edge is shown in blue. Each configuration is a 5-tuple $(s_1, s_2, s_3, s_4, s_5)$, where $s_i \in \{S, I, R\}$ is the state of node v_i, $1 \leq i \leq 5$. (**a**) A networked SIR system. (**b**) A sequence of configurations of the SIR system

infected. Thus, at $t = 2$, nodes v_2 and v_3 change to state R, node v_4 changes to state I while v_5 remains in state S.

Time $t = 3$*:* Here, v_4 is in state I. However, none of its neighbors is in state S. Thus, v_4 cannot cause any new infections. In other words, v_4 changes to state R while the other nodes remain in their previous states. Since this configuration at time $t = 3$ does not have any node in state I, no further state changes can occur. Thus, in this example, the system reaches a **fixed point** at $t = 3$.

Metrics for Forecast Evaluations

Given N forecast values $\{f_t\}_{t=1}^{N}$ and the true values $\{y_t\}_{t=1}^{N}$ with error $e_t = y_t - f_t$, several metrics are used when evaluating point predictions in epidemic forecasting:

(*i*) The Mean Absolute Error $(MAE) = \frac{1}{N}\sum_{i=1}^{N} | e_i |$,

(*ii*) Mean Squared Error $(MSE) = \frac{1}{N}\sum_{i=1}^{N} e_i^2$,

(*iii*) Root Mean Squared Error $(\text{RMSE}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} e_i^2}$,

(*iv*) Mean Absolute Percentage Error $(\text{MAPE}) = \frac{1}{N}\sum_{i=1}^{N} \frac{| e_i |}{y_i}$, and (*v*) Pearson Correlation (PCORR). MAE, MSE, and RMSE are scale dependent while MAPE is invariant to scale. PCORR, on the other hand, is a useful measure when the forecasts follow the same trend as the true values. Properties and shortcomings of these metrics and their variants are discussed in [24].

Similar to weather forecasting, epidemic forecasting efforts have moved towards probabilistic forecasts. In the *CDCFlusSight* [19] challenge, participants are required to submit forecast distributions as binned disease incidence measures, and are evaluated using standard scoring rules like the logarithmic score [95]. This approach requires the possible range of incidence to be known, to define a meaningful binning system. However, during outbreaks like COVID-19, where the possible peak in the number of cases is not known, this range can vary considerably, making reasonable binning schemes challenging.

A suggested alternative to the binning system is to employ predictive quantiles or intervals, and has been the *de facto* standard in *The Hub*. The interval forecasts can be evaluated by the weighted interval score (WIS), the community's agreed upon metric [96]. WIS is defined as

$$[(y<lk)+2\alpha k(y-uk)1(y>uk)],$$

$$WIS_{\alpha_{0:K}}(F,y) = \frac{1}{K+0.5}\left[\begin{array}{l}\frac{1}{2}|y-m| + \sum_{k=0}^{K}\frac{\alpha_k}{2}\left(u_k - l_k\right) + \frac{2}{\alpha_k}\left(l_k - y\right)1\left(y < l_k\right) \\ + \frac{2}{\alpha_k}\left(y - u_k\right)1\left(y > u_k\right)\end{array}\right], \tag{9.3}$$

where y is the observed value (the reported case count) for a given location and date, F is the forecast defined in terms of the median m, upper quantiles u_k and lower quantiles l_k of the predictive distribution, respectively. K is the number of quantiles considered and $1 - \alpha_k$ is the width of the central predictive interval k. Also, $1\left(\text{condition}\right)$ is the indicator function whose value is 1 when the condition is satisfied and zero otherwise. The WIS falls under the category of *strictly proper scoring rules* [96] and provides no incentive for forecasters to report altered versions of their true belief in the future predictions.

AI-Driven Engineering

As described in section "Introduction", several challenges exist in real-time forecasting of epidemiological targets. A general framework for the design of data-driven systems is illustrated in Fig. 9.3. In the forecasting of a *target*, the models typically consume ground truth data and other epidemiological signals (syndromic and genomics data). In addition, information from auxiliary data sources are leveraged to account for insufficient data (for a rich set of data streams see [97]). In the following section we provide two examples: (*i*) a deployed ensemble model producing high-resolution forecasts and (*ii*) a hybrid model leveraging the spatio-temporal information and the simulation outputs of mechanistic models.

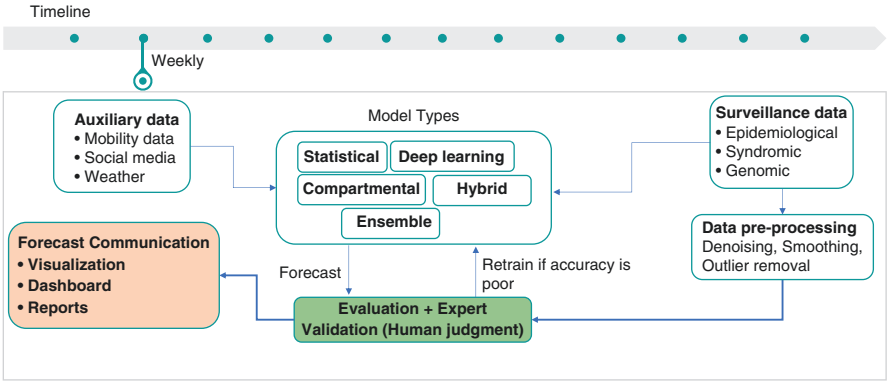


Fig. 9.3 A conceptual framework for real-time pandemic forecasting systems. Every week, data from various sources is downloaded, sanitized and converted into a format readable by the models. Multiple models use the available data and produce forecasts for the quantities desired. Ensembling techniques are typically used to combine the forecasts. The forecasts are constantly evaluated so that models can be updated based on the performance. Experts evaluate the forecasts and guide model modifications. Communicating the forecasts in a timely and appropriate manner is important

An Example of a Real-time Forecasting Model

In order to *understand* the state (national) level epidemic trajectories, it is important to consider the evolution at a finer resolution, such as the county (state) level. Forecasting at a higher geographical resolution can aid in targeted interventions and resource demand estimation. In addition, it can help to improve the overall forecasting quality and robustness at the state and national levels. However, accurate forecasting is challenging given the data-related issues at high resolution,. It is to be noted that of the many contributing teams to *The Hub*, only a handful consistently produce county-level forecasts. Since it is well accepted that ensemble forecasts tend to have better performance compared to individual forecasts [19, 98, 99], *The Hub* employs an ensemble approach to combine the probabilistic forecasts from multiple teams provided in the form of quantiles. *The Hub* has mostly relied upon an ensemble where each quantile of the ensemble forecast is obtained by taking the median of the forecast values provided for that quantile by the various models instead of a trained-ensemble that weights the models based on the past performance.

We present a real-time forecasting model, an example of a trained ensemble model that employs various classes of models. Although we present a US county-level forecasting model, this framework is applicable at multiple resolutions (county, state, multi-national levels, etc.). The modeling strategy remains the same at any resolution. This model was developed in the context of *The Hub*, and we generate

Table 9.1 A description of several classes of models incorporated in the ensemble model

Methods	Description
AR model and variants (AR, AR_spatial, ARIMA) (one model for each county, statistical)	• Lagged versions of observations (AR) + other county observations (AR_spatial) • ARIMA uses lagged versions of observations and forecast errors and requires fewer parameters compared to AR • Trained over a short observation window (8–10 weeks) • Each horizon modeled independently
Long short-term memory (LSTM) (data-driven)	• One model per state trained on short segments corresponding to respective counties • Observation window of four weeks and mean squared error loss function • Monte Carlo dropout method used for producing uncertainty • A highly non-linear model that can capture complex patterns in the time series
Ensemble Kalman filter (EnKF) (one model for each county, statistical)	• An approximate version of the Kalman filtering technique • EnKF updates computationally less expensive compared to Kalman filter updates • Trained on weekly cumulative case counts
SEIR model (one model for each county, compartmental model)	• Time-varying contact-rate parameter trained to match the infections (scaled) to the observed cases • Each county treated independently with homogeneous mixing within each county population • Incorporation of interventions is easier in this model

short-term forecasts of 1–4 weeks ahead, at a weekly resolution. Since we are dealing with real-time systems and highly non-stationary data, the individual models as well as the ensemble model need to be retrained every week. As mentioned previously, each method consumes data in a different format and trains in slightly different ways. We briefly describe the individual models, associated data preparation and the training procedures. The methods and some of their variants are described in Table 9.1. The motivation of the multi-class ensemble model is to leverage the benefits of various classes of methods.

Although the incident cases time series is non-stationary, in the AR models, we assume that any short segment of data is stationary and model it as an AR process. The data is log-transformed to nullify the large variations in variance across time, and the model is trained every week to account for non-stationarity. We also incorporate the lagged time series data of other counties corresponding to that state to capture the spatio-temporal correlation between counties. Unlike the recursive strategy of multi-step forecasting, where the prediction from the previous time step is used as an input for making the current time-step forecast, we use an independent model for each time step. This is to ensure that large errors in predictions from previous time steps do not propagate through the model (A trade-off is that accurate predictions from previous time steps are not incorporated in the prediction for the current time step.) Ensemble Kalman Filter (EnKF) [100] is an approximate form of Kalman filter that is particularly suited for practical applications. It represents the

distribution of the state with sample mean and covariance which makes the filter updates computationally less expensive than the update steps of the Kalman filter. LSTM networks are purely data-driven models; typically, they do not assume any statistical properties on the observations. Deep learning models usually require a large amount of training data which is not available in the context of COVID-19. For counties with small populations where the surveillance data is sparse and noisy, training a single model for each such region is highly susceptible to overfitting. We employed a clustering-based approach that simultaneously learns COVID-19 dynamics from multiple regions within the cluster and infers a model per cluster. Probabilistic forecasts are generated using MCDropout [101]. The compartmental model uses SEIR compartments to represent a county's population. Due to time-varying social distancing and variable prevalence/incidence across counties, we train each county model independently. The effects of social distancing measures and other *adaptations* are captured by the SEIR model's temporally varying transmissibility term β which is estimated sequentially using a simulation-optimization approach called the Golden Section Search (GSS) [102]. We refer the reader to [22] for a detailed description of models and the corresponding training schemes employed.

Bayesian Model Averaging Ensemble (BMA) Since there is considerable variation in the case counts across counties, unlike the strategies employed in [103, 104], we independently train a single BMA model per county. Considering K methods per county M_1, M_2, $\cdots$, M_K, we assume that each forecast f_k has a Normal distribution $\mathcal{N}\left(f_k, \tilde{\sigma}_k^2\right)$. In the BMA model, we assume that the probability of a forecast y given the mean of the individual forecasts is a Gaussian Mixture Model (GMM), $p\left(y|f_1|f_2|,\cdots|f_K\right) = \sum_{k=1}^{K} w_k g_k\left(y|f_k|\sigma_k\right)$, where w_k is the posterior probability of the k^{th} method's forecast being the best one and is determined using τ training samples. The term g_k is a normal probability density function with mean f_k and variance σ_k^2. We proceed to determine the weights w_k and σ_k. The weights and variance parameters of the GMM are obtained as the maximum likelihood estimate. It is to be noted that despite each method providing its uncertainty, we optimize further to refine it to obtain σ_k. Since the resulting log-likelihood function does not have an analytical solution, we employ the standard expectation-maximization (EM) algorithm [105, 106]. The optimization procedure is iterative and alternates between the E-step and the M-step with the updates for w_k and σ_k in the j^{th} iteration given by

$$z_{k,t}^{(j)} = \frac{w_k^{(j-1)} g\left(y_t|f_{k,t}|\sigma_k^{(j-1)}\right)}{\sum_{i=1}^{K} w_i^{(j-1)} g\left(y_t|f_{i,t}|\sigma_i^{(j-1)}\right)} \quad \left(\text{E}-\text{step}\right) \tag{9.4}$$

and

$$w_k^{(j)} = \frac{1}{\tau}\sum_{t=1}^{\tau} z_{k,t}^{(j)}, \quad \sigma_k^{(j)2} = \frac{\sum_t z_{k,t}\left(y_t - f_{k,t}\right)^2}{\sum_t z_{k,t}} \quad \left(\mathrm{M-step}\right) \tag{9.5}$$

The EM steps converge to a local minimum and the estimates are highly sensitive to initialization. For a detailed discussion on initialization we refer to [105, 106].

Results

Forecasts and Weights Distribution In an attempt to understand the spatio-temporal pattern of the dominant methods (i.e., methods with the highest weight) we present choropleth maps of US counties (Fig. 9.4a) where each county's color indicates the method with the highest weight for that particular week. In August 2020, we observed a nearly uniform distribution of weights. In November 2020, ARIMA appeared to dominate while in December 2020, the SEIR model is dominant. Figure 9.4b provides an insight into the number of counties for which a method is chosen for a particular week. The heat maps shown in that figure suggest that ARIMA gets picked by many counties consistently; while methods like EnKF dominate initially, LSTM starts to get higher weights for a lot more counties in the later months. Visual inspection indicates a spatio-temporal variation in the domination of

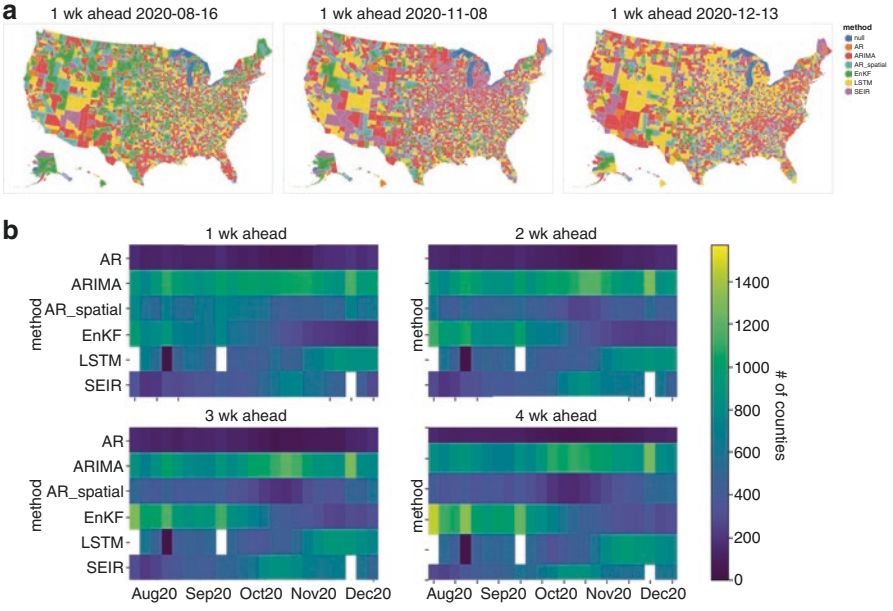


Fig. 9.4 Weights distribution: (**a**) Spatial distribution of methods with the highest weights for a county across three different forecast weeks. (**b**) A heatmap depicting the temporal evolution of the dominant methods across forecast weeks

methods. We next provide a quantitative evaluation of the ensemble and its constituent methods using standard metrics.

A Comparison of Individual Methods We evaluated 22 weeks of forecast data starting from the first week of August 2020 to the second week of January 2021 for all 3142 counties of the US. The performance across 1-, 2-, 3-, and 4-week ahead targets (horizon) are evaluated separately. Since county-level incident cases are typically small values, in all the evaluations, we consider only counties with observed forecast values greater than 10 cumulative cases. We employ MAE (cf. section "Metrics for Forecast Evaluations") for comparing the point forecasts.

First, we consider the overall performance of each method by computing the MAE across all the forecast weeks and counties. The results are shown in Fig. 9.5a, and they indicate that the average performance of individual models is similar. Also, the performance of the BMA ensemble is comparable to, if not better than, the best performing model which is in accordance with previous observations made in [99, 107]. The average performance of the methods across all counties for each forecast week is shown in Fig. 9.5b. As can be observed from that figure, in comparison to most methods, the BMA has lower variations.

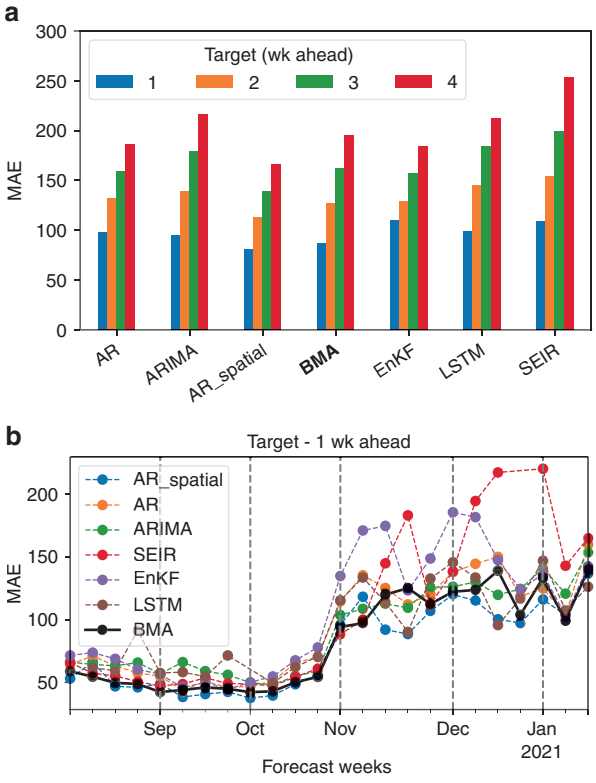
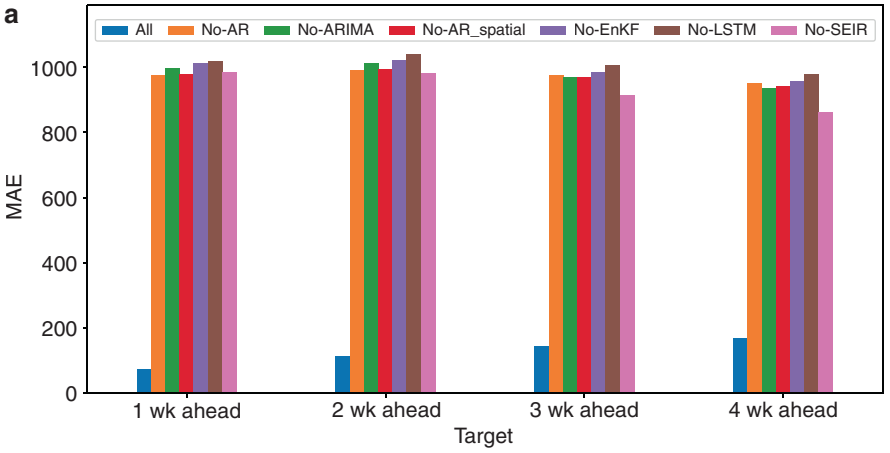


Fig. 9.5 The overall performance of individual methods and the BMA ensemble. (**a**) The overall performance is measured using the MAE computed across all counties and forecast dates. The results indicate that average performance of BMA is comparable to hte method with the least MAE. (**b**) Performance of BMA and the individual models across various forecast weeks since August 2020. The performance of individual models vary across weeks and while BMA shows relatively less variations across weeks

Relative Importance of Methods We performed ablation analysis by omitting the forecast of a specific method in the ensemble. The ensemble is retrained, and its performance is compared with the BMA method trained using all models. In Fig. 9.6a, we show the results of this experiment. The significant drop in performance seen in that figure can be attributed to large errors incurred by making poor forecasts for counties with large numbers of cases. Further, this can be explained using Fig. 9.6b, where we compute the percentage change in MAE after the removal of a method. Let $\text{MAE}^{(K)}$ denote the MAE for the BMA with all K models and let $\text{MAE}^{(K-1)}$ denote the MAE of the BMA with a specific method removed. The percentage change in performance is computed as $\frac{\text{MAE}^{(K)} - \text{MAE}^{(K-1)}}{\text{MAE}^{(K)}}$. We picked some of the top COVID-19 hit counties and determined the best performing method across all weeks for that county. For Maricopa County, Arizona, the SEIR model had the highest weights for eight weeks. When the SEIR model is dropped from the



b

County	Most picked method	MAE change on drop
Maricopa, AZ	SEIR (8)	1576.1%
Los Angeles, CA	AR_spatial (10)	1678.7%
San Bernardino, CA	ARIMA (11)	999.1%
Kings, NY	LSTM (10)	2085.4%

Fig. 9.6 Ablation analysis to study the relative importance of individual methods that feed into the BMA ensemble. The MAE is computed across all counties and forecast weeks. (**a**) A comparison of forecast performances of $\text{BMA}^{(K)}$ with the other K $\text{BMA}^{(K-1)}$ models. (**b**) Percentage change in MAE for a few county forecasts after dropping the most picked method (computed across forecast weeks) with respect to the MAE of BMA with all methods included. The number of times a method is picked is shown in parentheses

ensemble and the weights retrained, the relative change in performance is 1576.1%. This significant change indicates that the other methods could not compensate for the SEIR model.

A Comparison with The Hub Models To date, forecasts from over 100 models generated by dozens of teams have been submitted to *The Hub*, with the number of submissions varying every week. Only a handful of teams have been consistently providing county-level forecasts among the many teams. In order to make a fair comparison, we only consider teams that have been providing consistent forecasts across most counties and targets since August 2020. We employ the WIS metric defined by Eq. (9.3) for comparing the interval forecasts. Since there are over 3000 counties in the US, using the average of the scores computed across all the counties does provide a reasonable indication of a model's performance. However, it does not indicate the number of locations over which a model has provided low scores, which is essential for understanding a model's ability to provide quality forecasts across heterogeneous county-level data. Hence, for each forecasting week, we categorize the forecasts from different teams into three equal-width bins, denoted by low, medium and high, based on all the model WIS scores obtained for the particular week. Figure 9.7a, b, show the number of locations for which a each model had low WIS scores for one and four weeks ahead forecasts, respectively. Initially, the University of Virginia ensemble model (UVA-Ensemble (red)) had relatively poor performance; however, since December 2020, it has picked up and is comparable to

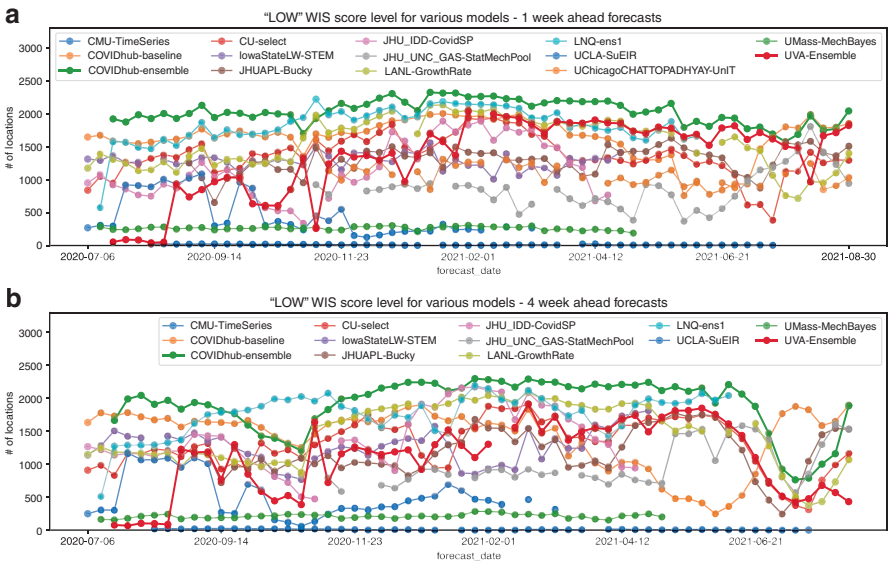


Fig. 9.7 A WIS-score-based performance comparison with several of *The Hub* models. (**a**) Number of locations for which a model had *low* WIS score for the 1 week ahead forecast. (**b**) Number of locations for which a model had *low* WIS score for the 4 weeks ahead forecast

the COVIDhub ensemble. It is to be noted that the COVIDhub ensemble consumes a suite of individually-tuned sophisticated models. Our trained BMA ensemble model can provide forecast performances comparable to the COVIDhub ensemble using relatively simple models. Since June 2021, *The Hub* has also deployed a trained ensemble with promising performance [108]; however, that ensemble is still not considered their primary model for forecast communication. We note that during July 2021 (the period of the *Delta wave*) most models were unable to capture the surge and exhibited poor performance. In fact, *The Hub* 's baseline model,the persistent forecast model (forecast value for the next week is the same as the current week's observed value) has a better performance.

A GNN-Based Spatio-Temporal Model

As discussed in section "Introduction", network-based compartmental models are a promising alternative to capturing the spatio-temporal dynamics of disease spread. However, it is challenging to employ those models due to their high computational requirements and data-related issues. Graph neural network models, also referred to as GNNs, are specifically designed to consume and extract patterns in graph data, where graphs are used to specify the connectivity among regions. However, GNNs have a large number of model parameters and require sufficiently large training data. Existing spatio-temporal forecasting models [109–114] have been shown to work well for applications such as traffic forecasting. However, these models tend to overfit when deployed on epidemiological signals as the data is sparser and noisier than traffic data. Also, some existing spatio-temporal epidemic forecasting models [65, 67] whose parameter size increases with the number of nodes in the graph failed to forecast over a large number of regions. In addition to the complexity issue, these models do not leverage the underlying physical phenomenon (i.e., the causal model) that governs the evolution of the target signal. The area of theory-guided data science [115] attempts to address this issue and demonstrate that incorporating domain knowledge into data-driven models can help to improve the performance of spatio-temporal forecasting algorithms. In the context of epidemiological signal forecasting, there have been a few efforts considering spatio-temporal forecasting [65, 67, 75]. We will now discuss one such effort that focuses on the specific task of COVID-19 forecasting.

A **Causal**-based **N**eural **N**etwork, referred to as a CausalGNN, attempts to jointly learn embeddings related to the spatio-temporal data (i.e., dynamic graphs) and the causal model (referred to as epidemiological context henceforth) represented by all the parameters and features corresponding to the disease models (e.g., S, I, R counts and other disease-specific parameters). Specifically, the framework has the following components.

- An attention-based dynamic GNN module that embeds spatial and temporal signals from disease dynamics using a connectivity network. The number of param-

eters in this module is agnostic to the number of regions and hence scale independent.
- A causal module encodes the single-patched compartmental model signals to provide epidemiological context. Unlike traditional network-based compartmental models, in this approach, the patches are connected via a learned GNN, and the calibration is done through computationally efficient GNN training.
- In order to ensure the flow of spatio-temporal information into the epidemiological model and vice versa, we allow for interaction between the causal and GNN modules. This is realized using a causal encoder which encodes causal features through node embedding. Similarly, the causal decoder is used to infer causal model parameters.

Additional Details Regarding the Framework

We assume N regions in total and define a dynamic graph $\mathcal{G}(\mathcal{V},\mathcal{E},\mathcal{T})$ on the N regions, where $\mathcal{V}$ is the set of N nodes, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of directed edges, and $\mathcal{T}$ is the set of T time points. At each time step t, the graph $\mathcal{G}$ is associated with a feature matrix $\mathbf{C}_t \in \mathrm{R}^{N \times C}$ where C is the number of features and the graph nodes connected via an adjacency matrix $\mathbf{A}_t \in \mathrm{R}^{N \times N}$. As mentioned earlier, it has two major modules: (i) an *attention-based dynamic GNN module* to capture the spatial and temporal disease dynamics via graph-based neural networks and (ii) a *causal module* to provide epidemiological context for GNN learning via ordinary differential equations that are applied at the node level. Each node runs its SIRD simulations using parameters inferred by the GNN module. Then at every time step, the nodes feed the simulation outputs back to the GNN module. The two parts learn mutually via the causal encoder and causal decoder layers. An illustration of the framework is shown in Fig. 9.8. Additional details regarding the modules, encoding mechanisms, optimization strategies, and the architecture are provided in [116].

Forecasting Performance

We evaluate our CausalGNN-based method and all baselines on forecasting COVID-19 daily new confirmed case counts at global, US state, and US county levels. The collected COVID-19 datasets are split into training-validation datasets (from May 3, 2020 to March 20, 2021) and testing datasets (from March 21, 2021 to April 23, 2021). For each targeted data point in a testing dataset, we make 7, 14, 21, and 28 days ahead forecasting of the data point. We compare the proposed method with a wide range of baselines including mechanistic causal models **SIR** and **PatchSEIR**; statistical models **AR** and **ARMA**; DL models **RNN**, **GRU**, and **LSTM**; spatio-temporal DL models **DCRNN**, **CNNRNN-Res**, and **LSTNet**, **STGCN**, **Cola-GNN**, and **STAN**. Major observations and implications are highlighted in the following paragraphs. The other details of experimental results can be found in our paper [116].

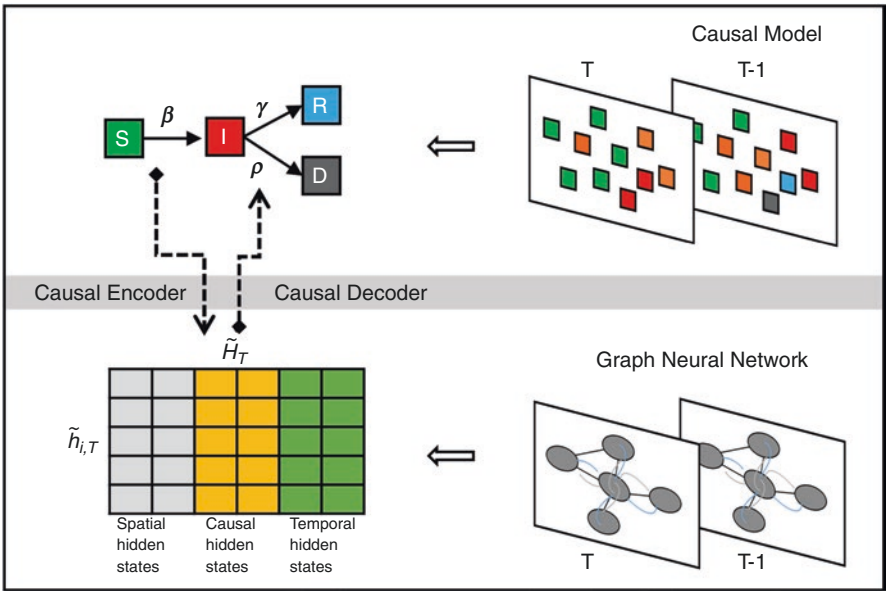


Fig. 9.8 A CausalGNN framework consisting of a causal module and an attention-based dynamic GNN module. Input to the GNN is a graph with time varying adjacency matrix and node features. The node embeddings are influenced by the local neighborhood properties and SIRD simulations

Major Observations The experimental results shown in the original paper [116] indicate that CausalGNN performs consistently better than the baselines across multiple scales and with increasing horizons. Two possible reasons for the better performance are the following: (i) our attention-based graph neural network architecture processes spatial and temporal signals efficiently and (ii) our method handles epidemiological context via an SIRD (SIR model with an additional compartment for Deaths) model at the node level. When considering spatio-temporal forecasting models, epidemic forecasting models (e.g., Cola-GNN, STAN, and CausalGNN) outperform models proposed for traffic forecasting (e.g., DCRNN, LSTNet, and STGCN). SIR and PatchSEIR perform worse than data-driven methods, especially for long-term forecasting. Compared with GNN-based models including STGCN, Cola-GNN, STAN, and CausalGNN, the vanilla RNN, GRU, LSTM models perform well when the time horizon is 7 or 14 days. However, as the horizon is increased, their performance is lower. This highlights the importance of capturing spatial disease transmission patterns in the input data for long term forecasting. In most cases, the classic statistical methods (AR, ARMA) show a poorer performance than the classic RNNs (RNN, GRU, LSTM). This implies the importance of modeling non-linear patterns for achieving good forecasting performance. We also conducted an ablation study on three critical components of our CausalGNN framework: the causal module, graph structure, and attention mechanism. The results show that all three components play important roles in improving our model's performance. Specifically, an adaptive adjacency matrix learnt by the attention

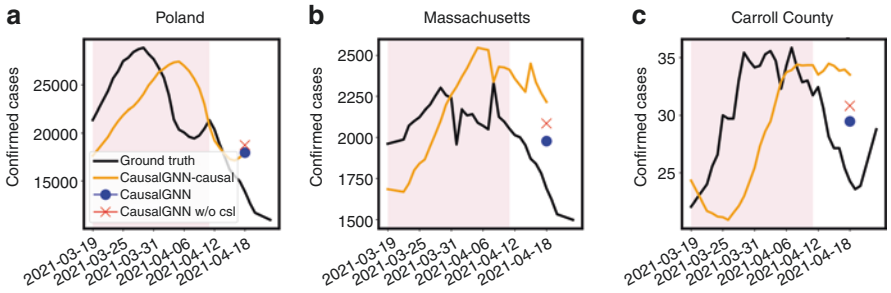


Fig. 9.9 Examples to demonstrate the impact of the causal module (csl). Locations at different spatial resolutions are considered. In all the examples, we observe that forecasts from CausalGNN are closer to the ground truth indicating the effect of the causal module. (**a**) Poland. (**b**) Massachusetts. (**c**) Carroll county

mechanism is crucial in capturing near-future dynamics. The Graph structure has the most impact in improving long-term forecasting performance. This agrees with the belief that incorporating cross-spatial signals is crucial for a good epidemic forecasting model. The results also demonstrate the effectiveness of the causal module in improving the epidemic forecasting performance.

Epidemiological Context In Fig. 9.9, we present examples of the causal module's (abbreviated as csl) impact by comparing 7 days ahead forecasts of 2021-04-18 by CausalGNN (blue dots) and CausalGNN w/o csl (red crosses) in Poland, Massachusetts, and Carroll County. Both solid lines and dots are smoothed, and the shaded area is the input window. One can observe the following: (i) CausalGNN makes better forecasts than CausalGNN w/o csl (i.e., the blue dots are closer to the black curves than the red crosses on the forecasting day). This means that the causal module proposed in our model can help in improving the model's performance; (ii) the causal module in CausalGNN can generate intuitively meaningful curves (orange curves) compared with the ground truth curves (black curves). This indicates that CausalGNN can infer a mechanistic causal process by producing meaningful causal parameters providing a meaningful epidemiological context for GNN learning. More examples and limitations are discussed in the paper [116].

Theoretical Foundations for Forecasting in Network Models

Overview

We present formal definitions of some short-term forecasting problems under the network model (introduced in section "Preliminaries") and observe that these problems are computationally intractable. Our goal is to point out that computational complexity imposes limits on the efficient computation of solutions to some

forecasting problems, even when one ignores the other difficulties (such as noisy and time-varying disease parameters) associated with forecasting. We also mention some restricted versions of forecasting problems that can be solved efficiently. Finally, we briefly discuss approaches used by researchers to deal with computationally intractable forecasting problems in practice.

Some Short-Term Forecasting Problems and Their Computational Intractability

Under the network-based SIR model discussed in section "Preliminaries", a typical forecasting problem involves the computation of an epidemic parameter. Rosenkrantz et al. [30] present rigorous formulations of many forecasting problems under that model. An example of such a problem, denoted by Pr-Num-Inf-at (t, q, S), is the following.

Problem Pr-Num-Inf-at (t, q, S):

Input: An undirected network $G(V, E)$, a transmission probability p_e for each edge $e \in E$, a subset $V \subseteq V$ of nodes that are infected at time 0, another subset $S \subseteq V$ of nodes, an integer $q \leq |S|$ and a time value t.

Requirement: Compute the probability that at least q nodes of S get infected *at* time t.

Several variants of the above problem have also been formulated in [30]. Here, we will mention two of these variants. For all of these variants, the input is the same as that of Pr-Num-Inf-at (t, q, S); the difference is in the parameter to be computed. For the variant denoted by Pr-Num-Inf-by (t, q, S), the quantity to be computed is the probability that at least q nodes of S get infected *by* time t. For the variant denoted by Pr-Peak-Inf-at (t), the goal is to compute the probability that the number of new infections reaches a peak *at* time t. In *short-term* forecasting problems, the parameter t is a small value (such as 2 or 3). Thus, all of these problems require the computation of epidemic parameter values for just a few time steps into the future.

Despite the short-term nature of these problems, it is observed in [30] that all of the above problems are *computationally intractable*. Specifically, it is shown that these problems are **#P** -hard [117]; that is, these are the hardest problems in the complexity class **#P**. A direct implication of this result is that if there is an efficient algorithm for any of these forecasting problems, then there is an efficient algorithm for every **#P** -hard problem. Some examples of **#P** -hard problems are given below; many other examples are presented in [117, 118]).

(a) **#SAT**: Given a Boolean formula F in conjunctive normal form[5] (CNF), compute the number of assignments of Boolean values to the variables of F for which F evaluates to true.

[5] A Boolean formula in conjunctive normal form [117] consists of clauses which are connected together by the AND operator. Each clause itself is the OR of negated or unnegated forms of a subset of variables.

(b) **#Perf-Match:** Given an undirected bipartite graph $H(V_1, V_2, E)$, where $|V_1| = |V_2|$, compute the number of perfect matchings[6] of H.

(c) **Path-Prob:** Given an undirected graph $G(V, E)$, a failure probability p_e for each edge $e \in E$ and two distinct nodes v_1 and v_2, compute the probability that there is a path between v_1 and v_2 when each edge e fails with probability p_e.

It is widely believed that none of the **#P** -hard problems can be solved efficiently [117, 119]. Thus, an efficient algorithm for any of the forecasting problems defined above would completely change our understanding of computational complexity.

Transforming a #P -Hard Problem into a Forecasting Problem We now illustrate through an example how a known **#P** -hard problem can be transformed into one of the forecasting problems mentioned above. This transformation is such that an efficient algorithm for the latter problem would imply a similar algorithm for the former. Following [30], we use a special version of the #SAT problem mentioned above. In this special version, which we refer to as #M2SAT, each clause contains exactly two *unnegated* variables. The #M2SAT problem was shown to be **#P** -hard in [120]. The target of the transformation is an instance of Pr-Num-Inf-at(t, q, S) problem with $t = 2$. Consider the following example of #M2SAT with four Boolean variables x_1, x_2, x_3 and x_4 and three clauses given by $C_1 = (x_1 \vee x_2)$, $C_2 = (x_3 \vee x_4)$ and $C_3 = (x_1 \vee x_4)$. The transformation from this instance of #M2SAT to Pr-Num-Inf-at(t, q, S) (with $t = 2$) is as follows.

1. The underlying graph has the following nodes: (i) a special node v_0, (ii) four nodes v_1, v_2, v_3 and v_4 that correspond to the four variables x_1, x_2, x_3 and x_4, and (iii) three more nodes w_1, w_2 and w_3 that correspond to the three clauses C_1, C_2 and C_3.
2. There is an edge between v_0 and v_i for $i = 1,2,3,4$. For each such edge, the probability value is 0.5.
3. For each node w_j corresponding to clause C_j, there are edges between w_j and the nodes corresponding to the variables in C_j. (Thus, in our example, node w_1 has edges to v_1 and v_2, node w_2 has edges to v_3 and v_4 and node w_3 has edges to v_1 and v_4.) The probability value for each of these edges is 1.0.
4. At time 0, node v_0 is in state I; all the other nodes are in state S.
5. To complete the specification of the Pr-Num-Inf-at(t, q, S) instance, we let $t = 2$, $q = 3$ and $S = \{w_1, w_2, w_3\}$. (Thus, for the resulting Pr-Num-Inf-at(t, q, S) instance, we need to compute the probability that all the nodes in S get infected at $t = 2$.)

The resulting instance of the Pr-Num-Inf-at(t, q, S) problem is shown in Fig. 9.10. In that figure, the graph on the left represents the occurrence of Boolean variables in the clauses. (For example, the presence of variables x_1 and x_2 in clause C_1 is represented by adding edges from the node C_1 to the nodes x_1 and x_2.) The graph on the right shows the SIR model network for the Pr-Num-Inf-at(t, q, S)

[6] A matching M of a bipartite graph $H(V_1, V_2, E)$ is a subset of edges so that no two edges are incident on the same node. When $|V_1| = |V_2| = n$, a perfect matching of H is a matching of size n.

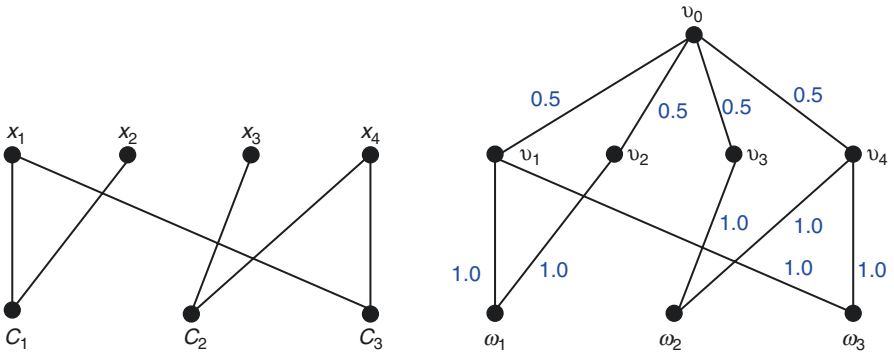


Fig. 9.10 Example to show a transformation from #M2SAT to Pr-Num-Inf-at (t, q, S), where $t = 2$, $q = 3$ and $S = \{w_1, w_2, w_3\}$. The graph on the left represents the occurrence of the Boolean variables x_1, x_2, x_3 and x_4 in the three clauses C_1, C_2 and C_3 of the #M2SAT instance. The graph on the right is the network corresponding to the Pr-Num-Inf-at (t, q, S) instance. The graph on the right is obtained by making a copy of the graph on the left (with nodes relabeled) and adding the node v_0 and the four edges from v_0. At time 0, node v_0 is in state I while all the other nodes of the graph on the right are in state S

instance. From the figure, the reader can see the one-to-one correspondence between the node sets $\{x_1, x_2, x_3, x_4\}$ and $\{v_1, v_2, v_3, v_4\}$, as well as that between the clause set $\{C_1, C_2, C_3\}$ and node set $\{w_1, w_2, w_3\}$.

It can be verified that of the $2^4 = 16$ possible assignments to the four variables x_1, x_2, x_3 and x_4, only 8 of them satisfy all the three clauses C_1, C_2 and C_3. Thus, the probability that an assignment to the four Boolean variables chosen uniformly randomly from the set of all 16 assignments satisfies the given CNF formula is 8/16 = 1/2. We can informally argue that the answer to the Pr-Num-Inf-at (t, q, S) created by the above construction is 8/16 = 1/2 as follows. Recall that $S = \{w_1, w_2, w_3\}$. From the graph in Fig. 9.10, it is seen that since v_0 is the only node that is in state I at time 0, any of the nodes in S can infected only at $t = 2$. Any subset of $Z = \{v_1, v_2, v_3, v_4\}$ that gets infected at time 1 can be thought of as an assignment of values to the Boolean variables x_1, x_2, x_3 and x_4 as follows: x_i is set to 1 iff v_i is infected at time 1, $1 \leq i \leq 4$. Further, since the probability value on each edge between a node in Z and one in S is 1.0, for all the the nodes in S to be infected at $t = 2$, the subset of Z that is infected at time 1 must correspond to a satisfying assignment to the CNF formula. Since the probability on each of the four edges $\{v_0, v_i\}$, $1 \leq i \leq 4$ is 0.5, the probability of any specific subset of Z getting infected at time 1 is 1/16. Since only 8 of the 16 assignments satisfy the CNF formula, the probability that the three nodes in S get infected at $t = 2$ is 8/16 = 1/2.

By presenting the above argument in a more general form, it is shown in [30] that if the CNF formula F (which is part of the #M2SAT instance) has n variables and the number of satisfying assignments to F is N, then the answer to the resulting Pr-Num-Inf-at (t, q, S) instance is $N/2^n$. Thus, knowing the answer to the Pr-Num-Inf-at (t, q, S) instance, one can efficiently find the value of N. This argument establishes

the **#P** -hardness of Pr-Num-Inf-at (t, q, S). By modifying this construction suitably, **#P** -hardness of other forecasting problems are established in [30]. Through more intricate constructions, this reference also shows that the forecasting problems remain **#P** -hard even when the underlying graphs are from special graph classes (e.g., power-law graphs, small-world networks). Further, these hardness results are also extended to other epidemic models (such as SIS, SI and probabilistic threshold models) and to other forecast parameters (e.g., takeoff value, takeoff time) defined in [24].

Some Efficiently Solvable Or Approximable Forecasting Problems While many short-term forecasting problems are computationally intractable, efficient algorithms or approximation algorithms are presented in [30] for some restricted versions. We mention some of these results below. For example, it is shown that the Pr-Num-Inf-at (t, q, S) and Pr-Num-Inf-by (t, q, S) problems can be solved efficiently when $t = 1$. This result, which points out the tightness of the computational intractability result for $t = 2$, is established by defining a more general formulation that captures many such forecasting problems and presenting an efficient dynamic programming algorithm for the general formulation. Further, efficient approximation algorithms are presented for some versions. For example, when t and $|S|$ are *fixed*, it is shown that the Pr-Num-Inf-by (t, q, S) problem can be efficiently approximated by transforming the problem to that of computing the number of satisfying assignments to a Boolean formula in disjunctive normal form[7] (DNF), and using a known approximation algorithm [121] for the latter problem.

Approaches to Handle Computationally Intractable Forecasting and Related Problems The computational intractability results for forecasting problems discussed above point out some fundamental limitations on the efficient solvability of such problems. However, these results do not rule out the possibility of algorithms that work well in practice. They suggest that practical approaches for dealing with such problems should try to exploit special properties of problem instances. For example, some restrictions that lead to efficiently solvable or approximable problems were mentioned above. Similar results for other computationally intractable problems involving epidemic dynamics have appeared in the literature. For instance, efficient approximation algorithms for the vaccine distribution problem have been developed [122, 123]; these algorithms rely on stochastic optimization techniques and spectral properties of the underlying networks. Researchers have also used heuristic approaches for forecasting and related problems that work well in practice. For example, highly effective approaches based on standard AI techniques (e.g., belief propagation, message passing) in conjunction with other network analysis techniques have been developed for identifying the infection sources in network models (e.g., [87, 88, 124]). Thus, in addition to their importance from a founda-

[7] A Boolean formula in disjunctive normal form consists of product terms which are connected together by the OR operator. Each product term is the AND of a subset of negated and/or unnegated variables.

tional point of view, computational intractability results for problems in epidemic dynamics may also be useful in selecting appropriate practical approaches for dealing with such problems.

Discussion

Infectious disease forecasting is a rapidly evolving discipline with significant scope for improvement across tools, techniques, platforms, and policy-making. As discussed in previous sections, forecasting COVID-19 dynamics yielded important lessons. First, our work and that of others showed that simpler models are easy to set up and useful during certain pandemic phases. On the other hand, it became difficult to keep them regularly updated to handle ever-changing surveillance data caused by epidemiological and socio-behavioral processes. Second, providing weekly high-resolution forecasts (e.g., at the county level) requires considerable scaling up and optimization of individual models. Understanding how the different data streams and modeling techniques can be integrated in a timely yet reliable fashion remains a challenge. The epidemic outcome being forecast (e.g., case counts, hospitalizations, deaths) can be used to guide interventions, supplement healthcare resources or shape public messaging. Third, the modeling community as a whole struggled to capture multiple waves of the pandemic; the speed at which cases increased and, in some cases decreased (e.g., the Omicron variant) made it very challenging to keep models updated. Several factors (e.g., the biology of the variants, surveillance and immunity data, and socio-political factors) contributed additional challenges to the modeling community. Existing frameworks such as the CDC Forecasting Hub (*The Hub*) did an outstanding job wrangling forecasts from multiple modeling teams to provide weekly updates to policymakers. Nevertheless, the community often failed to produce accurate and timely forecasts as a collective. Despite exhibiting good performance compared to individual models, an ensemble model's forecast quality is dependent on the quality of its constituent models. Notwithstanding the efforts of the forecasting community, there is a considerable lack of understanding of disease dynamics as teams have struggled to predict the onset, growth rate, peak size, and duration of the various waves. Consequently, the ensemble models have suffered from outlying forecasts and have failed to detect local peaks [76], a trend that has also been observed in our ensemble model forecasts. These efforts have also identified gaps in our understanding and need for specific real-time data streams.

Despite these challenges, our work and others have demonstrated how a forecasting pipeline based on a trained ensemble can still provide meaningful forecasts during a rapidly evolving pandemic. While we integrated methods from different modeling paradigms, an exhaustive search over the model space is challenging. Hierarchical model selection approaches along with ensembles that account for model complexity and diversity will help produce more robust frameworks. Methods to understand change points in pandemic dynamics also need to be

developed. Expert feedback as part of the loop is useful and can refine the objective functions by which the models are evaluated, leading to meaningful forecasts.

We conclude by discussing some related topics that we have not covered in detail but are essential; these topics will be the subject of future works.

Biosurveilliance Biosurveillance generally refers to the systematic collection of disease-specific information from various sources such as hospitals, laboratories, pharmacies, weather monitoring systems, water management systems, etc. The data collected in this manner can be very useful in the early detection and monitoring of pathogen and disease outbreaks [125]. *Syndromic* (e.g., US Outpatient Influenza-like Illness Surveillance Network (ILINet) [17]) and *epidemiological* surveilliance systems [126] help in monitoring and assessing the incidence of diseases. *Genomic* surveillance, which deals with pathogen sequences and studying of variants, is crucial in understanding the pathogen dynamics and detecting variants of concern (VoC) [127]. These systems have gained traction through several large-scale initiatives [128, 129]. Genomic surveilliance data holds crucial information related to transmissibility and competing pathogen strains; it is crucial to incorporate this information in forecasting syndromic or epidemiological data. In addition, there is considerable benefit in developing models for forecasting mutation rates and the evolution of variants. Many important questions arise when one focuses on forecasting viral (more broadly pathogen) dynamics; examples of such questions include: (*i*) when will a new variant emerge in a given region? (*ii*) when/if will a new version become a dominant variant? (*iii*) what are the chances of a spillover event for a new pathogen/virus? (*iv*) will the spillover event lead to a sustained spread in the human population? These are just some of the forecasting issues pertaining to viral dynamics.

Social Contagions As discussed earlier and observed during the COVID-19 pandemic, the spread of a virus through human populations is accompanied by the spread of corresponding social and behavioral contagions. Forecasting the spread of these co-evolving contagions is central in understanding the pandemic dynamics. Again, a multitude of questions arise when studying social and behavioral contagions during a pandemic. Examples of such questions include: (*i*) will vaccine acceptance reach a given threshold by a given time? (*ii*) when can one expect (say) 75% of the population to wear masks at all times outside their homes? (*iii*) how fast will misinformation related to a given drug spread? (*iv*) how long after a complete lockdown can we expect the infection rate to fall below a given threshold?

Forecast Communication The primary motivation for forecasting efforts is to help public health agencies understand and mobilize resources effectively. In addition, effectively communicating epidemic forecasts, similar to weather forecasts, can help create awareness and better behavioral adaptations among the general public. Summarizing the forecasts, especially during periods of high variability in the quality of forecasts, is particularly challenging as the summaries could create panic and distrust in epidemic modeling.

Acknowledgments We thank members of the Biocomplexity COVID-19 Response Team and the Network Systems Science and Advanced Computing (NSSAC) Division of the University of Virginia for their thoughtful comments and suggestions related to epidemic modeling and response support. This work has been partially supported by DTRA (Contract HDTRA1-19-D-0007), University of Virginia Strategic Investment Fund award number SIF160, VDH Grant PV-BII VDH COVID-19 Modeling Program VDH-21-501-0135, NSF XSEDE Grant TG-BIO210084, National Institutes of Health (NIH) Grants 1R01GM109718, 2R01GM109718, The James S. McDonnell Foundation twenty-first Century Science Initiative Collaborative Award in Understanding Dynamic and Multi-scale Systems, the C3.ai Digital Transformation Institute and Microsoft Corporation, Gift from Google, LLC, and the National Science Foundation (NSF) grants IIS-1633028 (BIG DATA), CMMI-1745207 (EAGER), OAC-1916805 (CINES), CCF-1918656 (Expeditions), CNS-2028004 (RAPID), OAC-2027541 (RAPID), CCF-2142997 (RAPID), CNS-2041952 (PREPARE), IIS-1908530, IIS-1955797, IIS-2027848, CNS-2027908 and CCF1917819. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

References

1. Barrios JM, Hochberg YV. Risk perceptions and politics: evidence from the COVID-19 pandemic. J Financ Econ. 2021;142(2):862–79.
2. Brzezinski A, Kecht V, Van Dijcke D, Wright AL. Science skepticism reduced compliance with COVID-19 shelter-in-place policies in the United States. Nat Hum Behav. 2021;5(11):1519–27.
3. Fancourt D, Steptoe A, Wright L. The Cummings effect: politics, trust, and behaviours during the COVID-19 pandemic. Lancet. 2020;396(10249):464–5.
4. Harman JL, Weinhardt JM, Beck JW, Mai I. Interpreting time-series COVID data: reasoning biases, risk perception, and support for public health measures. Sci Rep. 2021;11(1):1–11.
5. Levin R, Chao DL, Wenger EA, Proctor JL. Insights into population behavior during the COVID-19 pandemic from cell phone mobility data and manifold learning. Nat Computat Sci. 2021;1(9):588–97.
6. Van Bavel JJ, Cichocka A, Capraro V, Sjåstad H, Nezlek JB, Pavlović T, Alfano M, Gelfand MJ, Azevedo F, Birtel MD, et al. National identity predicts public health support during a global pandemic. Nat Commun. 2022;13(1):1–14.
7. Woods ET, Schertzer R, Greenfeld L, Hughes C, Miller-Idriss C. COVID-19, nationalism, and the politics of crisis: a scholarly exchange. Nations National. 2020;26(4):807–25.
8. Adiga A, Dubhashi D, Lewis B, Marathe M, Venkatramanan S, Vullikanti A. Mathematical models for COVID-19 pandemic: a comparative analysis. J Indian Inst Sci. 2020:1–15.
9. Borchering RK, Viboud C, Howerton E, Smith CP, Truelove S, Runge MC, Reich NG, Contamin L, Levander J, Salerno J, et al. Modeling of future COVID-19 cases, hospitalizations, and deaths, by vaccination rates and nonpharmaceutical intervention scenarios—United States, April–September 2021. Morb Mortal Wkly Rep. 2021;70(19):719.
10. Chen J, Levin S, Eubank S, Mortveit H, Venkatramanan S, Vullikanti A, Marathe M. Networked epidemiology for COVID-19. SIAM News, June 2020.
11. Eletreby R, Zhuang Y, Carley KM, Yagan O, Poor HV. The effects of evolutionary adaptations on spreading processes in complex networks. Proc Natl Acad Sci. 2020;117(11):5664–70.
12. Hu B, Guo H, Zhou P, Shi Z-L. Characteristics of SARS-CoV-2 and COVID-19. Nat Rev Microbiol. 2021;19(3):141–54.
13. Saad-Roy CM, Morris SE, Metcalf CJE, Mina MJ, Baker RE, Farrar J, Holmes EC, Pybus OG, Graham AL, Levin SA, Grenfell BT, Wagner CE. Epidemiological and evolutionary considerations of SARS-CoV-2 vaccine dosing regimes. Science. 2021;372(6540):363–70.

14. Wagner CE, Saad-Roy CM, Morris SE, Baker RE, Mina MJ, Farrar J, Holmes EC, Pybus OG, Graham AL, Emanuel EJ, et al. Vaccine nationalism and the dynamics and control of SARS-CoV-2. Science. 2021;373(6562):eabj7364.
15. Yagan O, Sridhar A, Eletreby R, Levin S, Plotkin JB, Poor HV. Modeling and analysis of the spread of COVID-19 under a multiple-strain model with mutations. Harvard Data Science Review. 2021;4. https://doi.org/10.1162/99608f92.a11bf693. URL https://hdsr.mitpress.mit.edu/pub/2q9jiymv.
16. Biggerstaff M, Alper D, Dredze M, Fox S, Fung IC-H, Hickmann KS, Lewis B, Rosenfeld R, Shaman J, Tsou M-H, et al. Results from the centers for disease control and prevention's predict the 2013–2014 influenza season challenge. BMC Infect Dis. 2016;16(1):1–10.
17. Biggerstaff M, Johansson M, Alper D, Brooks LC, Chakraborty P, Farrow DC, Hyun S, Kandula S, McGowan C, Ramakrishnan N, et al. Results from the second year of a collaborative effort to forecast Influenza seasons in the United States. Epidemics. 2018;24:26–33.
18. McGowan CJ, Biggerstaff M, Johansson M, Apfeldorf KM, Ben-Nun M, Brooks L, Convertino M, Erraguntla M, Farrow DC, Freeze J, et al. Collaborative efforts to forecast seasonal Influenza in the United States, 2015–2016. Sci Rep. 2019;9(1):1–13.
19. Reich NG, Brooks LC, Fox SJ, Kandula S, McGowan CJ, Moore E, Osthus D, Ray EL, Tushar A, Yamana TK, et al. A collaborative multiyear, multimodel assessment of seasonal Influenza forecasting in the United States. Proc Natl Acad Sci. 2019;116(8):3146–54.
20. DARPA. CHIKV Challenge announces winners, progress toward forecasting the spread of infectious diseases. 2015. http://www.darpa.mil/news-events/2015-05-27.
21. Muthiah S, Butler P, Khandpur RP, Saraf P, Self N, Rozovskaya A, Zhao L, Cadena J, Lu C, Vullikanti A, Marathe A, Summers KM, Katz G, Doyle A, Arredondo J, Gupta DK, Mares D, Ramakrishnan N. EMBERS at 4 years: experiences operating an open source indicators forecasting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining; 2016. p. 205–14.
22. Adiga A, Wang L, Hurt B, Peddireddy A, Porebski P, Venkatramanan S, Lewis B, Marathe MV. All models are useful: Bayesian ensembling for robust high resolution COVID-19 forecasting. In: KDD '21: proceedings of the 27th ACM SIGKDD conference on knowledge discovery and data mining; 2021. p. 2505–13.
23. Chakraborty P, Lewis B, Eubank S, Brownstein JS, Marathe M, Ramakrishnan N. What to know before forecasting the Flu. PLoS Comput Biol. 2018;14(10):e1005964.
24. Tabataba FS, Chakraborty P, Ramakrishnan N, Venkatramanan S, Chen J, Lewis B, Marathe M. A framework for evaluating epidemic forecasts. BMC Infect Dis. 2017;17(1):345.
25. Tabataba FS, Lewis B, Hosseinipour M, Tabataba FS, Venkatramanan S, Chen J, Higdon D, Marathe M. Epidemic forecasting framework combining agent-based models and smart beam particle filtering. In: 2017 IEEE international conference on data mining (ICDM). IEEE; 2017. p. 1099–104.
26. Drake JM. Fundamental limits to the precision of early warning systems for epidemics of infectious diseases. PLoS Med. 2005;2:461–2.
27. Drake JM. Limits to forecasting precision for outbreaks of directly transmitted diseases. PLoS Med. 2006;3:57–62.
28. May RM. Network structure and the biology of populations. Trends Ecol Evol. 2006;21(7):394–9.
29. Jacob F. Evolution and tinkering. Science. 1977;196:1161–6.
30. Rosenkrantz DJ, Vullikanti A, Ravi SS, Stearns RE, Levin S, Poor HV, Marathe MV. Fundamental limitations on efficiently forecasting epidemic measures in network models. Proc Nat Acad Sci (PNAS). 2022;119(4):1–7.
31. Nsoesie E, Mararthe M, Brownstein J. Forecasting peaks of seasonal Influenza epidemics. PLoS Curr. 2013;5.
32. Yang S, Santillana M, Kou SC. Accurate estimation of Influenza epidemics using Google search data via ARGO. Proc Natl Acad Sci. 2015;112(47):14473–8.

33. Rangarajan P, Mody SK, Marathe M. Forecasting Dengue and Influenza incidences using a sparse representation of Google trends, electronic health records, and time series data. PLoS Comput Biol. 2019;15(11):e1007518.
34. Kandula S, Hsu D, Shaman J. Subregional nowcasts of seasonal Influenza using search trends. J Med Internet Res. 2017;19(11):e370.
35. Soebiyanto RP, Adimi F, Kiang RK. Modeling and predicting seasonal Influenza transmission in warm regions using climatological parameters. PLoS One. 2010;5(3):e9450.
36. Paul MJ, Dredze M, Broniatowski D. Twitter improves Influenza forecasting. PLoS Curr. 2014;6.
37. Wang Z, Chakraborty P, Mekaru SR, Brownstein JS, Ye J, Ramakrishnan N. Dynamic Poisson autoregression for Influenza-like-illness case count prediction. In: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining; 2015. p. 1285–94.
38. Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, Igusa T, Rothman RE. Influenza forecasting with Google Flu Trends. PLoS One. 2013;8(2):e56176.
39. Radin JM, Wineinger NE, Topol EJ, Steinhubl SR. Harnessing wearable device data to improve state-level real-time surveillance of Influenza-like illness in the USA: a population-based study. Lancet Digit Health. 2020;2(2):e85–93.
40. Tibshirani R. Regression shrinkage and selection via the Lasso. J R Statist Soc Ser B (Methodol). 1996;58(1):267–88.
41. Tseng P. Convergence of a block coordinate descent method for nondifferentiable minimization. J Optim Theory Appl. 2001;109(3):475–94.
42. Brooks LC, Farrow DC, Hyun S, Tibshirani RJ, Rosenfeld R. Flexible modeling of epidemics with an empirical bayes framework. PLoS Comput Biol. 2015;11(8):e1004382.
43. Viboud C, Boëlle P-Y, Carrat F, Valleron A-J, Flahault A. Prediction of the spread of Influenza epidemics by the method of analogues. Am J Epidemiol. 2003;158(10):996–1006.
44. Van Panhuis WG, Hyun S, Blaney K, Marques ET Jr, Coelho GE, Siqueira JB Jr, Tibshirani R, da Silva Jr JB, Rosenfeld R. Risk of Dengue for tourists and teams during the World Cup 2014 in Brazil. PLoS Negl Trop Dis. 2014;8(7):e3063.
45. Petropoulos F, Makridakis S. Forecasting the novel coronavirus COVID-19. PLoS One. 2020;15(3):e0231236.
46. Evensen G. Data assimilation: the ensemble Kalman filter. Springer; 2009.
47. Anderson JL. An ensemble adjustment Kalman filter for data assimilation. Mon Weather Rev. 2001;129(12):2884–903.
48. Arulampalam MS, Maskell S, Gordon N, Clapp T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans Signal Process. 2002;50(2):174–88.
49. Yang W, Karspeck A, Shaman J. Comparison of filtering methods for the modeling and retrospective forecasting of Influenza epidemics. PLoS Comput Biol. 2014;10(4):e1003583.
50. Shaman J, Karspeck A. Forecasting seasonal outbreaks of Influenza. Proc Natl Acad Sci. 2012;109(50):20425–30.
51. Shaman J, Karspeck A, Yang W, Tamerius J, Lipsitch M. Real-time Influenza forecasts during the 2012–2013 season. Nat Commun. 2013;4:2837.
52. Yang W, Cowling BJ, Lau EH, Shaman J. Forecasting Influenza epidemics in Hong Kong. PLoS Comput Biol. 2015;11(7):e1004383.
53. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Netw. 1989;2(5):359–66.
54. Aburas HM, Cetiner BG, Sari M. Dengue confirmed-cases prediction: a neural network model. Expert Syst Appl. 2010;37(6):4256–60.
55. Wahyunggoro O, Permanasari AE, Chamsudin A. Utilization of neural network for disease forecasting. In: 59th ISI world statistics congress. Citeseer; 2013. p. 549–54.
56. Xu Q, Gel YR, Ramirez LL, Nezafati K, Zhang Q, Tsui K-L. Forecasting Influenza in Hong Kong with Google search queries and statistical model fusion. PLoS One. 2017;12(5):e0176690.

57. Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555. 2014.
58. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80.
59. Volkova S, Ayton E, Porterfield K, Corley CD. Forecasting Influenza-like illness dynamics for military populations using neural networks and social media. PLoS One. 2017;12(12):e0188941.
60. Venna SR, Tavanaei A, Gottumukkala RN, Raghavan VV, Maida AS, Nichols S. A novel data-driven model for real-time Influenza forecasting. IEEE Access. 2019;7:7691–701.
61. Zhu X, Fu B, Yang Y, Ma Y, Hao J, Chen S, Liu S, Li T, Liu S, Guo W, et al. Attention-based recurrent neural network for Influenza epidemic prediction. BMC Bioinform. 2019;20(18):1–10.
62. Adhikari B, Xu X, Ramakrishnan N, Prakash BA. EpiDeep: exploiting embeddings for epidemic forecasting. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining; 2019. p. 577–86.
63. Rodriguez A, Tabassum A, Cui J, Xie J, Ho J, Agarwal P, Adhikari B, Prakash BA. DeepCOVID: an operational deep learning-driven framework for explainable real-time COVID-19 forecasting. medRxiv. 2020.
64. Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos, Solitons & Fractals; 2020. p. 109864.
65. Wu Y, Yang Y, Nishiura H, Saitoh M. Deep learning for epidemiological predictions. In: The 41st international ACM SIGIR conference on research & development in information retrieval. ACM; 2018. p. 1085–8.
66. Ramchandani A, Fan C, Mostafavi A. DeepCOVIDNet: an interpretable deep learning model for predictive surveillance of COVID-19 using heterogeneous features and their interactions. arXiv preprint arXiv:2008.00115. 2020.
67. Deng S, Wang S, Rangwala H, Wang L, Ning Y. Cola-GNN: cross-location attention based graph neural networks for long-term ILI prediction. In: Proceedings of the 29th ACM international conference on information and knowledge management; 2020. p. 245–54.
68. Kapoor A, Ben X, Liu L, Perozzi B, Barnes M, Blais M, O'Banion S. Examining COVID-19 forecasting using spatio-temporal graph neural networks. arXiv preprint arXiv:2007.03113. 2020.
69. Wang L, Ben X, Adiga A, Sadilek A, Tendulkar A, Venkatramanan S, Vullikanti A, Aggarwal G, Talekar A, Chen J, et al. Using mobility data to understand and forecast COVID-19 dynamics. In: IJCAI 2021 workshop on AI for social good; 2021.
70. Wang L, Adiga A, Chen J, Lewis B, Sadilek A, Venkatramanan S, Marathe M. Combining theory and data driven approaches for epidemic forecasts. CRC Press (to appear). 2022. https://sites.google.com/vt.edu/sgml-book.
71. Zhao L, Chen J, Chen F, Wang W, Lu C-T, Ramakrishnan N. Simnest: social media nested epidemic simulation via online semi-supervised deep learning. In: 2015 IEEE international conference on data mining. IEEE; 2015. p. 639–48.
72. Hua T, Reddy CK, Zhang L, Wang L, Zhao L, Lu C-T, Ramakrishnan N. Social media based simulation models for understanding disease dynamics. In: Proceedings of the twenty-seventh international joint conference on artificial intelligence, IJCAI-18, International Joint Conferences on Artificial Intelligence Organization; 2018. p. 3797–804.
73. Wang L, Chen J, Marathe M. DEFSI: deep learning based epidemic forecasting with synthetic information. In: Proceedings of the AAAI conference on artificial intelligence, vol. 33; 2019. p. 9607–12.
74. Dandekar R, Rackauckas C, Barbastathis G. A machine learning-aided global diagnostic and comparative tool to assess effect of quarantine control in COVID-19 spread. Patterns. 2020;1(9):100145.
75. Gao J, Sharma R, Qian C, Glass LM, Spaeder J, Romberg J, Sun J, Xiao C. STAN: spatio-temporal attention network for pandemic prediction using real-world evidence. J Am Med Inform Assoc. 2021;28(4):733–43.

76. Ray EL et al. Challenges in training ensembles to forecast COVID-19 cases and deaths in the United States. International Institute of Forecasters; 2021.
77. Morgan JJ, Wilson OC, Menon PG. The wisdom of crowds approach to Influenza-rate forecasting. In: ASME international mechanical engineering congress and exposition, vol. 52026, page V003T04A048. American Society of Mechanical Engineers; 2018.
78. Taylor KS, Taylor JW. Harnessing the wisdom of the crowd to forecast incident and cumulative COVID-19 mortality in the United States. medRxiv. 2021.
79. Farrow DC, Brooks LC, Hyun S, Tibshirani RJ, Burke DS, Rosenfeld R. A human judgment approach to epidemiological forecasting. PLoS Comput Biol. 2017;13(3):e1005248.
80. Li EY, Tung C-Y, Chang S-H. The wisdom of crowds in action: forecasting epidemic diseases with a web-based prediction market system. Int J Med Inform. 2016;92:35–43.
81. Cheng J, Adamic LA, Kleinberg JM, Leskovec J. Do cascades recur? In: Proceedings of the 25th international conference on world wide web, WWW 2016, Montreal, Canada, April 11–15; 2016. p. 671–81.
82. Martin T, Hofman JM, Sharma A, Anderson A, Watts DJ. Exploring limits to prediction in complex social systems. In: Proceedings of the 25th international conference on world wide web, WWW 2016, Montreal, Canada, April 11–15, 2016. p. 683–94.
83. Hofman JM, Sharma A, Watts DJ. Prediction and explanation in social systems. Science. 2017;355:486–8.
84. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google Flu: traps in Big Data analysis. Science. 2014;343:1203–5.
85. Pinto PC, Thiran P, Vetterli M. Locating the source of diffusion in large-scale networks. Phys Rev Lett. 2012;109(6):1–4.
86. Karrer B, Newman ME. Message passing approach for general epidemic models. Phys Rev E. 2010;82(1):016101.
87. Altarelli F, Braunstein A, Dall'Asta L, Lage-Castellanos A, Zecchina R. Bayesian inference of epidemics on networks via Belief Propagation. Phys Rev Lett. 2014;112:118701.
88. Lokhov AY, Mézard M, Ohta H, Zdeborová L. Inferring the origin of an epidemic with a dynamic message-passing algorithm. Phys Rev E. 2014;90:012801.
89. Althouse BM, Wenger EA, Miller JC, Scarpino SV, Allard A, Hébert-Dufresne L, Hu H. Superspreading events in the transmission dynamics of SARS-CoV-2: opportunities for interventions and control. PLoS Biol. 2020;18(11):e3000897.
90. Scarpino SV, Petri G. On the predictability of infectious disease outbreaks. Nat Commun. 2019;10:1–8. https://doi.org/10.1038/s41467-019-08616-0.
91. Brauer F, van den Driessche P, Wu J. Mathematical epidemiology, volume 1945 of Springer Verlag, Lecture Notes in Mathematics. Springer; 2008.
92. Marathe MV, Vullikanti AK. Computational Epidemiology. Commun ACM. 2013;56(7):88–96.
93. Newman M. The structure and function of complex networks. SIAM Rev. 2003;45(2):167–256.
94. Lipsitch M, Cohen T, Cooper B, Robins JM, Ma S, James L, Gopalakrishna G, Chew S, Tan CC, Samore MH, Fisman D, Murray M. Transmission dynamics and control of severe acute respiratory syndrome. Science. 2003;300:1966–70.
95. Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc. 2007;102(477):359–78.
96. Bracher J, Ray EL, Gneiting T, Reich NG. Evaluating epidemic forecasts in an interval format. PLoS Comput Biol. 2021;17(2):e1008618.
97. Arnold T, Bien J, Brooks L, Colquhoun S, Farrow D, Grabman J, Maynard-Zhang P, Reinhart A, Tibshirani R. Covidcast: client for Delphi's COVIDcast epidata API. 2021. URL https://cmu-delphi.github.io/covidcast/covidcastR/. R package version 0.4.2.
98. Viboud C, Sun K, Gaffey R, Ajelli M, Fumanelli L, Merler S, Zhang Q, Chowell G, Simonsen L, Vespignani A, et al. The RAPIDD Ebola forecasting challenge: synthesis and lessons learnt. Epidemics. 2018;22:13–21.
99. Yamana TK, Kandula S, Shaman J. Superensemble forecasts of Dengue outbreaks. J R Soc Interface. 2016;13(123):20160410.

100. Burgers G, Jan van Leeuwen P, Evensen G. Analysis scheme in the ensemble Kalman filter. Mon Weather Rev. 1998;126(6):1719–24.
101. Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: International conference on machine learning; 2016. p. 1050–9.
102. Kiefer J. Sequential minimax search for a maximum. Proc Am Math Soc. 1953;4(3):502–6.
103. Raftery AE, Gneiting T, Balabdaoui F, Polakowski M. Using Bayesian model averaging to calibrate forecast ensembles. Mon Weather Rev. 2005;133(5):1155–74.
104. Yamana TK, Kandula S, Shaman J. Individual versus superensemble forecasts of seasonal Influenza outbreaks in the United States. PLoS Comput Biol. 2017;13(11):e1005801.
105. Bilmes JA, et al. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. Int Comp Sci Inst. 1998;4(510):126.
106. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Statist Soc Se B (Methodol). 1977;39(1):1–22.
107. Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial. Statist Sci. 1999;14(4):382–401.
108. COVID-Hub. The COVID-19 Forecast Hub community. 2021. https://covid19forecast-hub.org/.
109. Bai L, Yao L, Li C, Wang X, Wang C. Adaptive graph convolutional recurrent network for traffic forecasting. arXiv preprint arXiv:2007.02842. 2020.
110. Lai G, Chang W-C, Yang Y, Liu H. Modeling long-and short-term temporal patterns with deep neural networks. In: The 41st international ACM SIGIR conference on research & development in information retrieval; 2018. p. 95–104.
111. Li Y, Yu R, Shahabi C, Liu Y. Diffusion convolutional recurrent neural network: data-driven traffic forecasting. arXiv preprint arXiv:1707.01926. 2017.
112. Wu Z, Pan S, Long G, Jiang J, Zhang C. Graph wavenet for deep spatial-temporal graph modeling. arXiv preprint arXiv:1906.00121. 2019.
113. Wu Z, Pan S, Long G, Jiang J, Chang X, Zhang C. Connecting the dots: multivariate time series forecasting with graph neural networks. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining; 2020. p. 753–63.
114. Yu B, Yin H, Zhu Z. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875. 2017.
115. Karpatne A, Atluri G, Faghmous JH, Steinbach M, Banerjee A, Ganguly A, Shekhar S, Samatova N, Kumar V. Theory-guided data science: a new paradigm for scientific discovery from data. IEEE Trans Knowl Data Eng. 2017;29(10):2318–31.
116. Wang L, Adiga A, Chen J, Sadilek A, Venkatramanan S, Marathe M. CausalGNN: causal-based graph neural networks for spatio-temporal epidemic forecasting. In: Proceedings of the 36th AAAI conference in artificial intelligence (to appear). AAAI Press; 2022.
117. Garey MR, Johnson DS. Computers and intractability: a guide to the theory of NP-completeness. San Francisco, CA: W. H. Freeman and Co.; 1979.
118. Valiant LG. The complexity of enumeration and reliability problems. SIAM J Comput. 1979;8(3):410–21.
119. Arora S, Barak B. Computational complexity: a modern approach. New York, NY: Cambridge University Press; 2009.
120. Vadhan S. The complexity of counting in sparse, regular and planar graphs. SIAM J Comput. 2001;31(2):398–427.
121. Karp RM, Luby M. Monte-Carlo algorithms for the planar multiterminal network reliability problem. J Complex. 1985;1(1):45–64.
122. Saha S, Adiga A, Prakash BA, Vullikanti AKS. Approximation algorithms for reducing the spectral radius to control epidemic spread. In: Venkatasubramanian S, Ye J, editors. Proceedings of the 2015 SIAM international conference on data mining, Vancouver, BC, Canada, April 30—May 2, 2015. SIAM; 2015. p. 568–76.
123. Sambaturu P, Adhikari B, Prakash BA, Venkatramanan S, Vullikanti A. Designing effective and practical interventions to contain epidemics. In: Seghrouchni AEF, Sukthankar G, An

B, Yorke-Smith N, editors. Proceedings of the 19th international conference on autonomous agents and multiagent systems, AAMAS '20, Auckland, New Zealand, May 9–13, 2020. International Foundation for Autonomous Agents and Multiagent Systems; 2020. p. 1187–95.
124. Shah D, Zaman T. Detecting sources of computer viruses in networks: theory and experiment. In: SIGMETRICS 2010, proceedings of the 2010 ACM SIGMETRICS international conference on measurement and modeling of computer systems, New York, New York, USA, 14–18 June 2010. ACM; 2010. p. 203–14.
125. Wagner MM, Moore AW, Aryel RM. Handbook of biosurveillance. Elsevier; 2011.
126. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. 2020;20(5):533–4.
127. Robishaw JD, Alter SM, Solano JJ, Shih RD, DeMets DL, Maki DG, Hennekens CH. Genomic surveillance to combat COVID-19: challenges and opportunities. Lancet Microbe. 2021;2(9):e481–4.
128. The COVID-19 Genomics UK (COG-UK) Consortium. An integrated national scale SARS-CoV-2 genomic surveillance network. The Lancet Microbe. 2020;1(3):e99.
129. Shu Y, McCauley J. GISAID: global initiative on sharing all Influenza data–from vision to reality. Eur Secur. 2017;22(13):30494.

Chapter 10
Regulatory Aspects on AI and Pharmacovigilance for COVID-19

Vincenzo Salvatore

What Does Artificial Intelligence Mean According to Legal Definition?

AI is a computer science field where the computer is trained to perform tasks that normally require human intelligence.

Nevertheless, AI is an umbrella term entailing a broad meaning that covers a wide range of technologies, such as machine learning and automated operations.

In other words, AI can exist in software only (e.g., image analysis software, search engines, subject recognition systems) or can be embedded in hardware devices and interact with the physical world (e.g., robots, machines, sensor application).

The diversity of methods and applications of AI implies the still lack of a single agreed standard or definition of AI.

Pursuant to the legal definition set out in the recently proposed EU Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), 'artificial intelligence system' (AI system) means '*software that is developed with one or more of the techniques and approaches … and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with*'.

A broader non-legal notion of AI more commonly used in computer science encompasses iterative, 'learning' algorithms that utilise (big) data and high computing power to make interpretations, predictions, or decisions in an autonomous or semi-autonomous fashion that could be seen to imitate intelligent behaviour.

V. Salvatore (✉)
Department of Economics, University of Insubria, Varese, Italy
e-mail: vincenzo.salvatore@uninsubria.it

N. Lidströmer, Y. C. Eldar (eds.), *Artificial Intelligence in Covid-19*,
https://doi.org/10.1007/978-3-031-08506-2_10

AI and Health

AI is widely used to detect diseases and abnormalities through real-time pattern recognition and image processing, speeding up diagnosis and treatment.

Thanks to AI facilitated access to broader datasets increases the potential for healthcare professionals to better understand the patterns and symptoms of their patients and therefore provides better feedback, guidance and support to healthcare professionals and caregivers.

AI has also the potential to tailor treatments and medicinal products development to specific patient needs and enhance engagement with stakeholders and participants in the healthcare system.

Finally, AI systems have the potential to relieve healthcare systems, and especially medical staff, by supporting routine tasks such as patient transport and reminding patients of their medication, and to remedy challenges posed by rapidly ageing populations.

The European Union Legal Framework: A Work in Progress

The European Union approach to AI aims at ensuring that any AI improvements are based on rules that safeguard the functioning of markets and the public sector, and people's safety and fundamental rights.

It is not AI as a technology that should be regulated: the type, intensity and timing of regulatory intervention should solely depend on the type of risk incurred by the use of an AI system; underlining, in this regard, the importance of distinguishing between a minority of 'high-risk' and the vast majority of 'low-risk' AI use causes. Only the former category indeed demands legislative safeguards, businesses should self-regulate 'low-risk' technologies by choosing measures that deliver the best outcomes.

To help further define its vision for AI, the European Commission developed an AI strategy proposing measures to streamline research, as well as policy options for AI regulation, which fed into work on the AI package.

The European Commission's AI package published in April 2021 proposes new rules and actions to turn Europe into the global hub for AI and it mainly focuses on two areas: excellence in AI and trustworthy AI.

The EU AI package consists of a Communication on Fostering a European Approach to Artificial Intelligence, complemented by the Coordinated Plan with Member States: 2021 update and the proposal for an AI Regulation laying down harmonised rules for the EU (Artificial Intelligence Act).

The Commission's initiative intends to address the risks generated by specific uses of AI through a set of complementary, proportionate and flexible rules. This also feeds the ambition of Europe for playing a leading role in setting the global gold standard.

The proposed legal framework for AI introduces a clear, easy to understand approach, based on four different levels of risk: unacceptable risk, high risk, limited risk, and minimal risk.

The legislative initiative was anticipated by a White Paper on Artificial Intelligence published by the European Commission in February 2020 followed by a public consultation between February and June 2020.

In its White Paper the Commission identifies the need to establish a risk-based legal framework for AI, covering high-level ethical standards combined with appropriate liability rules and sector-specific provisions, while at the same time providing the private sector with enough flexibility, practicability, and legal certainty to develop new business models based on AI technologies.

Stakeholders in their large majority shared the European Commission's approach to the design of a regulatory framework for AI (e.g. regulation, certification and labeling) and supported the revision of the existing Product Liability directive to cover particular risks engendered by certain AI applications.

For its part, the European Parliament adopted in October 2020 three resolutions on AI covering ethics, civil liability, and intellectual property (IP). In particular, it was the legislative resolution by MEP Iban García del Blanco which urged the Commission to establish a comprehensive and future-proof European legal framework of ethical principles for the development, deployment and use of AI, robotics and related technologies—including software, algorithms and data—in the Union.

Furthermore, the European Parliament's Special Committee on Artificial Intelligence in a Digital Age (AIDA) presented its draft report in November 2021. The draft report highlights that the EU action should focus on fostering the enormous potential of AI. It was voted by the committee in March 2022, and then followed by a plenary debate and vote in May 2022.

The Proposed EU Regulation (Artificial Intelligence Act)

The April 2021 AI regulation is the first ever attempt to enact a horizontal regulation of AI.

It lays down: (a) harmonised rules for the placing on the market, the putting into service and the use of artificial intelligence systems ('AI systems') in the Union; (b) prohibitions of certain artificial intelligence practices; (c) specific requirements for high-risk AI systems and obligations for operators of such systems; (d) harmonised

transparency rules for AI systems intended to interact with natural persons, emotion recognition systems and biometric categorisation systems, and AI systems used to generate or manipulate image, audio or video content; and (e) rules on market monitoring and surveillance.

The Commission proposes to establish a technology-neutral definition of AI systems in EU law and to lay down a classification for AI systems with different requirements and obligations tailored on a 'risk-based approach'.

The legal scenario the European Commission is willing to frame focuses on the specific utilisation of AI systems and associated risks.

The risk-based approach differentiates between uses of AI that create (1) an unacceptable risk, (2) a high risk, and (3) low or minimal risk.

Some AI systems presenting 'unacceptable' risks would be prohibited. A wide range of 'high-risk' AI systems would be authorised, but subject to a set of requirements and obligations to gain access to the EU market.

Those AI systems presenting only 'limited risk' would be subject to very light transparency obligations.

The list of prohibited practices comprises all those AI systems whose use is considered unacceptable as contravening Union values, for instance by violating fundamental rights.

The prohibitions cover practices that have a significant potential to manipulate persons through subliminal techniques beyond their consciousness or exploit vulnerabilities of specific vulnerable groups such as children or persons with disabilities in order to materially distort their behaviour in a manner that is likely to cause them or another person psychological or physical harm.

Other manipulative or exploitative practices affecting adults that might be facilitated by AI systems could be covered by the existing data protection, consumer protection and digital service legislation that guarantee that natural persons are properly informed and have free choice not to be subject to profiling or other practices that might affect their behaviour.

The proposal also prohibits AI-based social scoring for general purposes done by public authorities. Finally, the use of 'real time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement is also prohibited unless certain limited exceptions apply.

As mentioned above, while generally supporting the Commission's proposal, stakeholders and experts call for a number of amendments, including revising the definition of AI systems, broadening the list of prohibited AI systems, strengthening enforcement and redress mechanisms and ensuring proper democratic oversight of the design and implementation of EU AI regulation.

The proposed legislation will apply *erga omnes*, i.e.: to providers placing on the market or putting into service AI systems in the Union, irrespective of whether those providers are established within the Union or in a third country; to users of AI systems located within the Union and providers and users of AI systems that are located in a third country, where the output produced by the system is used in the Union.

The Use of AI in Research and Developing Medicinal Products and Monitoring Their Quality, Safety and Efficacy

Pharmaceutical companies and medical devices manufacturers are constantly and increasingly challenged by the impact of new technologies in performing their respective activities.

Opportunities to apply AI occur at all stages of a medicinal product's lifecycle.

The benefits stemming from the use of AI have been highlighted across the different phases a medicinal product goes through from target validation and identification of biomarkers to annotation and analysis of clinical data in trials, to pharmacovigilance activities and clinical use optimisation.

As recently highlighted by the International Coalition of Medicines Regulatory Authorities (ICMRA), this range of applications brings with it regulatory challenges, including the transparency of the algorithms themselves and their meaning, as well as the risks of AI failures and the wider impact these would have on its uptake in pharmaceutical development and ultimately on patients' health.

In principle, AI systems appear suitable for the detection of safety signals.

Actually, it has been noted that the current signal detection and management tools have a heavy manual component that may be hard to sustain in the future, partly due to the growing data from increased ADR reporting worldwide.

The challenge in implementing AI solutions will lie in getting the balance right between AI and human oversight of signal detection and management, which may vary according to the risk of the medicinal product(s).

Regulatory guidelines for AI development and use with medicinal products should be developed in several areas, including data provenance, reliability, transparency and understandability, validity (construct, content, external etc.), development and use for pharmacovigilance purposes, real-world performance and monitoring.

In addition, there is also potential for such AI use to discover safety signals that are more difficult to detect with current methods, such as drug-drug interactions, drug-disease interactions, medication errors, secondary malignancies, changes in frequency and severity of known events, patterns of use in medications, and misuse.

Nonetheless, the application of the concerned software would require the marketing authorisation holder company to gain specialist expertise in AI, data quality and pharmacovigilance signal detection in order to best govern its use.

On the other hand, if the AI is operated by a third party, reassurance will be needed that the concerned third party does not trigger any dispute on any of the responsibilities of the sponsor and that the tool could be inspected by regulators. At this regard, it is crucial that software updates that affect the data are promptly communicated to regulators.

Of course, the company would be responsible to assess the effectiveness of updates: such as new methods, new algorithms, new tools for signal detection. This would be subject to internal audits and regulatory inspections.

The scientific or clinical validation of AI use would require a sufficient level of understandability and regulatory access to the employed algorithms and underlying datasets. Regulators may need to elaborate a risk-based approach to assessing and regulating AI, and this could be informed through exchange and collaboration in ICMRA. Legal and regulatory frameworks may need to be adapted to ensure such access options.

In addition, limits to validation and predictability may have to be identified and tolerated when, for example, the AI is to learn, adapt or evolve autonomously.

Another relevant important step forward would be related to the appointment of a Qualified Person responsible for AI/algorithm(s) oversight compliance (similar to what is required for pharmacovigilance or medical devices), clearly identifying its role and responsibilities.

The Added Value Brought Using Artificial Intelligence in Performing Pharmacovigilance Activities in General and During the COVID-19 Pandemic

The safety of medicinal products is one of the major concerns and at the same time one of the major threats once a new medicinal product has been placed onto the market.

Pharmacovigilance is generally defined as the science relating to the detection, assessment, understanding, and prevention of adverse effects and other drug-related safety problems.

Pharmacovigilance activities include the collection and assessment of safety related information to determine the benefit/risk impact of a given medicinal product, the adoption of risk management strategies to minimise risks associated with the use (and misuse) of medicinal products and the setting of provisions governing sharing of information and communication of any risk to the public and healthcare professionals.

The fight against COVID-19 has both accelerated research and development of new technologies, mainly relying on AI processes, with the aim to improve and speed up the signal detection and increased the need for industry and healthcare professionals and investigators to be able to use AI tools to improve the monitoring of current and future pandemics.

The way AI has been used as a tool to support the fight against the viral pandemic that has affected the entire world since the beginning of 2020 has been thoroughly analysed in an overview carried out by the Ad Hoc Committee on Artificial Intelligence (AHCAI) of the Council of Europe.

The use of AI tools in a pandemic or other health emergency situation has been and will have to be considered and adequately tailored in a way that will not generate excessive limitations on freedom of movement, undermine data protection principles or increase the risk of establishing excessive surveillance regimes.

The added value of pharmacovigilance compared to traditional safety monitoring activities is that AI can use sophisticated algorithms to extract useful information from the huge amount of healthcare data from the real world, dramatically improving the data collection and assessment speed and accuracy thanks to its learning and intrinsic self-correcting ability.

The accurate and timely handling the volume of adverse drug reaction generated by COVID-19 vaccines—due to the massive vaccination campaign around the world and the billions of people to whom vaccination was administered—would not have been and will not be possible without using AI tools.

AI additionally demonstrated to be a unique and indispensable tool to address the increasing demand for information adverse events of vaccines and medicines used either to treat or prevent the diseases generated by the COVID-19 pandemic and to promptly spread medical information answering to the uncountable number of live calls from the public that otherwise humans would have had to answer.

Ethical Issues: A Few Caveats

In addition to the legal issues, the use AI technologies should also not be underestimated with regard to its ethical implications.

Ethical guidelines would be welcome to introduce core values and principles such as the non-maleficence principle, the principle of respecting human dignity or the protection of the democratic process.

It is important to raise awareness on the need of a new regulatory framework for AI consisting of legal obligations and ethical principles for the development, deployment and use of artificial intelligence, robotics and related technologies which would ensure full respect of the principles set by the Charter of fundamental rights of the European Union and thereby respect human dignity, autonomy and self-determination of the individual, prevent harm, promote fairness, inclusion and transparency, eliminate biases and discrimination, including as regards minority groups, and respect and comply with the principles of limiting the negative externalities of technology used, of ensuring explainability of technologies, and of guaranteeing that the technologies are there to serve people and not replace or decide for them, with the ultimate aim of increasing every human being's well-being.

It should also be emphasized the asymmetry between those who employ AI technologies and those who interact and are subject to them; in this context, stresses that citizens' trust in AI can only be built on an ethics-by-default and ethics-by-design regulatory framework which ensures that any AI put into operation fully respects and complies with the Treaties, the Charter and secondary Union law.

Finally, it has to be noted that building on such an approach should be in line with the precautionary principle that guides Union legislation and should be at the heart of any regulatory framework for AI.

The Personal Data Protection Implications

In addition, pursuant to Article 16 of the TFEU, due attention shall be paid to the protection of individuals with regard to the processing of personal data.

The EDPB and the EDPS welcomed the risk-based approach underpinning the proposed EU regulation. However, this approach should be clarified and the concept of "risk to fundamental rights" aligned with the GDPR and the Regulation (EU) 2018/1725 (EUDPR), since aspects related to the protection of personal data come into play.

The EDPB and the EDPS agree with the Proposal when it states that the classification of an AI system as high-risk does not necessarily mean that it is lawful per se and can be deployed by the user as such.

Further requirements resulting from the EU data protection law may need to be complied with by the controller. Moreover, the compliance with legal obligations arising from Union legislation (including on personal data protection) should be a precondition to being allowed to enter the European market as CE marked product.

Considering the spread of AI systems across the single market and the likelihood of cross-border cases, there is a crucial need for a harmonized enforcement and a proper allocation of competence between national supervisory authorities. The EDPB and EDPS suggest envisaging a mechanism guaranteeing a single point of contact for individuals concerned by the legislation as well as for companies, for each AI system.

Regarding the codes of conduct, the EDPB and the EDPS consider it necessary to clarify if the protection of personal data is to be considered among "additional requirements" that can be addressed by these codes of conduct, and to ensure that the "technical specifications and solutions" do not conflict with the rules and principles of the existing EU data protection framework.

Provisional Conclusions

As we have witnessed during the COVID-19 pandemic, AI proved to be an unprecedented and unvaluable tool in monitoring the safety of medicines, unlocking a number of solutions that helped saving millions of lives.

It is undisputable that pharmacovigilance activities can make an additional great forward if driven by AI techniques.

The most important benefits of AI are definitely related to the reduction of cycle times in collecting and assessing a huge amount of safety information from the real world, managing different types of data formats and reducing the burden of case processing, and the mitigation of risk of human errors.

Nonetheless, automatic decision-making in healthcare applications may pose risks to patients and current liability frameworks do not provide sufficient legal certainty over who is accountable in the event of misinterpretation of relevant information through AI.

To realize the full potential of AI in health and achieve a widespread use of AI, maximizing its benefits, legal uncertainty, insufficient digital infrastructure and lack of AI skills will require additional attention from all the relevant stakeholders (legislator, industry, healthcare professionals) in order to be overcome.

Suggested Reading

1. European Commission. Proposal for a regulation of the European Parliament and of the Council of 21 April 2021 laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM (2021) 206). 2021a.
2. European Commission. Commission communication of 15 April 2021 on Fostering a European approach to Artificial Intelligence (COM (2021) 205). 2021b.
3. European Commission. Commission Coordinated plan on artificial intelligence 2021 review, Annex to Commission communication of 15 April 2021 on Fostering a European approach to Artificial Intelligence (COM (2021) 205). 2021c.
4. European Commission. Commission White Paper of 19 February 2020 entitled 'Artificial intelligence – a European approach to excellence and trust' (COM (2020) 0065). 2020.

Chapter 11
AI and the Clinical Immunology/ Immunoinformatics for COVID-19

Zikun Yang, Xiongye Xiao, and Paul Bogdan

Introduction

First detected in late 2019, the SARS-CoV-2 has spread all over the world for the past 2 years, causing more than five million deaths. Coronavirus is still undergoing serious mutations and the pandemic is still challenging global public health. The emerging variants of SARS-CoV-2 such as Delta and Omicron also continue to bring additional challenges to our health.

In our effort to combat the COVID-19 pandemic, biomedical science plays an important and critical role in providing proper strategies for detecting, controlling, treating, preventing virus infections, and maintaining the stability of our society. In modern days, biotechnologies are evolving in a digitalized era. In many cases the biochemical materials in the viruses can be quantified into computable and processable sequenced information, the biochemical reactions can be simulated in a computer through various computational frameworks, as well as the complex patient's biomedical symptoms can be evaluated by a computational system. In this context, Artificial Intelligence (AI) plays a vital role in helping the traditional biotechnologies onto a new stage. For instance, AI can learn and summarize experience and knowledge for extremely efficient practical judgment, as well as shade light on many new biochemical findings. AI can also help us to efficiently analyze a large number of COVID-19 datasets, which cannot be done by traditional methods.

Z. Yang
Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, SAR, China
e-mail: u3586534@connect.hku.hk

X. Xiao · P. Bogdan (✉)
Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, United States
e-mail: xiongyex@usc.edu; pbogdan@usc.edu

N. Lidströmer, Y. C. Eldar (eds.), *Artificial Intelligence in Covid-19*,
https://doi.org/10.1007/978-3-031-08506-2_11

In this section, we are going to introduce how AI helps us in the clinical immunology and immunoinformatics for COVID-19. More specifically, we show how AI-based methods can be applied to design vaccine candidates. We also demonstrate that the mutations of SARS-CoV-2 can be investigated by the computational model based on AI, which can help us design more robust vaccines against SARS-CoV-2.

Challenge for Traditional Vaccines in COVID-19

There are tens of different vaccines in use or that have passed phase III clinical trials for COVID-19 and many of them are practically widely used all over the world. Hundreds of millions of people have already embraced and received the Pfizer-BioNTech [1] or Moderna [2] vaccine, which are developed by the novel mRNA method. In China, more than 1.2 billion people were fully vaccinated with the Sinopharm's [3] or Sinovac's conventional inactivated vaccine. The viral-vector-based Janssen COVID-19 vaccine [4] and protein-subunit-based Novavax [5], are also officially authorized by the Word Health Organization (WHO) for COVID-19. All those vaccines are showing good efficacy against the SARS-CoV-2, offering people strong protections against the infections. We sincerely appreciate the biomedical scientists contributing to vaccine research and development. Their comprehensive works help us control the spread of infectious diseases and save lives all around the world.

While people benefit from those state-of-the-art vaccines in the COVID-19 pandemic, we may also recognize their shortcomings, which in turn bring severe challenges for the traditional vaccine design methodologies. Here, we raise some of the major concerns regarding the existing design strategies for COVID-19 vaccines:

Long Development and Design Period

If we recall the injection time of our first dose, those vaccines were put into practical use mostly in late 2020, while most people were fully vaccinated in 2021. By that time, the SARS-CoV-2 had already spread all around the world for more than a year and caused hundreds of millions of cases with several million deaths. It is impossible for the traditional vaccine design approaches to provide vaccine solutions while an infectious disease starts to emerge. Hence, the virus may already bring severe devastations before the vaccine solution starts to be used. For the most sophisticated conventional inactivated vaccine [6] strategies, the vaccine solution is developed based on growing pathogens that we first grow the virus then kill them and inject the inactivated viruses into human bodies. Before COVID-19, many other vaccines were developed in this way and were extremely successful even if the development was slow. However, growing the pathogens can take several months or even a year, not to mention the time to do clinical trials to test the efficacy, which is extremely

time-consuming. Unlike the traditional vaccines, mRNA vaccine [7] can skip the process of growing the virus, because researchers only need to focus and select their desired RNA fragments coding the vital proteins after the genome is sequenced out of the viral sequence and find a proper strategy to deliver the mRNA into the human bodies, based on various biomedical experiments. The mRNA techniques make it possible to start the clinical trials within 2–3 months, which is much faster than the inactivated vaccines. However, this time period is not short enough, which can still lead to numerous mutations and millions of cases. Biological and chemical experiments take time, and the traditional strategies have to wait for the experimental results. Consequently, it is natural to ask whether we can speed up this process and design vaccine candidates within seconds from the moment of detecting the emergence of a viral threat.

Difficulties in Knowing and Optimizing the Efficacy and Side Effects

The inactivated vaccine makes use of the whole virus to trigger the immune response, which is a mature and reliable way to provide basic antibodies. We have different inactivated vaccines for COVID-19 developed by different companies, and they might vary in the production procedure and quality. However, the vaccine solution injected into the human bodies is always the killed whole virus, which means that from the moment we start cultivating the pathogens, the final vaccine is already decided. It is difficult to know by traditional methods in advance how well the inactivated virus will work after the several-month growth especially when we do not know how the virus can evolve, what clades can emerge or what mutations are more likely to either exacerbate or diminish its virulence. It is also hard to predict what side effects it might have. In a word, for inactivated vaccines, researchers use several months to grow a certain virus, then conduct biological experiments to validate the efficacy during the clinical trials. If there exist any shortages after the pathogens are successfully grown, we have to live with that. Since the original vaccine is already decided at the very beginning, we cannot truly optimize the side effects and improve the efficacy after the clinical trials. If there are some serious issues during the validation, the previous months of hard and comprehensive work might go as wasted.

Similar things can happen to the mRNA vaccine, too. The selected mRNA fragment is still coding a certain part of the original virus proteins, which is large and fixed. For example, the Pfizer-BioNTech's mRNA is the genome that codes the spike protein of the SARS-CoV-2. We cannot predict how it works before clinical trials in traditional ways. In addition, we cannot change any components of the original mRNA, so we live with these shortcomings during the experimental validation. Consequently, those traditional vaccine design approaches are based on the logic that, we use the exact same thing in the original virus, and we hope it works when we inject them into the human bodies. The protein subunit vaccine is a way to

discard some original parts that might cause side effects, while using the vital parts to build the immune defense line. However, targeting those important protein fragments in traditional ways is extremely time-consuming.

Uncertainties with the Development and Other Costs During Production, Storage, and Transportation

Assuming we are lucky that the killed pathogens or the selected mRNA fragments work, and can be put into use, then we have to provide the required environmental conditions throughout the production of the injections. The conditions are decided by the vaccine, and since the original virus is fixed, it is very hard to let the produced vaccines satisfy certain environmental conditions. The mRNA vaccines usually need to be kept in an ultra-cold environment, which leads to high storage and transportation costs, even though the mRNA vaccines may be cheap and easy to produce. For inactivated vaccines, although the storage conditions for the killed viruses are not strict, the vaccine itself will be very expensive, due to the long-time production period. In traditional vaccine design, those situations need to be tested by a large number of experiments and are hard to be predicted before production.

Hard to Tackle Unknown and Emerging Mutations of Viruses

As we mention before, the developing period for traditional vaccine design is very long, hence, we cannot rapidly develop a new vaccine solution if the existing one cannot deal with the emerging mutations of a virus. Moreover, when the vaccine is developed based on the original virus, it is common that the virus might already mutate to a new generation. Hence, to tackle such virus mutations, we should either develop a more efficient vaccine design method that can rapidly update the vaccine solutions when mutations occur or investigate the evolution of the virus in order to predict the possible mutations in advance and design the vaccine while taking these potential mutations into account. The traditional design approaches do not satisfy those two points.

Existing AI Techniques Help the Traditional Vaccine Development in COVID-19

AI Makes the Practical Experimental Results Computational

Deep learning [8] is one of the hottest topics across many different scientific fields. Generally speaking, in deep learning, the neural network can be seen as a complex ‘formula’ capable of bridging or finding a mapping from the input information to

the desired output information. An easy-to-understand example is that we input the value of every pixel of an image to the neural network, and the output value can tell us what this image is about if the parameters in the 'formula' are well-adjusted. In deep learning strategies, researchers frequently use a large amount of data in which the input and output are rightly bridged to adjust the network's parameters and learn this "formula', which is called the training process.

In recent years, researchers found that it is possible and more and more robust to use neural networks to process biomedical information, so that it can tell us about some important biological features from the heterogeneous datasets automatically. For example, the ToxinPred [9] is a popular computational tool based on AI to compute whether the input protein sequence is toxic or not. Researchers use a large dataset with many toxic and non-toxic protein sequences, which are created and decided by chemical experiments, to train the neural network architecture. In this scenario, we can exploit this tool to simultaneously judge a large number of protein sequences and identify whether they are toxic or not in less than a second, without the need to conduct any biomedical experiments. The idea is to use the neural network as a computational strategy to replace the redundant and time-consuming biomedical analyses and experiments so that we can have better accuracy and higher speed. Similar methodologies have been used in computing and predicting the different features in protein analysis or genome analysis, the interactions between different molecules, the structural formation of big molecules, and many other biological or chemical aspects.

AI-Based Computational Tools Can Help the Traditional Vaccine Design

Here, we introduce some typical cases in which AI can help with the traditional vaccine design process.

1. Protein analysis. While the inactivated vaccines and some mRNA vaccines are using the whole or part of the whole original viruses, some other vaccines like subunit protein vaccines are selecting some fragmental parts of the viruses and constructing them together as the final vaccines. Some mRNA vaccines are also selecting out the wanted partial proteins to trace back to the mRNA fragments. In such a vaccine design process, researchers might want to know which parts of the virus's protein are the fragments they want. A very famous case is the Hepatitis B vaccine [10], in which the immunizing fragment was discovered as an antigen from the human blood, and of course, purifying the desired antigens from human blood is not easy. However, nowadays, the neural network architectures trained on known human antigens can help researchers to discover potential antigens directly from the virus protein (i.e., Vaxijen). We also have other tools which can help researchers to know the stability, allergenicity, structures, suitable environments in advance. When the researchers want to find the protein

subunits to design a vaccine candidate, they can scale down their targets by exploiting the AI-based tools with good accuracy, to skip some time-consuming experiments. Also, for some preliminary vaccine models, the researchers can use those AI tools to check the protein features before the trials to discard some bad ones.

2. Immune simulation. Many computational tools can now simulate how a vaccine will interact with the human immune system, and many of them take advantage of the AI technology to improve the accuracy (i.e., DeepImmuno [11]). The researchers can now use those AI tools to simulate the immune efficacy of their designed vaccines before clinical trials.
3. Structure prediction. For any designed vaccine, its 3-dimensional structure is important, because it might affect the stability, whether some important parts are exposed on the surface, and many other things. For example, in a subunit protein vaccine, we do not want the antigens to be buried. Hence, researchers can use structure prediction approaches to identify and then modify the vaccine structure, before going to real production, to avoid the waste of time and resources. Neural networks can help with this modeling effort in many perspectives, for example, by computing the distance between different molecules.

There are many other perspectives in which the deep learning techniques can help the vaccine design. The traditional vaccines are still safe and easy-to-accept, and they are relatively more stable and reliable. In the meanwhile, AI plays a role currently to shorten the time we need to evaluate different features in the vaccine components, and even for vaccine discovery. Those mature vaccine technologies can closely work with the modern AI.

AI-Based In Silico Vaccine Design

With more and more AI tools for the evaluation of biochemical features of genome or protein sequences, developing the vaccines fully by computational methods is a very popular topic. The idea is to design the vaccine formulation based on the virus protein sequence information through computational approaches so that we can have several good vaccine candidates for clinical trials at a very fast speed, which is called *in silico multi-epitope vaccine design*. Epitopes are small proteins of the virus that can be recognized by our immune system and trigger the immune response. Multi-epitope vaccines are constructed by several vaccine subunits and each of the subunits represents a virus protein particle-containing multiple epitopes (or simply being an epitope) and having good protein functions (e.g., high antigenicity, low allergenicity, good structures), which are promising to induce a robust immune response against potential viral infections. This new vaccine's design process is fast, its features and performance can be detailed and evaluated before biomedical experiments, the vaccine components can be optimized with purpose so that many problems with the traditional vaccine design can be addressed.

Figure 11.1a shows an overview of the *in silico* vaccine design process. The final vaccine candidate contains multiple protein fragments of the original virus, and we aim to select those fragments out, construct the vaccine with those fragments, and evaluate the final vaccine performance by AI-based computational methods. In what follows, we introduce the vital steps and the developed AI technologies:

1. Predicting the epitopes on the viral protein. Identifying the epitopes is the most important aspect for computational vaccine design, which is the essential basis for selecting the protein fragments. Different types of epitopes (i.e., B-cell epitope, CTL epitope, HTL epitope) that can be recognized by human immune cells can also be known by experimental experience and summarized into different datasets. Researchers train neural networks to classify the input protein sequences into two categories of epitope and non-epitope, based on the known epitope datasets. In this step, we input different overlapping viral protein fragments with different lengths into these AI-based epitope prediction tools to identify all the potential epitopes in the viral protein. There are many good prediction tools,

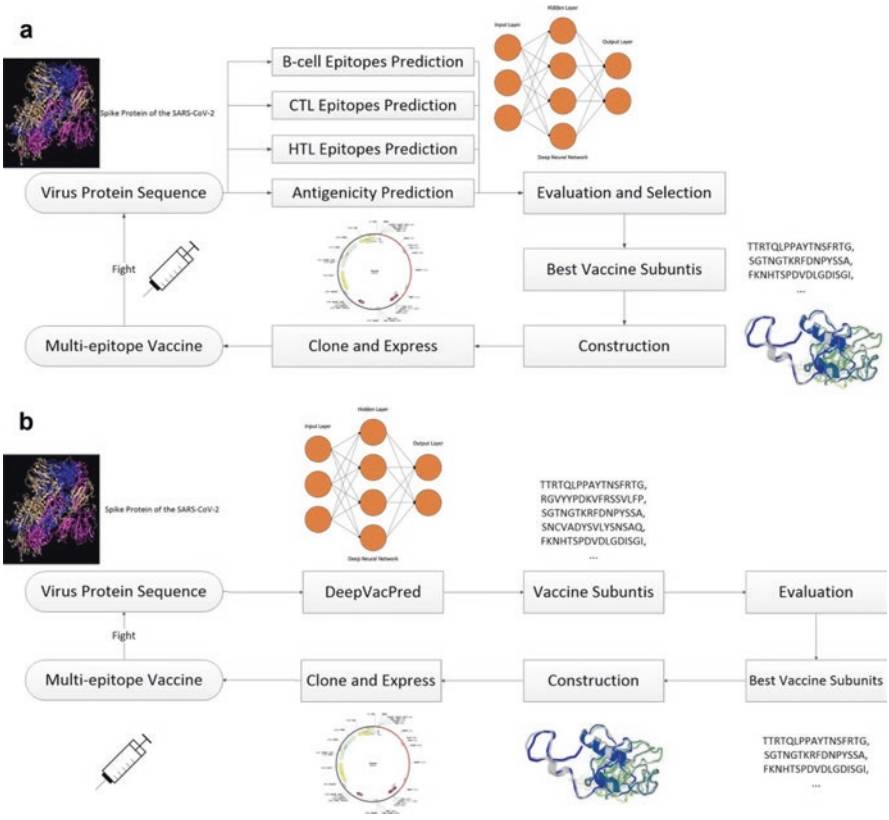


Fig. 11.1 (**a**, **b**) Schematic diagram of in silico vaccine design process. (Reprinted with permission from [12])

such as BepiPred [13] for B-cell epitope prediction, NetMHCpan [14] for CTL epitope prediction, NetMHCIIpan [15] for HTL epitope prediction. The deep learning or machine learning models in those AI tools can mostly achieve a very good accuracy from the computational validation, hence, the prediction performance largely depends on how many existing epitopes the dataset covers and how well the dataset is designed. Consequently, as biomedical researchers discover more and more human epitopes, AI prediction will be more and more reliable.

2. Selecting good viral protein fragments. The first thing is to target some of the good vital protein fragments with the desired length. We may select the protein fragments with multiple epitopes and have good epitope prediction scores, which is usually conducted by manual comprehensive evaluations based on the epitope prediction results. To further scale down to the final protein targets, we use different protein analysis tools on the selected candidates to check the protein features. Many existing AI-based tools can help us with this analysis, including Vaxijen [16] for predicting protein sequence antigenicity, AllergenPro [17] for predicting protein allergenicity, Solpro [18] for predicting the solubility, ExPASy ProtParam [19] for physicochemical properties analysis, etc. Based on the protein analysis results, we may select our final protein fragments with good protein features as the vaccine subunits, then construct them with suitable linkers into our final vaccine.
3. Structure predictions and other evaluations. Before the vaccine is finalized for clinical trials, we should check its structural features and the overall immune-related performance. The structure predictions are usually supported by mathematical models and deep learning methods, and popular tools include PSIPRED [20] for secondary structure prediction, RaptorX [21] for 3D structure modeling, GalaxyRefine [22] for 3D structure refinement, ProSA-web [23] for 3D structure validation, etc. We can further refine the constructed sequence based on the prediction results. For other aspects of vaccine performance, we have iMOD to check the molecular dynamics, ClusPro for molecular docking, C-ImmSim for immune simulation, etc.

After the above process, we shall get a protein sequence which contains multiple good epitopes and has good biochemical features and good physical structures to be considered as a vaccine candidate for further clinical trials. This type of vaccine addresses the vital parts of the virus, discards the parts which might have side effects, and makes sure of the theoretical efficacy before clinical trials. Also, the vaccine can be further optimized based on later clinical trials. Most importantly, the whole design process is much faster than traditional vaccines (e.g., for one researcher works alone, he/she might take days or a few weeks to finalize several good computational vaccine candidates). This means that we can easily construct another one if the clinical trial is not good or mutation happens. In the COVID-19 cases, many computational vaccines are designed in early 2020.

Our Recently Proposed DeepVacPred Vaccine Design Framework

As we mentioned in the previous section, to date, the researchers have developed many AI-based tools, which contain various deep neural networks (DNNs) trained on different datasets, realizing the prediction of whether an input sequence can be a certain kind of epitope or can have certain protein functions. In summary, supported by these tools, the *in silico* vaccine design process can be seen as performing different AI-based epitope predictions, evaluations and protein function analyses on every one of all the overlapping sequences from every virus protein fragment, then selecting good virus protein particles with a high number and quality of epitopes and good protein functions. Those AI-based tools do make the computational vaccine a relatively fast way to provide good vaccine solutions. However, the whole process is not efficient enough. In this process (see Fig. 11.2), most of the time will be spent on the comprehensive analyses of the prediction results from different AI frameworks, which needs to be done manually, it is slow and inaccurate, as well as it still requires days or weeks to design vaccine candidates.

In a recent publication [12], we proposed the prototype of DeepVacPred vaccine design framework. Figure 11.1b shows the efficient vaccine design process based on this framework. The idea is to use a one-time deep neural network (DNN) computation to replace the time-consuming comprehensive evaluation. In this framework, the direct prediction of the vaccine subunits from the virus can be realized by supervised learning on a subtly designed dataset in which all the positive samples are potential vaccine subunits (See Fig. 11.3). For the positive samples, we first create millions or even billions of subunit candidates by bridging different B-cell epitopes

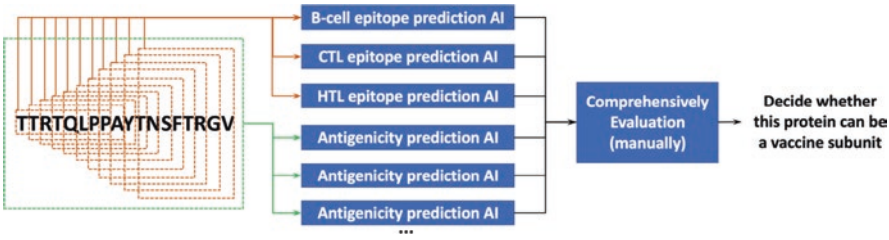


Fig. 11.2 Existing *in silico* vaccine design methodology and limitations: Current *in silico* vaccine design strategies evaluate whether a protein sequence can be a vaccine subunit by exploiting epitope prediction tools and checking on all of its overlapping fragments. Next, it evaluates the epitope scores. Consequently, when using them, we need to check the different protein functions of the whole sequence, which is very time-consuming. However, this process represents only one of the more than tens of thousands of protein fragments in a virus protein. The whole process requires a large amount of work, it wastes many computational resources, and the manual evaluation is highly inaccurate

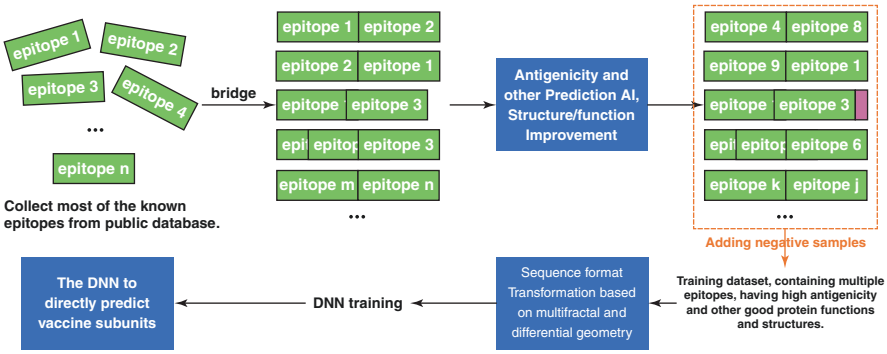


Fig. 11.3 The approach to building a DNN to directly predict vaccine subunits from the viral protein

and T-cell epitopes. We then use the required protein analysis AI tools in the common in silico vaccine design process to evaluate all the bridged candidates and to select the ones which have good protein features. Since the positive samples contain multiple epitopes as well as have good protein features, trained by this dataset, the DNN can directly identify the desired vaccine subunits, without the need to evaluate the AI outputs.

Since the speed of the DNN-based investigation is much faster and more accurate than the manually comprehensive evaluation, the current framework has been proved to largely increase the vaccine design efficiency and efficacy, compared with current computational solutions. In the DeepVacPred framework, we directly achieve a small number of vaccine subunits in less than 1 s, and the whole process can be done within an hour. Compared with the current *in silico* vaccine design approaches, we make the following contributions:

1. If any viral triggered pandemics like the COVID-19 occurs in the future, the DeepVacPred technology can help develop high-quality vaccines with high speed towards providing vaccine candidates to the clinical trials and medical therapies.
2. The DeepVacPred technology is extremely useful against the viral mutations during a pandemic event. Several COVID-19 vaccines cannot fight the recently mutated SARS-CoV-2 strains. With traditional vaccine design solutions, we might need another year to create robust vaccines for the newly mutated virus. With DeepVacPred technology, we can rapidly develop new vaccine candidates against any emerging viral mutations. The proposed AI techniques in the DeepVacPred framework bring faster speed in the vaccine design process. In addition the DeepVacPred has the potential to bring the computational vaccine design process to a much faster and more efficient level.

Scientists can easily use our framework to select the vaccine subunits to design the vaccine. For instance, the input to DeepVacPred framework is a quantification expression of the amino acids sequence (around 30 aa in length). We use Z-descriptors

and ACC transformation to transfer the amino acid sequence into a fixed 45-number sequence. The output is the 2-class positive or negative classification.

Artificial Intelligence for Investigating Viral Evolution and Mutations

An Algorithmic Information Theoretic Approach to Discover the State Machine Generator Governing the Viral Sequence Structure and Enabling AI Strategies for Viral Mutation Prediction

To efficiently deal with lots of genomic datasets of SARS-CoV-2 variants, an efficient mathematical model combining concepts from algorithmic information theory and AI has been proposed by using a state machine [24]. More precisely, the proposed approach consists of two strategies: (1) To decode and capture the hidden long-range dependencies among the distant nucleotides in a viral RNA sequence, as well as to avoid the need for alignment or overcome noise related challenges, an algorithmic information theory inspired approach learns the state machine generator governing a viral sequence and represents the genomic sequence into fixed-size transition probability matrix. The state machine generator will represent contiguous nucleotide sequences as state probabilities and encode their short- and long-range dependencies through state-to-state transition probabilities. (2) Based on this state machine generator representation of viral sequences, a data-driven AI model can expeditiously analyse the data and identify nucleotide regions of likely mutations. From a computational perspective, the state machine generators help us compress complex genomic information which further allows the AI-based model to make efficient statistical comparisons amongst large genomic datasets to detect mutation regions, characterize the temporal and spatial evolution of SARS-CoV-2 in a continuous manner, as well as to aid the alignment technique for fast discovery of mutations.

The genomic sequence is mapped onto a generator modeled using the state machine. The state machine consists of predefined states (i.e., contiguous groups of k nucleotides) and the transition probabilities between these states. The state-to-state transition probability encodes important information about predicting the next b nucleotides from previous k contiguous nucleotides in a genomic sequence. Consequently, a Markov chain consisting of 4^k states transitioning into 4^b states is developed for each viral sequence. Figure 11.4a represents the transition $acg \rightarrow t$ with the generator $k = 3$, $b = 1$, and Fig. 11.4b shows the transition $acgt \rightarrow cg$ with the generator $k = 4$, $b = 2$. After traversing the genomic sequence, the generator encodes the information needed to predict the next b nucleotides from previous k nucleotides. There are 4^{k+b} transitions between the states and the empirical transition probability matrix T is estimated from the sequences using the add-β estimator:

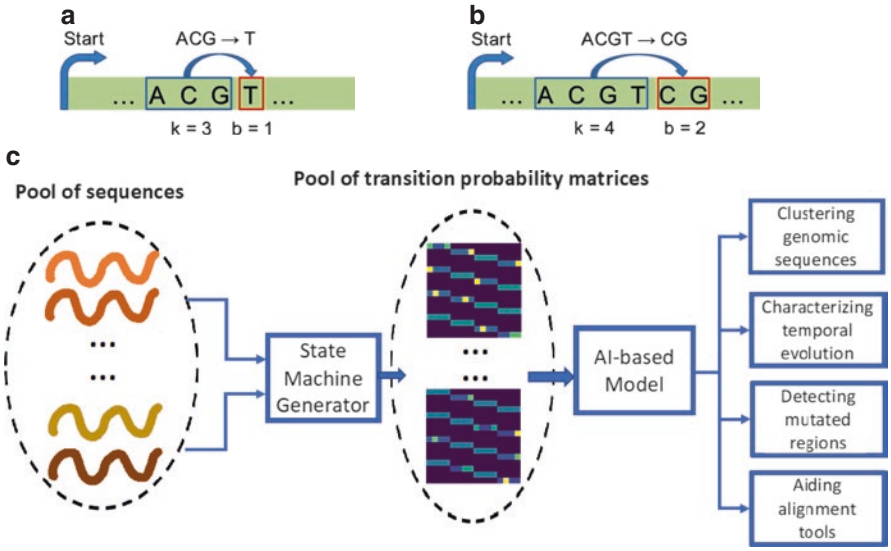


Fig. 11.4 (**a**–**c**) The architecture of the state machine and the flow diagram of the proposed method. (Reprinted with permission from [24])

$$\begin{cases} T_{ij} = \dfrac{N_{ij} + \beta}{N_i + 64\beta}, j\, is\, a\, neighbor\ of\ i. \\ 0, otherwise. \end{cases}$$

Here, $i \in \{1, 2, \ldots 4^k\}$, and $j \in \{1, 2, \ldots 4^b\}$. N_{ij} denotes the number of occurrences of the transitions from the i-th k-mers to j-th b-mers. N_i denotes the number of occurrences of the i-th k-mers in the viral sequence. Therefore, the state machine generator is independent of the sequence length and encodes the generation of the genomic sequence. In addition, the transition probabilities capture important context-sensitive characteristics within the viral sequence.

Based on the state machine generator, we build an AI-based data-driven framework to analyze a large amount of genomic data from the generator, the flow diagram of the approach is shown in Fig. 11.4c. The transition probability matrix can be converted into a 4^{k+b} dimensional vector as the feature vector for building AI-based classifiers. In the AI-based model, we have two stages. In the first, we perform feature extraction by using the XGBoost [25, 26] model with maximum depth = 1. In the second stage, we use the top ten features identified in XGBoost model to build a logistic regression model for the classification task. Fourfold cross-validation is performed in both stages. When building these AI-based classifiers, the number of training samples for each class is equal to each other to account for the data imbalance.

Next, we discuss the efficacy of the proposed method based on AI tools for different applications including (1) characterizing the temporal evolution of SARS-CoV-2 in a continuous manner, and (2) performing mutation region detection to gain insights into the mutation activity within the genomes.

Characterizing the Temporal Evolution of SARS-CoV-2 in a Continuous Manner

The temporal continuous evolution of SARS-CoV-2 can be characterized by quantifying the randomness of transient mutations. However, the clade/variant/strain is defined using stable mutations. By implementing our proposed AI-based approach, we perform the intra-clade and intra-variant comparisons for different continuous months within SARS-CoV-2. We also show that our approach can efficiently quantify continuous evolution within SARS-CoV-2. In Fig. 11.5, we perform the intra-clade comparisons of SARS-CoV-2 sequences of GISAID clades L and G over different continuous months. More precisely, let P and Q represent the set of sequences categorized as clade L in months p and q, respectively. Through the state machine generator, we can obtain the transition probability matrices using the sequences for months p and q and then we build the AI-based classifier as mentioned above to compare the sequences in different months p and q using the

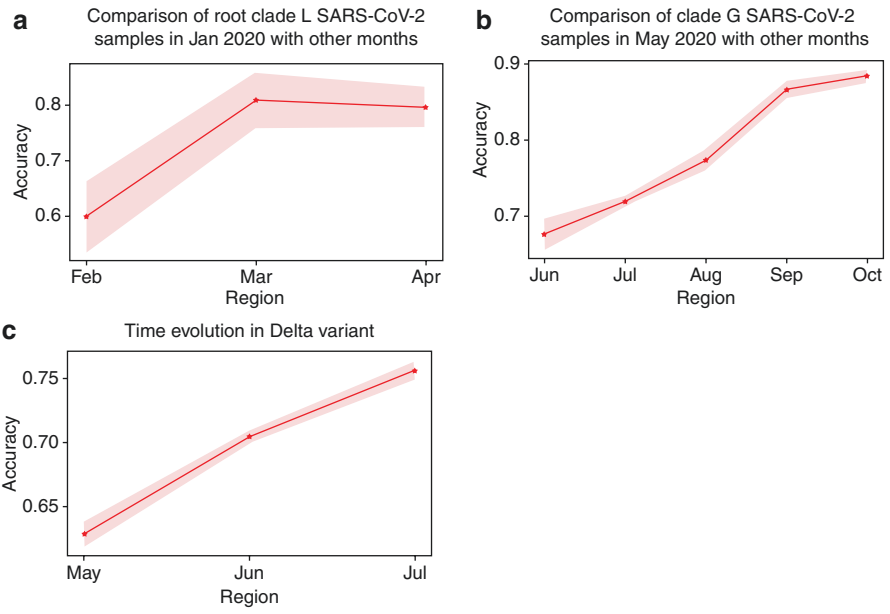


Fig. 11.5 (**a**–**c**) Continuous evolution of SARS-CoV-2. (Reprinted with permission from [24])

transition probability matrices as the features. The pairwise validation accuracy of the AI-based classifier is considered as the measure of the distance between sequences compared over different months.

In what follows, we consider the state machine generator with $k = 4$, $b = 1$ to transform the sequences into specific transition probability matrices. Figure 11.5a shows the validation accuracy of the comparison when 209 clade L samples from January 2020 are compared with clade L samples from February 2020 (209 samples), March 2020 (1739 samples) and April 2020 (788 samples), respectively. Figure 11.5b shows the validation accuracy of the comparison when 2433 clade G samples from May 2020 are compared with clade G samples from June 2020 (2309 samples), July 2020 (1465 samples), August 2020 (1605 samples), September 2020 (2038 samples) and October 2020 (1818 samples), respectively.

Figure 11.5c shows the validation accuracy of the comparison when 5699 Delta variant samples from April 2021 are compared with Delta variant samples from May 2021 (4163 samples), June 2021 (5092 samples) and July 2021 (5284 samples), respectively. The validation accuracies show the increasing trend revealing the power of the approach to characterize continuous evolution of SARS-CoV-2 within the same clade without any need for sequence alignment.

Detecting Regions Within Viral Sequences Likely to Exhibit Mutations

By using the proposed AI-based framework to estimate the distance between the regions, our approach can be used to efficiently detect the regions in the viral sequence with mutation activity. More precisely, given two sets of sequences *Set* 1 and *Set* 2, we transfer the fixed regions of all these sequences into the state-to-state transition probability matrices through the state machine generator approach, and the AI-based classifier model is used to compare these state-to-state transition probability matrices from the two sets. For a given region, the high validation accuracy implies that the region has mutations differentiating the sequences in the two sets.

We apply this approach on SARS-CoV-2 genomic datasets to detect the regions with mutations in different clades. We select the SARS-CoV-2 sequences in four clades—19A, 20A, 20B and 20C. Each SARS-CoV-2 sequence is divided into 30 non-overlapping regions of length 1000 each. The state-to-state transition probability matrices of each region are obtained using the generator with $k = 4$, $b = 1$. Then, we use the AI-based classifier models for comparing different clades in the 30 regions. The regions with mutations should be detected by the classifiers showing high validation accuracy. Figure 11.6a shows the results for comparing clade 19 A (558 samples) and 20A (605 samples). The regions with mutations can be accurately detected. We also perform the approach for comparing 20A clade with clades 20B (561 samples) and 20C (232 samples) and the results are shown in Fig. 11.6b, c. Therefore, our approach can efficiently process a large number of genomic sequences to detect the regions with mutations without any need of sequence alignment.

The region length l can be considered as a parameter, and the proposed approach can be used to detect regions with mutations differentiating the two groups of sequences. The method can be also used to aid alignment techniques as it can identify regions with mutations.

The state machine generator also provides new strategies to transform genomic sequences into new representations which encode short- and long-range dependencies, reveal the distance between the sequences, and preserve the contextual information. The proposed approach can be viewed as an interpretable tool to apply computational analysis on a large number of genomic sequences.

The higher-order interactions between the k-mers could be computationally analyzed to characterize the genomic sequences, which can help us to investigate sequences at the genetic level. As shown in Fig. 11.6, the approach can be used to compare different clades and detect the regions with highly likely mutations. The approach can be also applied to characterize the continuous evolution within a clade as shown in Fig. 11.5. Moreover, the approach can infer the local statistics of genomic sequences, which help us to identify actively mutating regions in the genomic sequence. The efficient performance of the state machine generator in dealing with a large number of genomic sequences to capture virus evolution and detect mutations allows us to detect the virus mutations efficiently and even predict the direction of virus mutations, thereby helping to design and develop more robust vaccines.

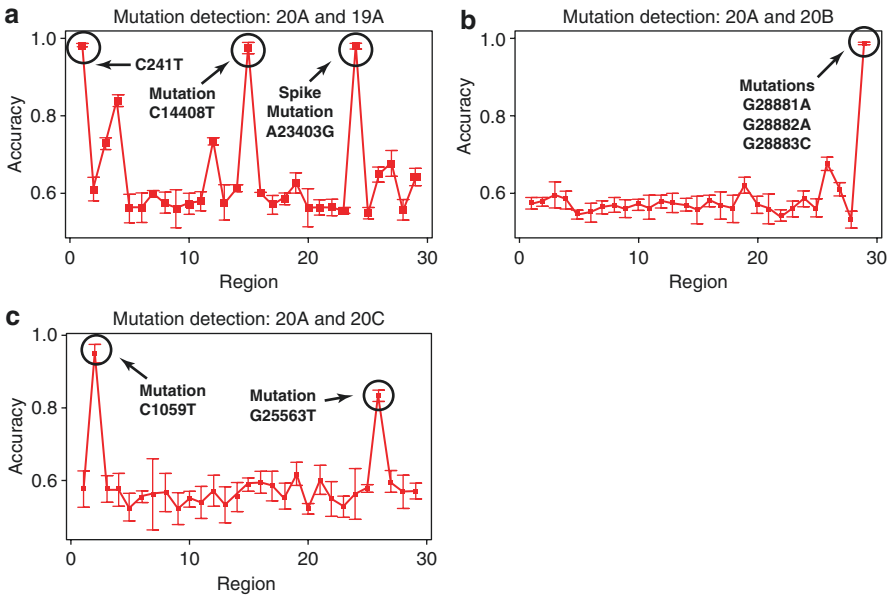


Fig. 11.6 (**a**–**c**) Mutation region detection in SARS-CoV-2 clades. (Reprinted with permission from [24])

Summary

The vaccine is a vital strategy against infectious diseases. Traditional vaccine design methods (e.g., inactivated, mRNA) are safe, reliable, and recognized by our governmental agencies, in which the existing AI tools can accelerate some of the experimental steps which require a long time in using usual strategies. While AI becomes more and more sophisticated and covers more and more biological analysis perspectives, now we can design vaccine candidates with numerous features following completely computational ways. In the COVID-19 case, it is very important to address the vaccine design speed, and many in silico vaccine candidates are designed for it, validating the strong values of this approach. Additionally, our recent work on an AI-based automatic vaccine design framework, DeepVacPred, brings vaccine design efficiency to a new level. The deep learning model exhibits very high classification accuracy; hence, the future research should be more focused on the development of the vaccine subunit dataset (i.e., keep on updating the newest epitopes to the database or finding more complex combinations consisting of more numbers of different overlapping epitopes). We shall also consider more aspects of the vaccine subunit metrics.

AI-based approach utilizes the information decoded through the state machine generator to analyze the large amount of the COVID-19 data efficiently. The new AI-based method provides not only an alternative way to statistically compare the SARS-CoV-2 sequences without any need of alignment, but also opens up new practical and analytical directions for scientific inquiry. The state machine approach generates the data for the AI-based model, then this model can analyze the information hidden in the genomic sequences, which can help us to comprehensively analyze the COVID-19 data including capturing the evolution of the SARS-CoV-2 and detecting the mutations. The investigation of the mutations of the virus has the potential to guide the development of more robust vaccines. In order to improve the AI-based computational model to predict the mutations accurately, specific models should be trained for different regions of the gene sequences and more powerful AI models such as transformer [27] and neural operator [28, 29] should be used to analyze the data.

In conclusion, AI can help with some biological redundancies in vaccine development, while on the other hand, the vaccines developed by computational methods also require robust biotechnologies to be constructed and realized into practical use. We hope that researchers in the computer science, computational biology, system biology and biomedical engineering fields can collaborate to enhance these AI tools and help with future vaccine development for emerging viral threats. The future work can focus on how to use the AI-based computational model to predict the possible mutations and use them to design and prepare the future vaccine through the DeepVacPred framework. The flow diagram of the comprehensive framework combining these strategies is shown in Fig. 11.7 and left for future work.

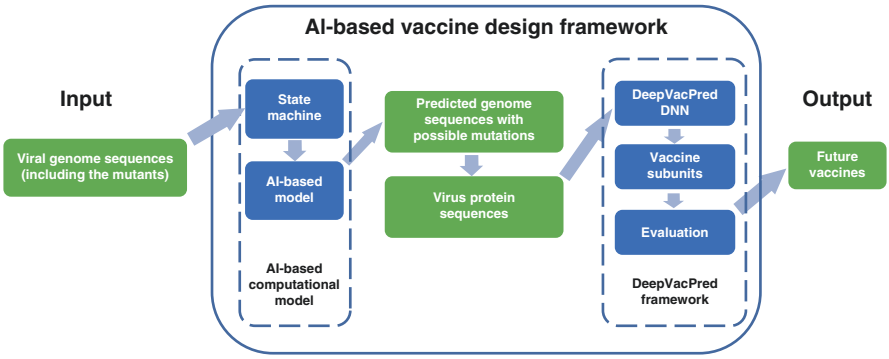


Fig. 11.7 The flow diagram of the AI-based vaccine design framework

Acknowledgements The authors gratefully acknowledge the copyright permission provided by Nature Scientific Reports: RightsLink1 (Fig. 11.1), RightsLink2 (Figs. 11.4, 11.5, and 11.6).

The authors gratefully acknowledge the support by the National Science Foundation Career award under Grant number CPS/CNS-1453860, the NSF awards under Grant numbers CCF-1837131, MCB-1936775, CNS-1932620, CMMI-1936624, the U.S. Army Research Office (ARO) under Grant No. W911NF-17-1-0076, the Okawa Foundation research award, the Defense Advanced Research Projects Agency (DARPA) Young Faculty Award and DARPA Director's Fellowship Award under Grant No. N66001-17-1-4044, a 2021 USC Stevens Center Technology Advancement Grant (TAG) award, an Intel faculty award and a Northrop Grumman grant. The views, opinions, and/or findings contained in this article are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied by the Defense Advanced Research Projects Agency, the Department of Defense or the National Science Foundation.

References

1. Jackson L, Anderson E, Rouphael N, et al. An mRNA vaccine against SARS-CoV-2 – preliminary report. N Engl J Med. 2020;383:1920–31. https://doi.org/10.1056/NEJMoa2022483.
2. Hacisuleyman E, Hale C, Saito Y, et al. Vaccine breakthrough infections with SARS-CoV-2 variants. N Engl J Med. 2021;384:2212–8. https://doi.org/10.1056/NEJMoa2105000.
3. Wang H, Zhang Y, Huang B, et al. Development of an inactivated vaccine candidate, BBIBP-CorV, with potent protection against SARS-COV-2. Cell. 2021;182:713–21. https://doi.org/10.1016/j.cell.2020.6.008.
4. Oliver SE, Gargano JW, Scobie H, et al. The advisory committee on immunization practices' interim recommendation for use of Janssen COVID-19 vaccine – United States, 2021. MMWR Morb Mortal Wkly Rep. 2021;70:329–32. https://doi.org/10.15585/mmwr.mm7009e4.
5. Chen WH, Strych U, Hotez PJ, et al. The SARS-CoV-2 vaccine pipeline: an overview. Curr Trop Med Rep. 2020;7:61–4. https://doi.org/10.1007/s40475-020-00201-6.
6. Kim HW, Canchola J, Brandt C, et al. Respiratory syncytial virus disease in infants despite prior administration of antigen inactivated vaccine. Am J Epidemiol. 1969;89:422–34. https://doi.org/10.1093/oxfordjournals.aje.a.120955.
7. Schlake T, Thess A, Fotin-Mleczek M, et al. Developing mRNA-vaccine technologies. RNA Biol. 2012;9:1319–30. https://doi.org/10.4161/rna.22269.

8. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–44. https://doi.org/10.1038/nature14539.
9. Gupta S, Kapoor P, Chaudhary K, et al. In silico approach for predicting toxicity of peptides and proteins. PLoS One. 2013;8:e73957. https://doi.org/10.1371/journal.pone.0073957.
10. McAleer W, Buynak E, Maigetter R, et al. Human hepatitis B vaccine from recombinant yeast. Nature. 1984;307:178–80. https://doi.org/10.1038/307178a0.
11. Li G, Lyer B, Prasath VBS, et al. DeepImmuno: deep learning-empowered prediction and generation of immunogenic peptides for T-cell immunity. Brief Bioinform. 2021;22:bbab160. https://doi.org/10.1093/bib/bbab160.
12. Yang Z, Bogdan P, Nazarian S. An in silico deep learning approach to multi-epitope vaccine design: a SARS-CoV-2 case study. Nat Sci Rep. 2021;11:3238. https://doi.org/10.1038/s41598-021-81749-9.
13. Jespersen M, Peters B, Nielsen M, et al. BepiPred-2.0: improving-sequence-based B-cell epitope prediction using conformational epitopes. Nucl Acid Res. 2017;45:W24–9. https://doi.org/10.1093/nar/gkx46.
14. Jurtz V, Paul S, Andreatta M, et al. NetMHCpan-4.0: improved peptide-MHC class I interaction predictions intergrading eluted ligand and peptide binding affinity data. J Immunol. 2017;199:3360–8. https://doi.org/10.4049/jimmunol.1700893.
15. Reynisson B, Alvarez B, Paul S, et al. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res. 2020;48:W449–54. https://doi.org/10.1093/nar/gkaa379.
16. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumor antigens and subunit vaccines. BMC Bioinformatics. 2007;8:4. https://doi.org/10.1186/1471-2105-8-4.
17. Kim CK, Seol YJ, Lee DJ, et al. AllergenPro: an integrated database for allergenicity analysis and prediction. Bioinformation. 2014;10:378–80. https://doi.org/10.6026/97320630010378.
18. Magnan C, Randall A, Baldi P. SOLPro: accurate sequence-based prediction of protein solubility. Bioinformatics. 2009;25:2200–7. https://doi.org/10.1093/bioinformatics/btp386.
19. Gasteiger E, et al. Protein identification and analysis tools on the ExPASy server. In: The proteomics protocols handbook. Totowa, NJ: Humana Press; 2005. https://doi.org/10.1385/1-59259-890-0:571.
20. McGuffin L, Bryson K, Jones D. The PSIPRED protein structure prediction server. Bioinformatics. 2000;16:404–5. https://doi.org/10.1093/bioinformatics/16.4.404.
21. Källberg M, Wang H, Wang S, et al. Template-based protein structure modeling using the RaptorX web server. Nat Protoc. 2012;7:1511–22. https://doi.org/10.1038/nprot.2012.085.
22. Heo L, Park H, Seok C. GalaxyRefine: protein structure refinement driven by side-chain repacking. Nucleic Acids Res. 2013;41:W384–8. https://doi.org/10.1093/nar/gkt458.
23. Wiederstein M, Sippl M. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 2007;35:W407–10. https://doi.org/10.1093/nar/gkm290.
24. Jain S, Xiao X, Bogdan P, et al. Generator based approach to analyze mutations in genomic datasets. Sci Rep. 2021;11:21084. https://doi.org/10.1038/s41598-021-00609-8.
25. Chen T, Guestrin C. Xgboost. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016. https://doi.org/10.1145/2939672.2939785.
26. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2000;29:1189–232. https://doi.org/10.1214/aos/1013203451.
27. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Adv Neural Inf Proces Syst. 2017;30:1.
28. Gupta G, Xiao X, Bogdan P. Multiwavelet-based operator learning for differential equations. Adv Neural Inf Proces Syst. 2021;34:1.
29. Gupta G, Xiao X, Balan R, et al. Non-linear operator approximations for initial value problems. In: International Conference on Learning Representations; 2021.

Chapter 12
AI and Dynamic Prediction of Deterioration in Covid-19

Carolin Danker, Marat Murzabekov, David Forsberg, Niklas Lidströmer, Antoine Honoré, Susanne Rautiainen, and Eric Herlenius

C. Danker
Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden

IS Intelligent Systems, Kungliga Tekniska Högskolan, Stockholm, Sweden
e-mail: cdanker@kth.se

M. Murzabekov · D. Forsberg · E. Herlenius (✉)
Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden

Astrid Lindgren Children's Hospital, Karolinska University Hospital, Stockholm, Sweden
e-mail: marat.murzabekov@stud.ki.se; david.forsberg@ki.se; eric.herlenius@ki.se

N. Lidströmer
Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden
e-mail: niklas.lidstromer@ki.se

A. Honoré
Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden

IS Intelligent Systems, Kungliga Tekniska Högskolan, Stockholm, Sweden

Astrid Lindgren Children's Hospital, Karolinska University Hospital, Stockholm, Sweden
e-mail: antoine.honore@ki.se

S. Rautiainen (✉)
Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden

Astrid Lindgren Children's Hospital, Karolinska University Hospital, Stockholm, Sweden

Department of Global Public Health, Karolinska Institutet, Stockholm, Sweden

Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA
e-mail: Susanne.Rautiainen@ki.se

N. Lidströmer, Y. C. Eldar (eds.), *Artificial Intelligence in Covid-19*,
https://doi.org/10.1007/978-3-031-08506-2_12

Introduction

In December 2019, a novel coronavirus, severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2), emerged in Wuhan, China. The disease caused by SARS-CoV-2 was named coronavirus disease 2019 (COVID-19).

As we are entering a third calendar year since the start of the COVID-19 pandemic, there have been more than 490 million reported cases which has resulted in at least 6.2 million deaths worldwide (https://coronavirus.jhu.edu/). This figure is the official death toll—a recent article in *Nature*, shows a modelling, which suggests three times higher figures, i.e., surpassing 18 million deaths in the pandemic until the end 2021 [1].

Although remarkable vaccine initiatives in the world have proven to be effective and provide substantial protection against severe disease and death, new variants of the virus continue to emerge. One study suggests COVID-19 could continue in a pandemic form, surging in one or more regions and disrupting daily life, likely until the world reaches herd immunity [2]. Other projections, that COVID-19 would remain as an endemic disease with stable infection rates, do not mean the virus would somehow disappear or necessarily develop into a more benign form [3]. Viral seroarchaeology research has suggested that the Russian Flu of 1889 held the common ancestor of the two SARS outbreaks in the recent one and a half century, in which case the Russian Flu was preserved as, e.g., a common cold until our times [4].

A majority of infected patients exhibit mild illness, with symptoms which usually disappear within 2 weeks. Still several individuals develop severe disease, which necessitates hospitalization and therapeutic medical support, trying to avoid death and morbidity. With resources temporarily strained by the pandemic situation it is crucial to rapidly identify those with an onset of severe and critical illness and the risk of subsequent death.

When patients enter the emergency department triaging is performed. Pre-existing routines at emergency department are adapted to swiftly identify patients with critical illness. These include prompt identifications of patients with possible sepsis, and initiate treatments with broad-spectrum antibiotic agents and intravenous fluids. Under standard sepsis management protocols, every hour's delay in completion of the administration of the 3 h bundle is associated with a 4% increase in mortality [5]. However, do these established scores help identify COVID-19 patients at risk of critical illness and death?

The measurements of vital parameters play central roles in the identification of patients running risk of serious complications, albeit only a handful of studies have investigated how changes, i.e., the dynamics of manual or continuous measurements of vital parameters associate with critical illness, adverse events and death in COVID-19 hospitalized patients, e.g., [6].

Here we depict if and how Clinical Decision Support Systems (CDSS) and AI (Artificial intelligence assisted methods), applied on dynamic vital parameters in individual patients, could provide personalized medicine. Hence, to enable patient-specific prediction of critical illness, sepsis and mortality in patients hospitalized for COVID-19.

COVID-19: A Novel Disease—Usage of Newer or Older Clinical Decisions Support Systems?

COVID-19 infections were one of the leading causes of death 2020–2021.

COVID-19 was a novel disease in 2020, but early associations, in the first months of the pandemic, demonstrated the risk of rapid deterioration and death, i.e., higher age, but also male gender [7]. Other features included comorbidities (e.g., hypertension, diabetes, or cardiovascular diseases) and specific laboratory results [8]. Due to the high impact on healthcare systems worldwide, and for an adequate use of resources, risk assessment of patients became crucial [8].

Early detection of patient deterioration and life-threatening events in patients are challenging. Thus, providing effective life-saving interventions in due time is essential. During rapid progression of viral replication and infection, the immune and autonomic control systems of COVID-19 patients are continuously adapting to maintain homeostasis. Hereby, the dynamic disease progression and patient response are under constant development and change over time. Therefore, the initial signs preceding potentially life-threatening events can be subtle. Clinical detection is aided by biomarker analysis, which are unfortunately laborious, requiring manual interventions, invasive sampling and is time consuming. Infection and inflammation interfere with the autonomic control systems and consequently affect vital signs. Constant automatic or semiautomatic monitoring of vital signs with high frequency may enable the immediate detection of discrepancies and is therefore a key non-invasive instrument in modern emergency units, on hospital wards and intensive care units.

Clinical Decisions Support System Stable Parameters/ Features Using Threshold Values

Patient Deterioration

Identifying patients at risk of clinical deterioration is important, yet difficult [9, 10]. Biomarkers are used to follow disease development, but are often slow to analyze, requiring a biological sample, such as blood, cerebrospinal fluid or urine, and can only be measured at low frequency [11]. Instead, scoring of a variety of clinical parameters in hospitalized patients can be used as a proxy for disease progress and has been used over the last decades [12, 13]. Abnormal vital signs can indicate clinical deterioration in real-time, and therefore help to detect cardiac, respiratory shock, and sepsis [14, 15]. Thromboembolism, respiratory distress and sepsis are associate with severe COVID-19 The latter is a life-threatening organ dysfunction caused by a dysregulated host response to infection [16]. This definition, independent of infection, but depending on organ dysfunction, is called Sepsis-3 criteria. Notably, the performance of blood cultures in detection of microorganisms is poor, with less than

40% of sepsis patients exhibiting clinically relevant bacterial growth, take time, often days, before replies [17]. If patient deterioration stays undetected or is detected too late, it can lead to death [14].

Therefore, vital signs are used to measure and assess the physiological conditions of a patient.

Widely used vital signs are pulse/heart rate (HR), blood pressure (BP), respiratory rate (RR), peripheral oxygen saturation (SpO2) and temperature [18, 19]. Abnormalities in vital signs indicate an increased risk of death, hence a thorough analysis is critical to determine the strategy for therapeutic interventions and individual patients support [20].

Vital signs are checked manually by the nursing staff every few hours. Without continuous monitoring, abnormalities may be missed, thus putting the patient at risk. Therefore, particularly in intensive care units (ICUs), patients are continuously monitored to detect complications at an early stage. Commonly used signals are the blood oxygen saturation levels, indirect assessment via transcutaneous probes and the electrocardiogram (ECG).

Notably, established criteria to determine the risk of patient deterioration, use **threshold values**, such as cyanosis (blood oxygen saturation), heart frequency, blood pressure and respiratory rate.

Together with laboratory values, thresholds for vital signs form the basis of prediction scores for patient deterioration. In contrast to patients vital signs, which exhibits continuous variation and adaption in response to disease, the set, i.e., fixed, threshold values are not dynamic.

General Prediction Scores

Several severity scores to evaluate the patient's condition when entering the ICU have been developed [21]. These can be divided into generic and disease-specific scores [22]. Generic scores are applied to assess the severity at ICU admission to predict an outcome, to assess the severity of organ dysfunction, and lastly, scores that assess the nursing workload [22] Table 12.1. In the former category, three widespread scores have been developed over the last 40 years: the *Simplified Acute Physiology Score* (SAPS), the *Acute Physiology and Chronic Health Evaluation* (APACHE) II score, and the *Mortality Probability Model* (MPM) [21]. All three score systems aim to predict mortality within the first hours of admission.

A lot of effort has been made to further the development of scoring systems to predict the risk of mortality for critically ill patients. In the aetiology of hospital mortality, sepsis is leading. Hence, reduction of mortality is dependent on early sepsis prediction. A major problem is the similarity between these symptoms and signs and other conditions of lesser criticality. Two relevant terms in this context are sepsis and septic shock.

To identify critical patients, the Sequential Organ Failure Assessment score (SOFA) is used. SOFA is a mortality prediction score, which is based on the degree

Table 12.1 Scoring systems in clinical use for mortality and sepsis prediction

System	Aim	Input parameters	Model	Results	Clinical use
APACHE II	To classify patients according to the severity of illness	Temperature, mean arterial pressure, heart rate, respiratory rate, oxygenation, pH, sodium, potassium, creatinine, hematocrit, white blood count, age, previous health status	Logistic regression	AUC: 0.86 [23]	ICU patients [24]
SAPS 3	To predict in-hospital mortality	Age, comorbidities, circumstances of admission, Glasgow coma scale, heart rate, systolic blood pressure, bilirubin, temperature, creatinine, leukocytes, platelets, hydrogen ion concentration (pH), and ventilatory support and oxygenation.	Logistic regression	AUC: 0.85 [25]	ICU patients
MPM_0 III[a]	To predict in-hospital mortality	Glasgow coma scale, age, systolic blood pressure, metastatic neoplasm, cirrhosis, cardiac dysrhythmia, intra-cranial mass, cardiopulmonary resuscitation	Logistic regression	AUC: 0.82 [26]	ICU patients at admission
SOFA	To describe the degree of organ dysfunction/ failure over time	Respiration PaO2/FiO2, platelets, bilirubin, blood pressure, Glascow Coma Scale, creatinine.	Logistic regression	AUC range 0.61–0.88 [27][b]	Acute morbidity in ICU patients [13]
qSOFA	To predict in-hospital mortality	Glascow Coma Scale, Systolic Blood Pressure, respiratory rate	Logistic regression	AUC: 0.81 [28]	Suspected infection outside of the ICU
CURB65	To predict mortality secondary to community acquired pneumonia	Confusion, Urea, Respiratory rate, Blood pressure, age ≥65 years old.	Logistic regression	AUC: 0.80 [27]	Triage of patients with community acquired pneumonia [29]

[a] MPM_0 is the score at ICU admission, while MPM_{24} is the score used 24 h after admission to the ICU

[b] SOFA score was not originally developed for prediction

of dysfunction of six organ systems. The SOFA score takes into account respiration, coagulation, liver, cardiovascular, nervous, and renal values. Moreover, several scoring systems to predict sepsis instead of mortality per se, exist [30]. Sepsis exhibits a high mortality and morbidity. However, if sepsis is detected early adequate therapies may save life of patients. Notably, the multifactorial characteristic of the disease makes early diagnosis a challenging task for physicians. As with

mortality prediction, most studies process static features or aggregated dynamic features and thus make non-continuous predictions. Notably, the performance of blood cultures in detection of microorganisms is poor, with less than 40% of sepsis patients exhibiting clinically relevant bacterial growth [17].

The SOFA score is the basis of the Sepsis-3 criteria. Septic shock is a form of sepsis in which acute abnormalities of the circulation and cellular metabolism lead to an even higher risk of death [16]. There is also a Quick SOFA (qSOFA) score to identify high-risk patients for in-hospital mortality, with suspected infection outside the ICU (Table 12.1).

Critical illness induced by SARS-CoV-2 in COVID-19 patients fulfils the SOFA and Sepsis-3 criteria with Sequential organ dysfunction and the risk of adverse events and death.

Recent evidence provides support for assessing respiratory function with peripheral oxygen saturation in the Sequential Organ Failure Assessment score and the Sepsis-3 criteria. Threshold peripheral oxygen saturation thresholds 94% and 90% to get 1 and 2 Sequential Organ Failure Assessment respiratory points, respectively. This may have important implications for emergency practice, rapid response teams, surveillance, research, and resource-limited settings [31].

Another clinically widely used tool is the CURB65 score [32]. The CURB65 score is developed and might be used for assessment of the severity of community-acquired pneumonia and include characterization of Confusion, Urea, Respiratory rate, Blood pressure, Age 65; Acute Physiology and Chronic Health Evaluation [33] (Table 12.1).

The presented scores e.g., SOFA, APACHE II and CURB65, are established, validated and may help in the triage problem [32]. However, each score has its own problems and limitations, including the need for laboratory variables that are hard to obtain at hospital admission [34].

Early Warning Systems (EWS)

Real time prediction will save critical time for therapeutic interventions, much ahead of clearly visible clinical symptoms. This is called early warning systems (EWS), which aid accurate medical decisions. An optimal EWS will perform real time data acquisition, storage, analysis, and provide automatic monitoring and appropriate warning notifications. But are these in use in the clinic, and can they be applied to or adapted for COVID-19? (Fig. 12.1).

Early warning scores (EWS), such as the National EWS (NEWS), have been extensively validated and proven useful to predict general, non-sepsis-related, deterioration [10, 35, 36]. Additionally, NEWS outperforms qSOFA in detecting seriously ill patients, even in those with infectious disease [37]. Even in Covid-19 patients, NEWS can give a high reliability in detecting patient deterioration [38]. However, since all EWS rely on manually annotated vital signs, there is a great

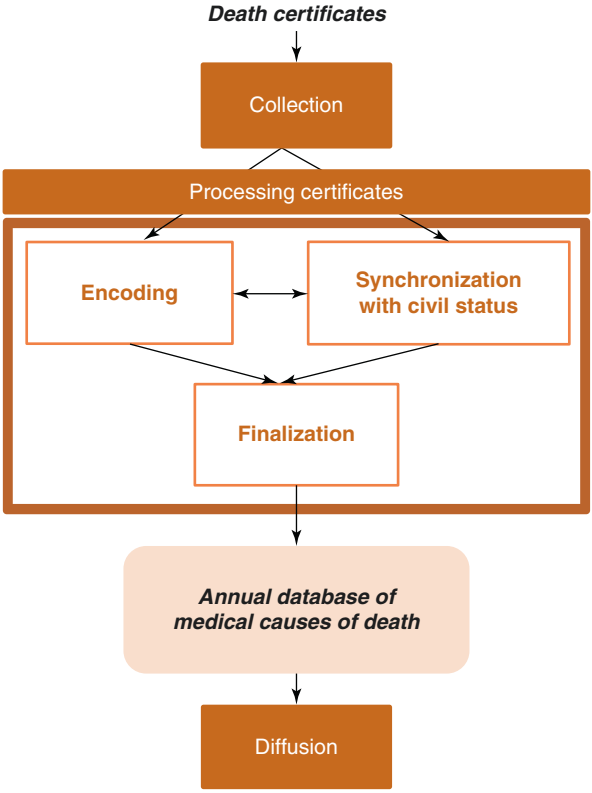


Fig. 12.1 Multimodal data sources from patients. Despite surveillance, late diagnosis of adverse events often occur in hospitalised patients, e.g., with some 100,000 deaths/year in hospitalized patents only in US. This in spite of initial subtle early signs, but the massive amount of data are often in an undigestible format. E.g., Multimodal data >75 million data points/parameter/week from high frequency (40–500 Hz) vital sign monitors. This in contrast to 168 data points/parameter/week if recorded manually per hour. High cognitive demand for data integration (Electronic Health records, tests, and surveillance data). Thus, integrated assessment of individual patients benefits from interpretable Clinical decision Support Systems

potential in assessing continuous vital signs monitoring with more advanced analysis techniques, such as the neonatal clinical decision support systems e.g. [39].

Traditionally, linear models such as logistic regression analysis have been used to construct such prognostication tools. However, in recent years, scores based on non-linear machine learning have emerged [40].

AI for Prediction of Deterioration

Artificial Intelligence (AI) can be of great use for assisting in the analysis of continuous vital parameter, demographic, laboratory and other manually achieved patient variables integrated into a Clinical Decision Support System (CDSS). The use of AI in the early prediction and diagnosis of severe complications, such as sepsis and hospital mortality, with the use of unstructured healthcare data, has been elaborated by several teams and accelerated in the recent few years. In the following, a brief summary of commonly used AI models is given.

Machine Learning (ML) is a subfield of artificial intelligence aiming to make a prediction based on input data. ML methods can either output numeric values (regression) or categorical values (classification). A model is trained by minimizing a loss function, which measures the discrepancy between the trained predictions and the true outcomes [41].

A simple regression model is a linear regression, that predicts an output based on one or more input parameters using a linear function. For binary classification logistic regression (LR) can be used instead. Logistic regression is similar to linear regression, but uses a logistic function for classification. Next, support vector machines (SVM) are models able to perform non-linear classifications. Similar to regression models, a decision boundary is learned, but uses a different kind of loss function. By applying a kernel function to the data, it is possible to create non-linear decision boundaries [41]. Another ML method is the decision tree (DT), which consists of nodes to predict an outcome for given features. A single decision tree tends to overfit; instead, an ensemble of decision trees, called random forest (RF), can be used that combines the predictions of the individual decision trees [41]. To train the random forest efficiently, boosting can be used. One commonly used algorithm for scalable tree boosting is XGBoost (extreme gradient boosting) [42]. Finally, a statistical-based approach for processing sequential data is the Hidden-Markov Models (HMM).

Machine learning approaches tend to perform poorly on complex problems with a large number of features, therefore, deep learning methods based on artificial neural networks (ANNs) can be applied [43]. ANNs replicate the idea of biological neural networks, that use neurons to store and forward information [41]. A basic ANN has an input layer, an output layer, and hidden layers. A deep neural network (DNN) is an ANN with many hidden layers [41]. Commonly used deep learning methods for analyzing time-series data are convolutional neural networks (CNNs) and recurrent neural networks (RNNs) (Fig. 12.2).

CNNs are neural networks that process data in a grid-like topology, such as time-series data (1D topology) or images (2D topology). A CNN consists of several layers, including convolution layers, nonlinear activation functions, pooling layers, and fully-connected layers. The sequence of convolution, pooling, and activation functions allows it to capture abstract features from the input data, that are then used by the fully-connected layer to compute an output [43] (Fig. 12.3).

Like CNNs, RNNs process grid-like data, but are specialized for sequential data. The basic idea of RNNs is that at each time step a prediction is made based on previous information and a new input. Because the network can process feedback from previous steps, it is called recursive. Apart from the basic RNN cell architecture, other architectures have been developed. One is the long short-term memory architecture (LSTM) that uses additional parameters to control the forwarded information. This allows certain information to be forgotten and stored, respectively, making it easier to account for long-term dependencies [43]. A similar architecture is the gated recurrent unit (GRU).

For the use of AI in design of efficient and trusted EWS, it is important to develop explainable machine learning (EML) for robust, reliable, and comprehensible data analysis. Explanations could help examine whether a machine learning method has

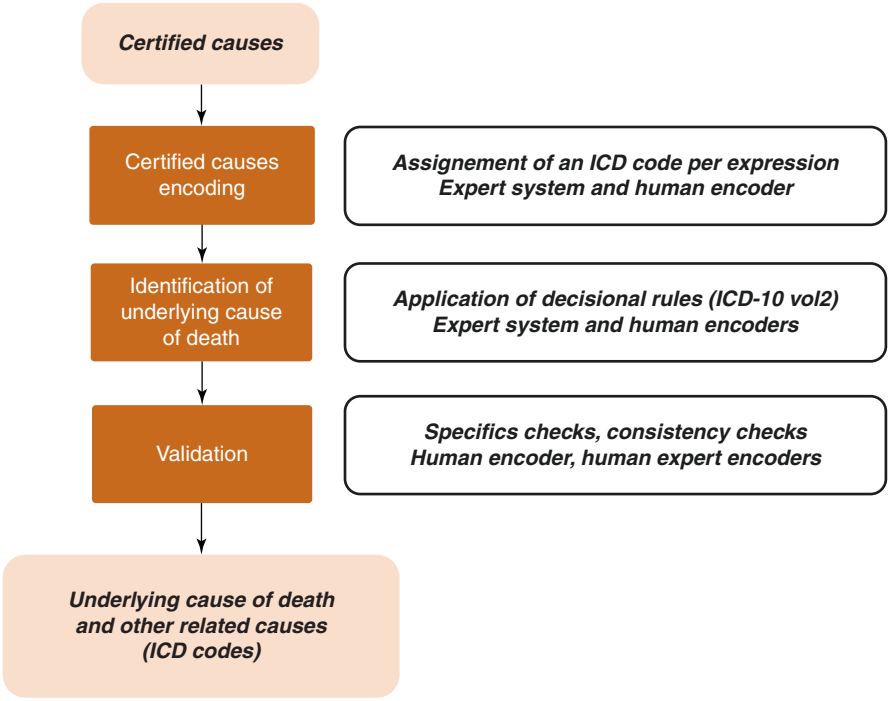


Fig. 12.2 Basic architecture of (**a**) Artificial Neural Network (ANN). (**b**) Basic architecture of a Deep Neural Network (DNN). Both models consist of an input layer one (ANN) or several (DNN) hidden layers, and an output layer

employed true evidence instead of biases that widely exist in training data. The ideas from unrolling iterative estimation algorithms are attracting high attention for developing neural networks. The growing popularity of unrolled deep networks is due in part to their potential in developing efficient, high-performance and yet interpretable NN architectures from reasonable size training sets [45].

AI Assisted Patient-Specific Risk Prediction

A possible use of machine learning assisted analysis of continuous vital parameter, demographic, laboratory and other manually achieved patient variables integrated into a clinical decision support system (CDSS).

A targeted breakthrough would be the development of novel paradigms for EWS with focus on deep systems. This breakthrough might be achieved by combining model-driven systems and data-driven learning methods. Models will help to design necessary constraints for analysis, in turn explainability. Our hypothesis is that data-driven optimization of proposed deep systems composed of many model-driven components will offer high quality performance along with explainability (Fig. 12.4).

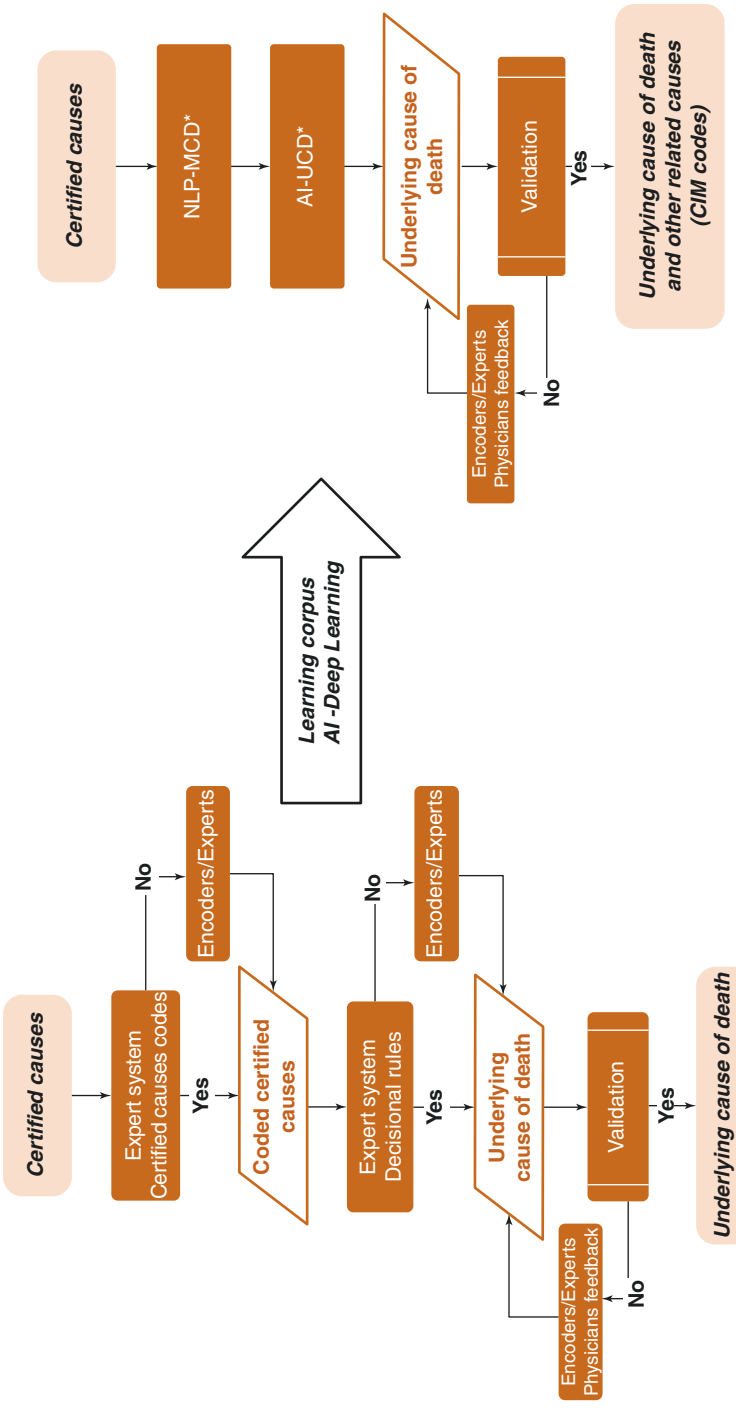


Fig. 12.3 Example of a convolutional neural network architecture for time-series data. The architecture take a time-series data as input and applies convolution and pooling steps, before using a fully connected layer to compute an output. (Adapted from [44])

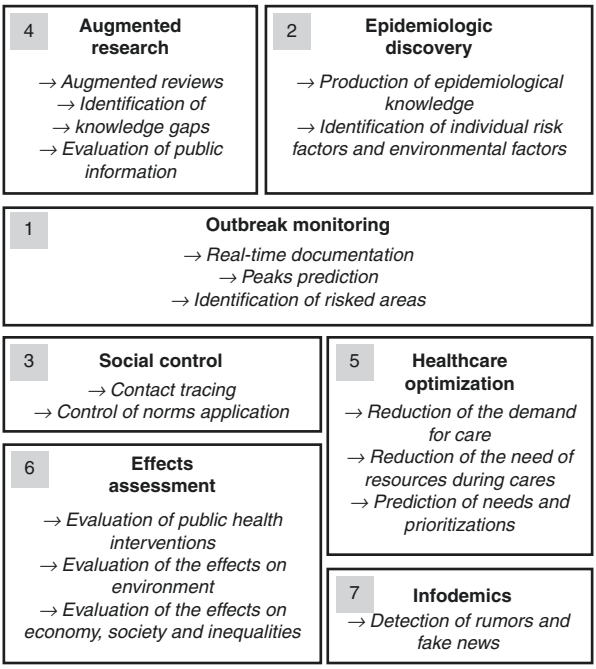


Fig. 12.4 DeepNEWS—clinical decision support system. Adverse clinical events may quickly lead to cardiorespiratory dysfunction and potentially death in patients in hospital wards and intensive care units (ICU). Changes in the vital parameters appear early and can be detected using novel machine learning based approach—a physiomarker toolbox. The physiomarker toolbox provide an alert to the medical team and suggest a risk reduction strategy on which the clinical staff can base their interventions to prevent serious disease

The use of artificial intelligence in the early prediction and diagnosis of severe complications, such as sepsis, and hospital mortality, with the use of unstructured healthcare data has been elaborated by several teams and accelerated in the recent few years. Some of these methods have also been applied to COVID-19 populations.

AI Assisted Prediction of Critical Illness and Deterioration in COVID-19 Patients

Mortality Prediction Models for Covid-19

Many studies have been published in which machine learning and deep Learning techniques have been used to predict clinical deterioration and mortality for Covid-19 patients to provide more specific care [46, 47]. Commonly used features are demographics, vital signs, comorbidities, medication, laboratory values, and imaging (X-ray, CT) [47].

Among all these available factors, machine learning methods have been shown to help identify particularly important factors influencing COVID-19. For instance, in the COVIDAge Study the predictors of in-hospital mortality in elderly patients were analysed, and it was concluded that in COVID-19, respiratory crackles, male gender, increased fraction of inspired oxygen (FiO2) and the level general functionality were all independently prognostic risk factors for mortality, and not the age per se [48]. Among sepsis patients subject to resuscitation in the emergency department all vital parameters decreased, with the exception of peripheral saturation, which increased. Nearly all standard blood samples decreased too, except for CRP, K, Troponin T and Bilirubin, which were unchanged. The non-descending parameters were directly affected by the treatment. Most patients improved, and it was concluded that vital parameters and routine biomarkers can mitigate the clinical course prediction in the earlier stages of sepsis resuscitations [48].

In the ICU there are already modus operandi with APACHE (acute physiology and chronic health evaluation) II and Sequential Organ Failure Assessment (SOFA) to deem the mortality, but as has been notoriously seen, these have proved to be less useful for severely affected COVID-19 individuals' survival.

Machine learning approaches can also be used to predict mortality and several studies have been published differing in terms of input features and algorithms (Table 12.2). Predictions can be made using blood samples [58, 59], clinical parameters and vital signs [49] or CT images [49]. While most models require a variety of input features, it has been shown that already a reduced amount of features can be used for a prediction model, such as age, oxygen saturation, and type of patient encounter [50]. Using machine learning based model, in a small ICU patient material, and an extensive analysis of proteomic profile can predict which COVID-19 infected individual will be most severely affected [60]. This based solely on a single blood sample, containing a series of blood plasma proteins taken upon arrival in the hospital ICU. Machine learning methods can be logistic regression [6, 49, 50], random forest [49], survival analysis [49], or XGBoost models [49, 51, 52].

However, most of these research models are based on advanced blood sample analysis that might help in the acute triaging, when resources are available to perform these tests. But these invasive and complicated blood sample analysis cannot be performed in most emergency settings worldwide. Neither do they allow us to screen, detect and protect patient deterioration over time. For this patient-specific prediction dynamic, continuous time series data is needed (Fig. 12.1).

Important to consider is the time between prediction in relation to admission and time of death. To be able to decide for the right treatment, a reliable prediction should be made as close to the hospital admission as possible. Machine learning methods can improve traditional prediction scores in the first days [52]. For an adequate prediction already the vital signs of the first 2 h after admission can be used [51].

In addition to machine learning approaches, deep learning methods can be of great help to identify important predictors for mortality in Covid-19 patients (Table 12.2). For example, it has been shown that a DNN with five fully connected dense layers can be used to predict ICU admission and mortality using EHR data.

Table 12.2 Comparison of exemplary studies applying machine learning methods to Covid-19 populations to predict clinical adverse events

Author	Aim	Size of dataset	Input parameters	Prediction model	Results
Machine learning					
Mendes [46]	Identify predictors of in-hospital mortality in older patients; predict mortality	235	DG, CM, VS, lab, scores, clinical status	LR	AUC: 0.82
Kamran [4]	Identify predictors of mortality; predict mortality	9291	DG, VS, lab, data from nursing flowsheets	LR	AUC range: 0.78–0.84
Yan [47]	Predict mortality	485	DG, lab, symptoms, epidemiological information	XGBoost	AUC: 0.95
Sánchez [49]	Predict mortality	1696	DG, VS	Survival analysis, LR, BN, DT, RF	AUC: 0.89 (LR)
Yadaw [50]	Predict mortality	4802	DG, CM, VS	LR, XGBoost, SVM, RF	AUC: 0.91 (XGBoost)
Burdick et al. [51]	Prediction of ventilation requirements	197	VS, lab	XGBoost	AUC: 0.87
Bolourani [52]	Prediction of respiratory failure	11,525	DG, VS, lab	LR, XGBoost	AUC: 0.77 (XGBoost) AUC: 0.70 (LR)
Tang [53]	Predict sepsis	2453	DG, lab	XGBoost	AUC: 0.95
Deep learning					
Li [54]	Identify top predictors; predict ICU admission and in-hospital mortality	1108	DG, CM, VS, lab, symptoms	DNN	**ICU Admission** AUC: 0.78 **Mortality** AUC: 0.84
Zhu [55]	Identify top predictors, predict mortality	181	DG, CM, VS, lab, symptoms	DNN	AUC: 0.97
Villegas [56]	Mortality prediction	6087	DG, VS, lab, medication	Ensemble of RNNs	AUC range: 0.89–0.90
Cheng [57]	Predict in-hospital mortality of ICU patients	654	DG, CM, VS, lab, X-rays	DNN, LTBN	**Clinical Data:** AUC: 0.65 **With X-Ray:** AUC: 0.73

Note: *DG* demographics, *CM* comorbidities, *VS* vital signs, *lab* laboratory test values, *DNN* deep neural network, *LR* logistic regression, *BN* Bayesian network, *DT* decision tree, *RF* random forest, *SVM* support vector machine, *RNN* recurrent neural network, *LTBN* longitudinal transformer-based network, *AUC* area under the curve, *ACC* accuracy

Age, SpO2, laboratory values and comorbidities were shown to be the top predictors for mortality [54]. The use of another DNN architecture on EHR data has revealed similar key predictors, including the oxygen index, and four laboratory values [55].

In addition to vital signs and laboratory values, other features can be included in the prediction model. For example, X-ray images can improve the mortality predictions in COVID-19 patients [57]. The X-ray images can be processed in a separate suitable model, while the vital signs and laboratory values can be processed in a similar way to the methods presented above.

A main drawback of ANN and DNN architectures is **the limited use for sequential data**. One method is to average and discretize measurement into intervals instead of using all data at once, but using the same architectures [61, 62]. Another method is the application of RNNs that are designed to process sequential data and can be used to update the prediction when new data is available. This makes it possible to combine static values and dynamic values [56].

Mortality Prediction Models Using High-Frequency Data

Most COVID-19 mortality predictions are mainly based on demographics and laboratory values and just a few include sequential data or even calculated features from high-frequency waveform data to provide dynamic predictions. Research using non-COVID-19 datasets has shown the potential for these models. Table 12.3 summarizes the most important aspects of the models. The models can be roughly divided into non-temporal and continuous prediction models.

Non-temporal models use different methods to aggregate sequential data. Statistical features (e.g., average, standard deviation, skewness, kurtosis) are calculated from waveform data and average before being processed by non-sequential machine learning model such as logistic regression, random forest or NN [56, 63, 64]. Waveform data can be used from the mean arterial pressure (MAP) [56] or ECG [63].

The main drawback of non-temporal models is the limited use for continuous prediction. Although it is possible to aggregate sequential to shorter intervals and make separate predictions [68], other architecture can be used that are specifically designed for taking sequential input and give continuous output. Architectures can be RNN [66], LSTM [40] or HMM [67].

Just like for the non-temporal models, statistical features are calculated from the time-series data first. Instead of aggregating the features, they are processed sequentially by the model, thus taking into account also development over time [40, 66, 67]. Furthermore, the combination of more than one RNN makes it possible to combine time-series data with different time-scales for example, monitored data that is continuously available with more sparsely available data such as medications, demographics, and laboratory values [66].

Table 12.3 Comparison of relevant studies for mortality prediction using time-series data

Author	Input parameters	Processing of time-series data	Model	Results
Non-temporal models				
Todd et al.	EMR data	Five calculated features of 24 h of waveform data	LR	**ACC:**
	Pulse wave			EMR: 0.91
	MAP wave			EMR + pulse: 0.92
				EMR + MAP: 0.92
Sadeghi et al. [63]	ECG Wave	Twelve calculated features from 1 h of waveform data	i.e. DT, RF, SVM, LR	AUC DT: 0.93
Kock [64]	EMR data	Median values of calculated features from a 1-min PPG waveform signals and its derivatives	LR, ANN	AUC LR: 0.85
	PPG wave			AUC ANN: 0.90
Morid et al. [65]	EMR data	Average of a 2-h interval of 48 h of time-series data.	k-NN	Precision: 0.65
				F-Measure: 0.66
Temporal models				
Ghanvatkar et al. [66]	EMR data	RNN processes mean and standard deviation of ECG signal with a frequency 1 Hz. A second RNN combines the output with lower frequency data with a frequency of 0.02 Hz	RNN	AUC: 0.71
	ECG wave			
Thorsen-Meyer et al. [40]	EMR data	One variable is aggregated for every 1-h interval.	RF, BN, PART	AUC RF: 0.82
Gupta [67]	EMR data	Aggregation of time-series data in 2-h intervals.	HMM	AUC: 0.87

Note: *EMR* electronic medical record, *MAP* mean arterial pressure, *ECG* electrocardiography, *PPG* photoplethysmography, *LR* logistic regression, *DT* decision tree, *RF* random forest, *SVM* support vector machine, *ANN* artificial neural network, *k-NN* k-nearest neighbors, *RNN* recurrent neural network, *LSTM* long short-term memory, *BN* Bayesian network, *HMM* hidden Markov model, *ACC* accuracy, *AUC* area under the curve

A key challenge remains to achieve good results as early as possible. Methods developed so far that have investigated this give the best results only closer to discharge [40, 67, 68].

Instead of calculating features from waveform data, the raw data could be used directly as input to the model, so that the model can learn the features by itself. A field where this type of feature learning was studies is ECG-based feature extraction, disease detection, and sleep staging [44, 69]. The most common networks applied are CNNs and RNNs [69].

Prediction Models for Sepsis

Even though prediction models focusing solely on mortality can be of great use in supporting triage, early prediction of sepsis is crucial in preventing mortality, given that sepsis management is highly time sensitive. Machine learning based models for the prediction and diagnosis of sepsis have been developed in the last few years.

For instance the SERA algorithm; trained on both structured and unstructured electronic health record contents, to predict sepsis [70]. In retrospective studies comparing how the SERA performs in comparison to the predictions of doctors. Here, in retrospect, the algorithm has the capability to improve sepsis prediction and reduce false positive sepsis diagnosis significantly [70]. The algorithm's correctness was increasing with the augmented mining of unstructured medical record notes, when compared to the usage of traditional clinical measures (NEWS) for early warning 12–48 h before the onset of sepsis.

Another approach was during the PhysioNet/Computing in Cardiology Challenge in 2019, where the unmet need for clinically useful algorithms led researchers to directly compare a range of sepsis prediction methods, hitherto not feasible due to disparate patient cohorts, sepsis definitions, scales and criteria, clinical variables, prediction tasks, evaluation metrics, and other obstacles. To deal with this issue the event orchestrated the fruition of automatised and open-source algorithms for the earliest sepsis detection based on data from clinics, making it comparable. It was broadly proven that algorithms can predict the onset of sepsis several hours earlier than the clinical eye, however a remaining issue was shown to be the lack of universality between different hospital platforms [71].

Recently retrospective analysis of adult patients admitted to the intensive care unit (from the MIMIC III database) demonstrate that a recurrent neural network is superior to previous developed models considering the prediction performance [40].

It would also be useful to have early predictions of emergency department (ED) patients requiring ICU transfer, since patients with a prolonged transfer have a much higher mortality and extended hospitalisation. It has been shown that it is possible to predict transfer based on the monitored data readily available in the ED.

To further explore the vital sign trajectories into the present health resource straining COVID-19 pandemic, where patients have shown rapid worsening after ICU admittance, a handful of teams have created models of vital parameters, in two patient cohorts with either COVID-19 in 2020 or viral pneumonia a few years earlier. A novelty score was created when comparing the groups. The patients with COVID-19 showed a much more rapid deterioration with respiratory failure, decreased SpO2 and elevated FiO2, but notably mere minor changes [72].

Finally, it has recently been demonstrated how AI models might help to predict survival in COVID-19. Furthermore, it also has been shown that coagulation function indicators are strongly related to a sepsis caused by COVID-19 including glomerular filtration rate, hematocrit, creatinine, and total bilirubin [53].

However, these models require advanced blood sample analysis and neither allow us to screen, detect and protect patient deterioration over time. To achieve patient-specific predictions continuous time-series data is needed as described for the mortality prediction models (Figs. 12.1 and 12.4).

Explainable and Interpretable Machine Learning Methods for Clinical Decision Support Systems

Overall aim is to design explainable machine learning algorithms for data analysis and decisions in critical health care. Health care data provide unique opportunities to implement novel scientific technologies to improve predictions and diagnostics of severe, eventually fatal outcomes in hospitals. The current COVID-19 pandemic is a real-life evidence that society has an urgent need for new technologies to screen, detect and protect patients via early analysis of patient data. These are early warning systems, which will process patient data, analyze the data and then provide aid to the medical care team for their clinical decisions far ahead in time, to avoid fatalities. Explainable machine learning is at the core of developing these new technologies—early warning systems (Fig. 12.4).

Explainable machine learning algorithms provide explanations of their outcomes. Explanations are necessary for the medical care team to understand clinical scenarios and rely on assistive technologies with trust and social acceptance. We propose analytically driven systematic design principles to achieve the main technical objective—development of a set of deep learning systems that offer explanations by combining model-driven signal processing and data-driven machine learning. It is necessary to address robustness against noise and model mismatch, stability for tracking time-series patient-data, understanding reasons for failures and developing their mitigation approaches, and relative influence of input observations for a system output.

Aided with explainable deep learning algorithms, the primary application of the early warning systems is to predict infection buildup based on time series patient data in hospital wards and intensive care units (ICUs). Potential of the early warning systems can be further explored in prediction of infection buildup and consequences for COVID-19 patients admitted to ICUs. The secondary application is early diagnosis of cancer using various image modalities.

Development of trustworthy explainable deep system based novel EWS (DeepNEWS) requires a strong cooperation between medical caregivers, medical researchers, ML researchers, data scientists and IT engineers in the hospital environment. Thus, we have some endeavours ahead of us in multidisciplinary teams to prepare for the next crisis.

Development of more effective predictive monitoring systems to early detect risk for critical illness, adverse events and death, will save lives through earlier and more efficient treatment. The machine-learning approach might be utilized in to allow for

individualized detection of adverse events that automatically takes into account patient characteristics such as age, sex, weight etc. Patient-specific interpretable prediction using individual dynamic heath care and vital signs data may provide medical professionals with a novel machine learning assisted Clinical Decision Support system.

This novel, explainable AI based Clinical Decision Support system for dynamic predictions, would benefit the care of COVID-19 patients. In addition, it could be applied in several clinical settings to facilitate decision making. Such as anesthetic monitoring during surgery, monitoring of critically ill patients in the emergency room and long-term surveillance at various pediatric and adult hospital wards.

Novel interpretable machine learning assisted Clinical Decision Support systems may help to facilitate the triage and early patient-specific therapeutic support. Thus, screen, detect and protect against critical illness in daily emergency and hospital care will help us be better prepared for the next crisis.

References

1. Adam D. COVID's true death toll: much higher than official records. Nature. 2022;603(7902):562.
2. Nature.com. COVID is here to stay: countries must decide how to adapt. Nature (London). 2022;601(7892):165.
3. Katzourakis A. COVID-19: endemic doesn't mean harmless. Nature. 2022;601(7894):485.
4. Vijgen L, Keyaerts E, Moes E, Thoelen I, Wollants E, Lemey P, et al. Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. J Virol. 2005;79(3):1595–604.
5. Seymour CW, Gesten F, Prescott HC, Friedrich ME, Iwashyna TJ, Phillips GS, et al. Time to treatment and mortality during mandated emergency care for sepsis. N Engl J Med. 2017;376(23):2235–44.
6. Kamran F, Tang S, Otles E, McEvoy DS, Saleh SN, Gong J, et al. Early identification of patients admitted to hospital for COVID-19 at risk of clinical deterioration: model development and multisite external validation study. BMJ. 2022;376:e068576.
7. Jin JM, Bai P, He W, Wu F, Liu XF, Han DM, et al. Gender differences in patients with COVID-19: focus on severity and mortality. Front Public Health. 2020;8:152.
8. Tian W, Jiang W, Yao J, Nicholson CJ, Li RH, Sigurslid HH, et al. Predictors of mortality in hospitalized COVID-19 patients: a systematic review and meta-analysis. J Med Virol. 2020;92(10):1875–83.
9. Gold DL, Mihalov LK, Cohen DM. Evaluating the Pediatric Early Warning Score (PEWS) system for admitted patients in the pediatric emergency department. Acad Emerg Med. 2014;21(11):1249–56.
10. Silcock DJ, Corfield AR, Gowens PA, Rooney KD. Validation of the national early warning score in the prehospital setting. Resuscitation. 2015;89:31–5.
11. Brown JVE, Meader N, Wright K, Cleminson J, McGuire W. Assessment of C-reactive protein diagnostic test accuracy for late-onset infection in newborn infants. JAMA Pediatr. 2020;174:260.
12. Spangfors M, Bunkenborg G, Molt M, Samuelson K. The national early warning score predicts mortality in hospital ward patients with deviating vital signs: a retrospective medical record review study. J Clin Nurs. 2019;28(7–8):1216–22.

13. Lambden S, Laterre PF, Levy MM, Francois B. The SOFA score-development, utility and challenges of accurate assessment in clinical trials. Crit Care. 2019;23(1):374.
14. Jones D, Mitchell I, Hillman K, Story D. Defining clinical deterioration. Resuscitation. 2013;84(8):1029–34.
15. Joshi M, Ashrafian H, Aufegger L, Khan S, Arora S, Cooke G, et al. Wearable sensors to improve detection of patient deterioration. Exp Rev Med Devices. 2019;16(2):145–54.
16. Singer M, Deutschman CS, Seymour C, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA. 2016;315(8):801–10.
17. Yu D, Unger D, Unge C, Parke A, Sunden-Cullberg J, Stralin K, et al. Correlation of clinical sepsis definitions with microbiological characteristics in patients admitted through a sepsis alert system; a prospective cohort study. Ann Clin Microbiol Antimicrob. 2022;21(1):7.
18. Sapra A, Malik A, Bhandari P. Vital sign assessment. Treasure Island, FL: StatPearls; 2021.
19. Orphanidou C. Signal quality assessment in physiological monitoring. Cham: Springer International Publishing; 2018.
20. Marik PE. Understanding the vital signs: BP, HR, RR, TEMP, SaO2 … and SV. Evidence-based critical care. Cham: Springer International Publishing; 2015. p. 169–96.
21. Le Gall J-R. The use of severity scores in the intensive care unit. Intensive Care Med. 2005;31(12):1618–23.
22. Vincent JL, Moreno R. Clinical review: scoring systems in the critically ill. Crit Care. 2010;14(2):207.
23. Wong DT, Crofts SL, Gomez M, McGuire GP, Byrick RJ. Evaluation of predictive ability of APACHE II system and hospital outcome in Canadian intensive care unit patients. Crit Care Med. 1995;23(7):1177–83.
24. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985;13(10):818–29.
25. Moreno RP, Metnitz PG, Almeida E, Jordan B, Bauer P, Campos RA, et al. SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. Intensive Care Med. 2005;31(10):1345–55.
26. Higgins TL, Teres D, Copes WS, Nathanson BH, Stark M, Kramer AA. Assessing contemporary intensive care unit outcome: an updated Mortality Probability Admission Model (MPM0-III). Crit Care Med. 2007;35(3):827–35.
27. Minne L, Abu-Hanna A, de Jonge E. Evaluation of SOFA-based models for predicting mortality in the ICU: a systematic review. Crit Care. 2008;12(6):R161.
28. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, et al. Assessment of clinical criteria for sepsis: for the third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA. 2016;315(8):762–74.
29. Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax. 2003;58(5):377–82.
30. Kausch SL, Sullivan B, Spaeder MC, Keim-Malpass J. Individual illness dynamics: an analysis of children with sepsis admitted to the pediatric intensive care unit. PLoS Dig Heath. 2022;1:e0000019.
31. Valik JK, Mellhammar L, Sunden-Cullberg J, Ward L, Unge C, Dalianis H, et al. Peripheral oxygen saturation facilitates assessment of respiratory dysfunction in the sequential organ failure assessment score with implications for the sepsis-3 criteria. Crit Care Med. 2022;50(3):e272–e83.
32. Richards G, Levy H, Laterre PF, Feldman C, Woodward B, Bates BM, et al. CURB-65, PSI, and APACHE II to assess mortality risk in patients with severe sepsis and community acquired pneumonia in PROWESS. J Intensive Care Med. 2011;26(1):34–40.
33. Alan M, Grolimund E, Kutz A, Christ-Crain M, Thomann R, Falconnier C, et al. Clinical risk scores and blood biomarkers as predictors of long-term outcome in patients with community-acquired pneumonia: a 6-year prospective follow-up study. J Intern Med. 2015;278(2):174–84.

34. Zou X, Li S, Fang M, Hu M, Bian Y, Ling J, et al. Acute physiology and chronic health evaluation II score as a predictor of hospital mortality in patients of coronavirus disease 2019. Crit Care Med. 2020;48(8):e657–e65.
35. Churpek MM, Snyder A, Han X, Sokol S, Pettit N, Howell MD, et al. Quick sepsis-related organ failure assessment, systemic inflammatory response syndrome, and early warning scores for detecting clinical deterioration in infected patients outside the intensive care unit. Am J Respir Crit Care Med. 2017;195(7):906–11.
36. Dawes TR, Cheek E, Bewick V, Dennis M, Duckitt RW, Walker J, et al. Introduction of an electronic physiological early warning system: effects on mortality and length of stay. Br J Anaesth. 2014;113(4):603–9.
37. Redfern OC, Smith GB, Prytherch DR, Meredith P, Inada-Kim M, Schmidt PE. A comparison of the quick sequential (sepsis-related) organ failure assessment score and the national early warning score in non-ICU patients with/without infection. Crit Care Med. 2018;46(12):1923–33.
38. Kostakis I, Smith GB, Prytherch D, Meredith P, Price C, Chauhan A, et al. The performance of the National Early Warning Score and National Early Warning Score 2 in hospitalised patients infected by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Resuscitation. 2021;159:150–7.
39. Sullivan BA, Fairchild KD. Vital signs as physiomarkers of neonatal sepsis. Pediatr Res. 2021;91:273.
40. Thorsen-Meyer HC, Nielsen AB, Nielsen AP, Kaas-Hansen BS, Toft P, Schierbeck J, et al. Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. Lancet Dig Health. 2020;2(4):e179–e91.
41. Jung A. Machine learning. Singapore: Springer Singapore; 2022.
42. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13–17 August 2016; 2016. p. 785–94.
43. Lan G. BYCA. In: Deep learning; 2016.
44. Ebrahimi Z, Loni M, Daneshtalab M, Gharehbaghi A. A review on deep learning methods for ECG arrhythmia classification. Exp Syst Appl X. 2020;7:100033.
45. Monga V, Li Y, Eldar YC. Algorithm unrolling: interpretable, efficient deep learning for signal and image processing. IEEE Signal Process Mag. 2021;38:8–44.
46. Khan IU, Aslam N, Aljabri M, Aljameel SS, Kamaleldin MMA, Alshamrani FM, et al. Computational intelligence-based model for mortality rate prediction in COVID-19 patients. Int J Environ Res Public Health. 2021;18(12):6429.
47. Islam MM, Poly TN, Alsinglawi B, Lin MC, Hsu MH, Li YC. A state-of-the-art survey on artificial intelligence to fight COVID-19. J Clin Med. 2021;10(9):1961.
48. Mendes A, Serratrice C, Herrmann FR, Genton L, Périvier S, Scheffler M, et al. Predictors of in-hospital mortality in older patients with COVID-19: the COVID age study. J Am Med Dir Assoc. 2020;21(11):1546–54.e3.
49. Sánchez-Montañés M, Rodríguez-Belenguer P, Serrano-López AJ, Soria-Olivas E, Alakhdar-Mohmara Y. Machine learning for mortality analysis in patients with COVID-19. Int J Environ Res Public Health. 2020;17(22):8386.
50. Yadaw AS, Li Y, Bose S, Iyengar R, Bunyavanich S, Pandey G. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. Lancet Dig Health. 2020;2(10):e516.
51. Burdick H, Lam C, Mataraso S, Siefkas A, Braden G, Dellinger RP, et al. Prediction of respiratory decompensation in COVID-19 patients using machine learning: the READY trial. Comput Biol Med. 2020;124:103949.
52. Bolourani S, Brenner M, Wang P, McGinn T, Hirsch JS, Barnaby D, et al. A machine learning prediction model of respiratory failure within 48 hours of patient admission for COVID-19: model development and validation. J Med Internet Res. 2021;23(2):e24246.
53. Tang G, Luo Y, Lu F, Li W, Liu X, Nan Y, et al. Prediction of sepsis in COVID-19 using laboratory indicators. Front Cell Infect Microbiol. 2020;10:586054.

54. Li X, Ge P, Zhu J, Li H, Graham J, Singer A, et al. Deep learning prediction of likelihood of ICU admission and mortality in COVID-19 patients using clinical variables. PeerJ. 2020;8:e10337.
55. Zhu JS, Ge P, Jiang C, Zhang Y, Li X, Zhao Z, et al. Deep-learning artificial intelligence analysis of clinical variables predicts mortality in COVID-19 patients. J Am Coll Emerg Physic Open. 2020;1(6):1364.
56. Villegas M, Gonzalez-Agirre A, Gutiérrez-Fandiño A, Armengol-Estapé J, Carrino CP, Fernández DP, et al. Predicting the evolution of COVID-19 mortality risk: a recurrent neural network approach. medRxiv. 2021:2020.12.22.20244061.
57. Cheng J, Sollee J, Hsieh C, Yue H, Vandal N, Shanahan J, et al. COVID-19 mortality prediction in the intensive care unit with deep learning based on longitudinal chest X-rays and clinical data. Eur Radiol. 2022;1-11
58. Yan L, Zhang H-T, Goncalves J, Xiao Y, Wang M, Guo Y, et al. An interpretable mortality prediction model for COVID-19 patients. Nat Mach Intell. 2020;2(5):283–8.
59. Karthikeyan A, Garg A, Vinod PK, Priyakumar UD. Machine learning based clinical decision support system for early COVID-19 mortality prediction. Front Public Health. 2021;9:626697.
60. Demichev V, Tober-Lau P, Nazarenko T, Lemke O, Kaur Aulakh S, Whitwell HJ, et al. A proteomic survival predictor for COVID-19 patients in intensive care. PLoS Dig Health. 2022;1(1):e0000007.
61. Ryan L, Lam C, Mataraso S, Allen A, Green-Saxena A, Pellegrini E, et al. Mortality prediction model for the triage of COVID-19, pneumonia, and mechanically ventilated ICU patients: a retrospective study. Ann Med Surg. 2020;59:207.
62. Parchure P, Joshi H, Dharmarajan K, Freeman R, Reich DL, Mazumdar M, et al. Development and validation of a machine learning-based prediction model for near-term in-hospital mortality among patients with COVID-19 PT and AK contributed equally. BMJ Support Palliat Care. 2020:1–8.
63. Sadeghi R, Banerjee T, Romine W. Early hospital mortality prediction using vital signals. Smart Health. 2018;9–10:265–74.
64. Kock KS, Marques JLB. Use of photoplethysmography to predict mortality in intensive care units. Vasc Health Risk Manag. 2018;14:311.
65. Morid MA, Sheng ORL, Abdelrahman S. PPMF: A patient-based predictive modeling framework for early ICU mortality prediction. arXiv. 2017;
66. Ghanvatkar S, Rajan V. Deep recurrent neural networks for mortality prediction in intensive care using clinical time series at multiple resolutions. In: 40th International Conference on Information Systems, ICIS 2019; 2019.
67. Gupta A, Liu T, Crick C. Utilizing time series data embedded in electronic health records to develop continuous mortality risk prediction models using hidden Markov models: a sepsis case study. Stat Methods Med Res. 2020;29(11):3409–23.
68. Awad A, Bader-El-Den M, McNicholas J, Briggs J, El-Sonbaty Y. Predicting hospital mortality for intensive care unit patients: time-series analysis. Health Informatics J. 2020;26(2):1043–59.
69. Hong S, Zhou Y, Shang J, Xiao C, Sun J. Opportunities and challenges of deep learning methods for electrocardiogram data: a systematic review. Comput Biol Med. 2020;122:103801.
70. Goh KH, Wang L, Yeow AYK, Poh H, Li K, Yeow JJL, et al. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. Nat Commun. 2021;12(1):711.
71. Reyna MA, Josef CS, Jeter R, Shashikumar SP, Westover MB, Nemati S, et al. Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019. Crit Care Med. 2020;48(2):210–7.
72. Dhont S, Derom E, Van Braeckel E, Depuydt P, Lambrecht BN. The pathophysiology of 'happy' hypoxemia in COVID-19. Respir Res. 2020;21(1):198.

Chapter 13
AI, Epidemiology and Public Health in the Covid Pandemic

Thomas Lefèvre, Hélène Colineaux, Claire Morgand, Laurent Tournois, and Cyrille Delpierre

Introduction

The management, at all scales and with a view to control or eradication, of a known or emerging contagious disease is a founding textbook case of public health. Public health is also the "practical" or applied side of epidemiology: epidemiology largely provides the framework, tools, and knowledge necessary for public health measures, vis-à-vis this or that phenomenon. Thus, epidemiology and public health are intimately linked, and are particularly so, historically, around infectious diseases.

The case of the Covid pandemic then looks like a textbook case, to which it would have been or has been possible to apply knowledge and skills in epidemiology in the service of public health measures to be taken, in order to manage the

T. Lefèvre (✉)
IRIS Institut de Recherche Interdisciplinaire sur les enjeux Sociaux, UMR8156 CNRS - U997 Inserm - EHESS—Université Sorbonne Paris Nord, Villetaneuse, France

AP-HP, Jean Verdier hospital, Department of forensic and social medicine, Bondy, France
e-mail: thomas.lefevre@univ-paris13.fr

H. Colineaux · C. Delpierre
CERPOP UMR 1295, Université Toulouse III Paul Sabatier, INSERM, Toulouse, France
e-mail: colineaux.h@chu-toulouse.fr; cyrille.delpierre@inserm.fr

C. Morgand
Centre for Epidemiology on the Medical Causes of Death, Inserm, Le Kremlin-Bicêtre, France

Data and Health Studies Department, Regional Health Agency, Saint-Denis, France

L. Tournois
BioSilicium, Riom, France

CNRS UMR 8045 BABEL, Université Paris Cité, Paris, France
e-mail: laurent.tournois@biosilicium.fr

N. Lidströmer, Y. C. Eldar (eds.), *Artificial Intelligence in Covid-19*,
https://doi.org/10.1007/978-3-031-08506-2_13

Epidemic and preserve the state of health of populations: define the characteristics of the most affected populations, the risk factors, the degree of contagiousness and lethality, evaluating interventions to manage the outbreak … so many indicators and classic measures of epidemiology.

The development of digital technology, whether through the digitization of many aspects of everyday life, through the collection of the data generated, or through the processing of this data, has been a relatively new element in the context of this pandemic: to serve epidemiology, and also public health, new or simply faster, more reliable means or complementary to conventional surveys, were likely to be used.

Today, when we talk about data processing, we increasingly think of artificial intelligence. So here is an additional element still capable of modifying the face and the contributions of epidemiology, but also of the way of "doing public health" in the context of the pandemic. An intersection of epidemiology, digital technology and public health has even had a name of its own for a little less than 10 years: precision public health, a kind of counterpart to precision medicine or 4P medicine. For the sake of effectiveness, and even efficiency, there were hopes of resorting to precision public health, that is, to a set of targeted measures or whose estimates are on scales smaller than usually—for example, in terms of geographic information, and measurements of spatial indicators, monitoring of movements of populations or individuals. The development of contact tracing is an example of tools that can be used for precision public health, by providing information on a certain number of parameters on an individual scale, and at a much higher temporal and spatial frequency than in the past.

Also, we are interested here, beyond the concepts of epidemiology, public health, precision public health, in the conditions of existence of the use of AI in these two disciplines: digitization and decision support, developments around data and devices processing these data, in particular from a regulatory point of view and from the point of view of the real appropriation of the devices necessary for the operation of these new epidemiologies.

These conditions met, we are interested in what epidemiology, and public health, could expect from AI. We then look at one of the most historic and classic use cases of epidemiology, but also among the most critical in the case of a pandemic linked to an emerging disease: the information system under-tending the collection and analysis of medical causes of death. Where and how does AI intervene in this system? What are the specificities in the case of emerging diseases and Covid? We take the concrete example of the French system.

Beyond mortality, but also a few words on the modeling of infectious diseases, we more specifically address the identified use cases of AI, in the service of epidemiology and public health, in the context of pandemic. We identify a typology of six categories of uses, plus a seventh, emerging, called infodemics.

Finally, we place these findings in a broader context, and discuss some major issues around AI, epidemiology, and public health for the future.

The chapter is therefore divided into two main parts: a first part in which we briefly establish the evolution of epidemiology and public health until the digital

age, the conditions of existence of an epidemiology and public health taking advantage of digital and AI; a second part where we explore how all these elements are articulated and applied in the context of the Covid19 pandemic.

Epidemiology: Definition and Purposes

Cyrille Delpierre and Thomas Lefèvre

Multiple definitions have been proposed to define epidemiology. John Last, in the Dictionary of Epidemiology [1], has defined epidemiology as "The study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health problems." Depending on the population concerned a distinction can be made between clinical epidemiology, which is concerned with populations of "patients" with a view to improving the medical management of these patients, and population epidemiology, which is concerned with the general population with a view to developing public health strategies. As such, epidemiology constitutes a basic discipline of public health since it aims to quantify the health status of populations in order to identify its causes and to propose interventions to improve the health of the population concerned and to evaluate them. The nature of the purposes can allow to distinguish three main fields in epidemiology: (1) descriptive epidemiology which aims to describe the frequency and distribution of health phenomena or health determinants in populations, according to human, spatial and temporal characteristics; (2) analytical epidemiology which aims to identify and estimate the link between exposure to certain factors and the subsequent occurrence of disease (or health event). The object here is the search for determinants, to study the patterns, the "causes"; (3) evaluative epidemiology which aims to test the effectiveness of a health program or policy intervention on the disease or health condition under study.

Initially focused on the disease, the objects of investigation in epidemiology now correspond to any factor that may influence the state of health of the human being (i.e. biological, clinical factors, environmental exposures including physico-chemical, behavioural and psychosocial factors).

Whether it is to describe a phenomenon, to understand mechanisms, to make inferences or to test an intervention, evidence in epidemiology is based on the protocolized/standardized production of data. The basic brick of epidemiology is then the data that must be as valid and precise as possible to ensure that they measure what they are supposed to measure (validity) and limit measurement error (reliability). Protocols specify the volume and the scope of the data to be collected. The logic consists of working with a limited set of data that are precisely characterized and traced, and even certified.

In clinical epidemiology, these data most often come from clinical and biological data collected routinely in medical records, for example, or ad-hoc through the implementation of specific studies (clinical trials, cohorts), or from specific systems

that ensure the validity and exhaustiveness of the data, such as cancer registries. In population epidemiology, data can also come from trials, registries or cohorts, which have the advantage of including all the desired data but for a restricted population (the one included in the study) and from data routinely collected in different systems, such as medico-administrative data (reimbursement of expenses/medical acts), administrative data (civil status, taxes, family allowances, retirement, etc.), environmental data (air pollution) likely to provide information on the living conditions of a very large part of the population (or even the entire population). Once these data have been collected, they are then modeled. Epidemiology has always used numerical methods, including some of which are now labelled "machine learning", to produce results, whether to help describe, understand, or evaluate parameters hidden from observation, and to help with experimentation [2].

The definition, collection and analysis of data is therefore at the heart of epidemiology, which makes it a discipline particularly concerned by the development of what is called massive data and artificial intelligence (AI). The development of data digitization, including in the field of health (biology, clinical, imaging …), connected tools, social networks, opens a new era of availability of massive, varied data, never known before. These massive data may constitute opportunities to better describe, understand and intervene on the health of populations, and therefore a tool for epidemiology and its development. But massive data and their rapid processing through tools and approaches some of which being part of AI, raise also concerns.

Epidemiology and Public Health: How They Relate to Each Other and the Concept of One Health

Cyrille Delpierre and Thomas Lefèvre

Public health is both a collective project and a field straddling the science of the state of population health—how to define it, measure it, research its determinants—and the means of action to maintain and prevent the degradation if not improve the state of population health, political means included [3]. In this first introductory part, we see the main and characteristic features of public health and individual health—their articulation. It is possible to examine these two dimensions of health according to different models of representations, which will condition the type of knowledge and action that can be mobilized. Closer in time to us, we then discuss how the concepts of public health and individual health seem to be able to integrate certain technological innovations essentially, through respectively precision public health and P4 medicine [4]. Finally, we briefly illustrate all of these elements through an archetypal example in health and medicine, that of communicable diseases. We also understand, that both through predominant knowledge models and through the structuring of nations and contemporary policies, that a certain conception of health is no longer reduced to either individual health or population health, but extend beyond the human person and their societies, to encompass all the

dimensions of their conditions of existence: we are talking about the concept of One health [5].

Individual Health and Population Health

Since a population is made up of individuals, there are necessarily intersections between the state of health of a population and the state of health of individuals: to a certain extent, by preventing degradation, maintaining or improving the state of health of individuals, individuals can prevent, maintain and improve the state of health of populations—to a certain extent only, since this would amount to considering that a population is only the sum of individuals independent of each other, without direct interaction or indirect between them, or with their common environment. Indeed, in particular with constant and limited resources, favoring the state of health of certain individuals can have the consequences of degrading that of others, by competition for access to resources for example, or by degradation of the environmental conditions of other individuals.

The Articulation Between Individual and Population Level

Also, how do population health and individual health fit together? In terms of knowledge and in terms of means of action? Part of the answer lies in models of health representation, and what they say about it; we tackle them soon after. However, another part of the answer also lies in the "places" where individuals meet and the implementation of collective means for health: first and foremost, the health system. Public health, within the scope of its domain, includes the organization of the health system, both the training of health actors (medical, paramedical, administrative agents, executives, etc.), and the conditions of access to care (system health insurance, or social security more broadly), or the functional and geographical structure of places where we provide information or care [6]. Finally, the organization of the health system is also linked to the regulation of this system and its various components: regulation of tariffs and fees, regulation of types of acts, regulation of diplomas, regulation of innovation.

The health system is therefore this "place", in the broad sense, where both individual demand and collective response meet: we try to orchestrate access to individual resources, while aiming at collective health objectives. Health policies are thus an essential social element in the articulation between the individual and the population level.

We can also mention other systems which are independent of the healthcare system but which are likely to have an important influence on health, or even more important than the health care system [7], such as the education system, employment policy, urban policy, spatial planning, etc.

Biomedical and Biopsychosocial Models of Health: Individual, Environmental and Social Determinants of Health

Historically, there are two main models of health representation: the biomedical or biological model [8] and the biopsychosocial model [9]. There are others—variations, mainly, of the first two statements—but we will consider here only these two main models in the sense that they clearly shed light on the majority type of scientific production concerning health in all its dimensions (prevention, therapy, etc.), as well as the means of action thought out and put in place.

The biomedical model intends to define health on the basis of individual health, this health itself being defined and determined by the biology of the individual, at different scales: genetics at the most basic scale today, molecular scale, cellular scale, histological and anatomical scale, etc. This individual-centered representation is supplemented, to take into account everything that is not the individual, by the notion of the environment. To refer to health as a whole, we then speak of a gene × environment interaction model. It is therefore an initially both deeply reductionist and materialist representation.

Another model proposed in the 1970s by Engel is the biopsychosocial model. As its name seems to indicate, it defines individual health by taking into account at least three complementary dimensions: the biological, psychological and social dimensions. Note that if this model, at least these dimensions, differs significantly from the biomedical model, these two models are nonetheless models of individual health, taking into account differently on the one hand the environment, and on the other hand, the social contributions to individual health.

It is important to know which model we are referring mainly to, because the dimensions defining individual health being different depending on the model, this means that what we will seek to identify as mechanisms and causes of good or bad health of an individual or targets for potential interventions, regardless of the methodological means implemented, will not necessarily be the same from one model to another, nor considered to be able to interact together. Thus, in the biomedical model, the cognitive, psychological and social dimensions are seen as a consequence of biology, even if we do not know precisely through what means. It is not foreseen, initially, an autonomy of an object which would be the psychological life of the individual, which could maintain links starting from psychology to biology, for example. Psychological state therefore cannot be a determinant of health in general, nor a determinant of biological health. The same is true of the social determinants of health: possibly, in the biomedical model, the social can be absorbed in the environment, but does not benefit from a specific autonomy, nor from the status of an intrinsic component of health individual.

If the determinants sought are different, then the means of action deriving from them will also be. This has immediate repercussions in terms of health policies and regulation. Take for example the opposition between biomedical view where individuals are independent of each other, and biopsychosocial view, where individuals are by nature defined by social interactions. In the first case, one can tend more easily to policies which place a large part of the responsibility for the state of individual health on

the individual himself: beyond genetic chance, the responsibility of which can however be transferred on the parents, the individual is responsible for all the so-called modifiable factors contributing to his or her good or bad health, independently of the behavior of other individuals. Thus, if we recognize that smoking is bad for health, and that no one is supposed to ignore it, then contract a pathology which we can estimate with great certainty that it would be linked to the fact of having smoked, may be attributable to the sole responsibility of the individual [10]. This responsibility means that it may be no longer legitimate for the individual to benefit from collective and joint responsibility for the consequences of his pathology: the costs of care will be his sole responsibility. In contrast, in a model where the social component is an essential component if not a cause of individual health, it is no longer possible to immediately release collective responsibility for the deterioration of a person's state of health.

From the moment when the recognized determinants are not the same, and that some relate to the social or the environment, this implies that the policies to be put in place can no longer target only the individual, or even the health sector taken into account in a classic sense, but will necessarily concern, virtually, all dimensions of life in society: thus, the development of laws concerning housing, agriculture, the installation of certain activities, etc.

At the same time as this movement to broaden policies that we could qualify as participating in public health, we can see the desire to make more targeted, precise or individualized the definitions and means of action relating to population and individual health, essentially with a view to improving efficiency.

From Precision Medicine to Precision Public Health

The last 20 years have seen the emergence of concepts of P4 medicine [11]—this mainly in parallel with the development of clinical genetics and IT and digital means: digitization, storage capacities and calculation, accessibility of algorithms.

P4 medicine, for participatory, personalized, predictive and preventive, reflects fairly well the consideration of developments mainly technological rather than theoretical, by the biomedical model. It remains self-centered, and aims, through the use of new tools at its disposal, to be more effective in its means of action through increased relevance at the individual level. The participatory dimension is not a consideration of the social dimension per se, but above all a propensity to request greater participation from the individual in their health, passively (collection of personal data) or actively (education therapy, response to questionnaires, etc.); this goes in the direction of greater individual responsibility, as seen previously, even if these two evolutions are not necessary for each other [12]. P4 medicine is particularly based on the now almost ubiquitous nature of digital technology, whether in terms of data collection (via internet browsing, or simply the use of smartphones), storage and computing capacities (especially for data genome sequencing in clinical routine) or in terms of development and use of algorithms. It updates an old idea and desire, to have means of action that are more specific, more targeted and more suited to a given pathology and person. However, this approach is still mainly centered on

the use of large volumes of data, often biological (omics data) with the risk of reducing the individual to his genetic/biological characteristics and the illusion of perfect prediction. It is therefore presented as personalized medicine, but not as personal medicine that would take into account other dimensions likely to influence biological function (physico-chemical, behavioural, psychosocial exposures) [13].

The concept of precision public health is newer than that of P4 medicine. It is also largely based on the development and increased accessibility of technological means, mainly digital, with the aim here too of being more relevant and more effective in its prescribed actions. Precision can refer here to at least two quite different elements [14]. The most obvious here is to refer again to the individual. For example, health information, communication and education follow the same paths as advertising and marketing. Digital technology has seen the emergence of a new way of advertising, as practiced primarily by Google, but also Facebook to limit itself to these two examples. From broadcasting, or one-to-all, we have moved on to so-called targeted—or "personalized" advertising. How? Or what? By segmenting more or less finely the populations exposed to advertisements, for example on the basis of their internet browsing data or their expressed social links. Everyone using Google or Facebook is no longer uniformly exposed to the same ads at the same time, but virtually everyone is exposed to a "personalized" selection of ads. Public health messages have always faced the same limitations as advertising: how do you reach people who you think should be the first recipients of the message? Precision public health has or will have in its tools this a priori possibility of better targeting its prevention messages.

The second element covered by the notion of precision refers to precision in terms of scale, notably spatial, but also temporal. It has now been known for a long time by epidemiologists that ecological indicators can be dangerous, in the sense that they can show inverse links to those which might actually be observed at the individual level. Roughly speaking, it is as much a question of documenting certain measures and certain links between exposure and health event other than through an excessively imprecise average effect. An example is that of the birth and death indicators in certain African regions: an overall measurement can indicate an improvement in these indicators over time, while a measurement at a finer geographical scale, and more respectful of territorial heterogeneities or inequalities, can show a very different reality: the worsening of these indicators over time for a whole section of the population, and an improvement affecting only a minority [15]. Of course, the same phenomenon can be observed by varying the time scale of observation.

Epidemiology and Public Health in the Digital Era: Prerequisites

Thomas Lefèvre

Talking about AI in epidemiology and public health requires talking about the conditions of existence of this AI and its possible use cases. Indeed, from now on, it is no longer a question of imagining AI and its uses, but of examining whether

the practical conditions are met, today, so that its use is tangible. We are discussing four main conditions here, without claiming to exhaustively discuss all the determinants of the use of AI: we have announced it, no AI without data. Ubiquitous digitization brings closer to us the possibility of activating data by AI, for everyday actions or for already existing tasks, without mentioning new uses. We cannot use just any data, by anyone for any purpose: in recent years, many regulatory elements have evolved, particularly with a view to clearly defining what is legal or not to do in terms of data processing, but also to rule on the nature and place of AI, in particular in health. Third and fourth aspects, depending on the uses envisaged, that the pandemic has brought to light again, even if in the end this has been little taken into account: digital literacy, and the number of people equipped with digital access equipment, whether as a "producer" of data or as a user of data-based services. These two aspects have been particularly critical in the example of precision public health that has been the use of contact tracing and health pass applications.

A Ubiquitous Digitization

A fundamental element which makes or suggests that it is becoming possible to practice epidemiology and public health based on the use of more numerous, more varied, more complete, more relevant, more individualized and even more up-to-date data than what these two disciplines are historically constrained by the use of ad hoc studies—cohorts, case-controls, etc.—is due to an undeniable and generalized phenomenon: the growing and ubiquitous digitization of most of our human activities, passively or actively.

Of course, not all of the world's population is equally or similarly affected at all points, in particular there is this major risk of becoming invisible and therefore of an equally major bias, of a large part of the population: those who have little or no use of digital tools, possibly by choice, mostly by necessity. That said, even populations that do not have a smartphone, for example, or use it very little, the fact remains that their interactions with the social fabric: businesses, institutions, etc. leave more and more traces digital since these are becoming more and more "digitalized", all over the world.

Thus, the simple fact of shopping leads for many to digital traces on the side of merchants, physical or online; declare income and taxes too; have a salaried activity; consume energy; use the internet or a smartphone; have an activity on social networks, even minimal, even passive; go to the hospital or to a laboratory for additional examinations, etc. all this generates a little more data every day, a little more "precise"—precise does not necessarily mean "correct" here.

Whatever the quality of these data sources, they are nevertheless considered to be major opportunities for exploitation, both commercial and scientific, and the movement does not seem to be reversible.

The Evolutions of the Regulatory Framework on Personal Data

Another structuring element, necessary and making possible public action in particular, based on the use of data collected of various types and sources, is the evolution of the regulatory framework for the protection of personal data. As much as the private and commercial sector, in its use of legally miscategorized data, can or could do almost as it sees fit, including reselling the data collected to third parties, or offering "health" services of various kinds and extremely varied quality because without regulation or evaluation, the public sector is itself bound by its own rules. A particularly notable development in recent years is that of the development and adoption at European level of the General Data Protection Regulation [16], which contributes massively to redefining and constraining the possibilities inside and outside the European area, from the moment when services based on the use of data, collected or not from European citizens, can be addressed to European citizens.

The other important regulatory element is that which concerns algorithms whose claimed use is medical. It was necessary, it is necessary to decide the status of these algorithms, in particular, are they comparable to medical devices, and therefore fall under the same regulations, or should new provisions be provided, which potentially would respond to a differentiation between classical devices and algorithms? In the United States, we see the first wave of algorithms approved by the FDA [17].

Beyond the AI used in a context of diagnostic aid or clinical monitoring, for example, the question remains today as to a possible specific regulation of the algorithms which would certainly be intended for epidemiology, but above all for decision-making and public health in general. A recent example is the use of smartphones for contact tracing, and the algorithms that can be associated with it: to our knowledge, at best, these applications have "needed" no other authorizations for their massive deployment than those compliance with regulations concerning the use of personal data, which is a prerequisite, but certainly not sufficient.

Connected Devices and Equipment Rates

A third important point is that of a pivotal tool between data collection and means of action based on data and AI, from institutions—scientific or governmental—to people: these are connected devices, "medical" or not, the most typical example of which is the smartphone.

If personalized medicine can be based mainly either on devices intended for health professionals themselves, or on medical or similar devices intended for patients, epidemiology and above all public health, if they wish to achieve the greatest public, certainly currently have more to expect from non-medical devices, such as the smartphone: for example, it is very likely that in the long term, people will be more equipped with a multi-function device, without specific attachment to their health, only connected watches or wristbands.

The industry, but also part of the medical community or the public authorities, put forward for this a high rate of equipment in terms of smartphones: this was a more or less implicit argument for the deployment of contact tracing or health passes in the context of managing the pandemic. However, it is important to see the limits of this very general argument because it hides significant disparities in terms of the characteristics of the populations concerned. First, the equipment rate may not be sufficient in relation to the objectives pursued, particularly in a pandemic context. Indeed, the use of the smartphone for contact tracing requires, according to certain models, an equivalent of "vaccination coverage" to hope for an effect on the epidemic of at least 60% [18]. However, the rates of equipment can barely be of this order, which would therefore require that 100% of the people equipped adhere to the public health system, and more generally, this rate varies significantly according to age groups, within the same country [19]. In the case of Covid, this is particularly important, because if we want to target people at high risk, we must target the elderly: these are those with the lowest rate of equipment, and well below of the estimated necessary threshold.

Secondly, equipment rate does not mean sufficient and operational equipment, and even less homogeneous. A person may have a smartphone that is too old or does not have sufficient technical characteristics to be compatible with the deployed application. The diversity of models and operating systems, despite great progress in terms of standardization, is another obstacle to the deployment of a wide coverage application.

Thirdly, it is not because a person has a smartphone that this provides information on its actual use: many telephones can be in the name of a parent for example and be used by a child or other member of the same family. It may also not be used or used in a variable way and not adapted to the device envisaged—for example, the Bluetooth necessary switched off or active for something else during risky journeys. Or, quite simply, it can be turned off.

Digital and E-health Literacy

Finally, when we are in a situation where the digitization of all our activities is growing, if not omnipresent, when the regulatory framework concerning the conditions for collecting and using data is refined almost everywhere in the world and finally when the population general is very widely equipped with devices allowing the collection of data but also their use by AI, a fourth major point is that of the literacy of populations: digital literacy, but also literacy in health, a fortiori in digital health. This point is not as critical from the point of view of passive data collection: it is often sufficient for people to be equipped with a device which is known to be a minimum user, whatever the use (the smartphone for example), only from the point of view of active collection, and even more, from the point of view of the use of digital technology for public action.

If there is already a first divide in the population between those who have access to digital technology—physically—and those who do not have access to it, there is a second one based on the skills of these populations to use digital resources. This second divide is doubled in this specific case: it is a question of people having digital skills but also in health. Here again, the example of contact tracing or more broadly, the use of smartphone applications as public health tools in the context of the pandemic, is enlightening: a very large number of people are needed who are certainly equipped with a smartphone, but just as many who are able to use it, and to understand the issues and the health information conveyed. However, to date, these two literacies certainly remain very insufficient to envisage a wide-ranging action in the general population, and often depend very much on age [20].

Towards a Real Life Use of AI in Epidemiology and Public Health: Some First Examples

No Data Means No Artificial Intelligence: A Few Words About Data Federation and "New" Types of Data

The widespread and large-scale development of AI algorithms relies on an increased access to relevant data. This is necessary to develop the algorithms, to put them into production and to adapt them to work environments, by interfacing them with pre-existing information systems. AI appears to be a useful tool for organizing and federating data, as well as for processing and analyzing data. The diversity of data sources presents an opportunity for public health since there is often a lack of relevant data such as: (1) real and repeated data, (2) environmental data (physico-chemical and psychosocial exposures), (3) behavioral data and, more specifically, (4) data on exposure. AI in public health could benefit from data from geographic information systems and more broadly from geospatial/geolocated data [21], data on digital footprints and individual activity, data from social networks and online forums [22], and data from connected objects, with an emphasis on passive data collection. These connected objects can include medical devices (implanted devices), as well as non-medical devices, including smartphones.

Citizens and Patients as Producers, Actor and Manager of Their Own Health

Like other sectors basing their management and steering on data, it is estimated that one of the best sources of data for AI in public health would therefore be the people themselves, whether they are sick or in good health. Everyone is concerned by this collection of individual data, in particular in epidemiology, since the identification of associations between exposures and health events requires having both cases (patients) and control persons (healthy). Exposures can be biological, or they can be

social, cultural, economic, behavioral, etc. Each citizen thus becomes a potential "producer" of data, active or passive, while also becoming a "producer" and promoter of their own health. We can thus identify a real loop, which would start from the individual who generates data from his activities. This data is added to a pool of data collected in a similar way from other individuals. The data pool can be analyzed through AI techniques, and the circle can be closed when the original data "producer" is contacted, in the form of information or incentives regarding their health. A classic example would be the use of this model to improve therapeutic education [23]. Of course, AI can operate at a much higher level than the individual.

At the Population Level, Health Surveillance Systems and AI

There are a few pre-existing examples of the use of AI in the public health field, e.g., for surveillance of pathologies or health events, such as epidemics and the inappropriate use of pharmaceuticals or drugs, and of medical devices [24]. A textbook case is that of Google Flu Trends, which is no longer used today. The Google Flu team said the algorithm was able to predict flu outbreaks faster and more reliably than the Centers for Disease Controls (CDC), based on searches made by users of the Google search engine [25]. Several studies have highlighted the advantages and disadvantages of the Google Flu Trends algorithm, showing that it was not very precise, among other things. Rather than talking about the Google Flu algorithm, we should also talk about the algorithms, since the algorithm has been replaced several times, without it being known in what and how it has been modified.

Another example, monitoring the use of drugs and medical devices after they have been approved and placed on the market, which are increasing in number each year, is an important public health issue. Controlling each product systematically, manually and on a case-by-case basis, is not a realistic or viable solution. Without replacing human operators, AI can help analyze signals to determine which should be followed preferentially. However, more recent events have shown us that currently, and probably for a long time to come, the best early warnings rely on discerning humans being in the right place at the right time. It was not the data monitoring signals that sounded the alarm about the COVID-19 pandemic, but a human medical doctor.

Between the Individual and the Population, Healthcare Systems: Learning Healthcare Systems (LHS)

Yet another example relates to the healthcare systems themselves, with which individuals and populations interact. What are called learning healthcare systems (LHS) [26] are currently being developed, based on the use of data, and to a greater or lesser extent, AI. The concept of LHS was developed from 2007, in a context and with a goal of continuous improvement of the quality of care, the organization and the adequacy of the care provided by a health establishment to meet the needs of the

population concerned. These LHS were made possible by the prior collection of data, from medical and paramedical activities and practices (we speak of evidence-based practice) and are based on the use and processing of electronic health records and the use of flexible and continuous research capabilities. Because an existing data warehouse or data pool is required for an LHS to work, we see again that AI can be used at different points in the system: from data collection to specialized processing, their federation and their formatting. AIs used in an LHS can exhibit a varying degree of autonomy: from very high (full) autonomy in the data mining layers, to much lower (supervised) autonomy in the higher layers, where decisions are taken. The prospects offered by LHS still require human intervention, both technical and organizational, with the promise of a more responsive and less fixed system. Although the idea is interesting in principle, we currently lack sufficient information to accurately assess the realities and effectiveness of LHS.

What Contributions Could Be Expected from AI in Epidemiology and Public Health in the Context of a Pandemic?

Thomas Lefèvre

We have seen that the conditions for the existence of epidemiology and public health exploiting the possibilities of digital technology and AI are wholly or partly met. From these conditions, some concrete and current examples have been depicted. Given the goals pursued by both epidemiology and public health, what contribution can we expect AI to make to these two fields, particularly in the context of emerging or infectious diseases?

What Is Due to the Use of Non-classical Data Sources, and to the Comparison or Cross-Checking Between Sources

If we base ourselves on a sufficiently broad meaning of AI, taking into account, for example, natural language processing techniques within its scope, then we must recognize that in epidemiology and public health, as in any other field, there are areas or uses that have developed, and which have been made possible solely by the development of AI, and more broadly, by the development of certain digital-based activities.

This is the case for the large-scale analysis of data from social networks and the activity of people on these same networks: the links between people, in time and space, the analysis of feelings, the typology of dynamic networks of people or themes, the analysis of controversies … More generally, it is the analysis of vast

corpuses of textual data which has been made possible, which comes under "text mining": the analysis scientific literature, as well as the analysis of patient forums.

We owe, and will certainly owe AI even more tomorrow, the increase in our ability to match databases or data sources with each other, especially when they concern data of various kinds: voices, images, texts, etc. lack of native interoperability, through the adoption of common standards at the origin of data collection, can be partially filled by the use of AI. Health data warehouses are, for many, made up partly of structured data as soon as they are entered, but also of data extracted from textual data, electronic health records (EHR). In fact, all of these advances, which are not very visible for the researcher or the citizen who only sees the final database or the results of the studies, contribute to improving for some of the studies their generalizability or statistical power if is relevant, and to speed up the collection of additional data—or quite simply, to complete certain surveys with additional data that already exists, but is usually not easily accessible, in its native form.

The Real-Time or Refreshable Nature of the Information

Epidemiology, and by extension public health, is historically and traditionally based on the implementation of ad hoc surveys, specific to a problem to be explored—for example the search for explanatory factors of a given disease. This has not, of course, prevented them from sometimes resorting to already available data sources, typically census data, or exposure data collected by monitoring organizations, or even more recently in history, medico-administrative data, healthcare reimbursement. Epidemiology has also gradually deployed major population tools, which are increasingly general and bring together large numbers of people within population cohorts.

All these sources no doubt retain and will retain their relevance for a very long time to come, because in particular for ad hoc surveys and cohorts, these are systems whose implementation we control, the measuring instruments—in short, we control the experience and its quality. They will retain at least their reference value. These devices nevertheless remain expensive, and above all, can have characteristic times for collecting new data or simply updating these data that are very difficult to reconcile with the requirements of monitoring an emerging phenomenon, and of quantitative and above all qualitative evolution quickly — as has been the case with this pandemic. The digitization and processing by AI of data resulting from digitization is an opportunity to be considered by epidemiology and public health, in addition to other existing sources, for the constitution of information systems adapted to this type of phenomenon, and its refresh which can approach real time in some configuration. This is the case, for example, for the analysis of social networks, or of the spatial nature of such and such a variable, such as the study of the mobility of populations on a large scale via the analysis of telecommunications data or satellite images, or contacts between people on a finer scale (for example: contact tracing).

New Types of Analysis, in Particular on Massive, Incomplete or Even Poorly Balanced Data

It may have been surprising for some epidemiologists to see that several governments have based their political decisions on "SIR" type models of the spread of the epidemic. SIR models are so-called compartmental models, each letter here representing a compartment of the population studied: S for Susceptible, I for Infected and R for Recovered [27].

Several variants exist, in particular with other compartments including a deceased compartment in the case of a particularly lethal infectious pathology. These models can be deterministic or stochastic, and it is from these models that the famous parameter R0 comes, known as the basic reproduction number (or ratio), which reflects the number of new infected expected for an infected observed. Political decisions can be centered around this R0: a threshold value of R0 corresponding to a decreasing rate or extinction of the disease in the population can be an objective to be reached by different health policy measures; it can be an indicator to monitor reflecting the intensity and type of phase of the epidemic …

However, these models are quite old, which is not a fault in itself, but we know their advantages and disadvantages quite well. A definite advantage lies in the good mastery of their mathematical analysis and interpretation. Several major flaws, however, certainly make them unsuitable tools for crisis management, when faced with an emerging disease. First, like any model, it is necessary to be able to estimate its parameters, and the R0 is particularly difficult to estimate reliably in a short and prospective time, a little more easily retrospectively. It depends on the ability to screen and diagnose the infected, and beyond the availability and performance of screening and diagnostic tools, the number of asymptomatic but contagious people, or similarly, the period of contagiousness asymptomatic. Next, these are models which, unlike the majority of epidemiological tools, make it difficult to adjust or stratify population characteristics: for example, sex, age, comorbidities, socioeconomic position or country of birth. This is inevitably a major problem in the face of an infection that obviously targets and kills people very differently according to these characteristics. In addition, another major flaw is that these models are not or very rarely used in a spatialized way, which is also an important problem, in the sense that the beginning of the pandemic's history has clearly put highlight very delimited regional clusters in different countries, even though there was no containment measure or ban on population mobility at that time. Finally, another defect highlighted by other approaches, the treatment even by compartments: everyone in the same compartment ultimately has the same chances of being infected or contagious. However, it seems that there are people, for various and varied reasons, who are much more than others at the origin of a large number of contaminations, while other people will only contaminate little if any person [28].

In the age of digital technology, AI but also precision medicine or precision public health, the use of this type of model for public decision-making seems strangely conservative, whereas other techniques based on AI necessarily emerge in this type of context [29].

More generally, beyond modeling, much is expected of AI to manage more agile and more robust data sources that usually pose problems for conventional epidemiological techniques, such as large volume data, presenting a lot of missing data (which is almost an inherent characteristic of big data), or even data whose completeness is far from being balanced according to the sub-populations studied, which unbalances the analyzes and the quality of the results obtained according to these subpopulations.

Aid in the Search for Causality or Networks of Causality

The question of causality is a central question in epidemiology, and a fortiori in infectious disease. Otherwise, the issue of risk or vulnerability factors is just as important in a context where, failing to control the chain of transmission and the specific causes, it may become possible to implement targeted preventive and protection measures.

Causality in medicine and public health remains something delicate to establish and relies mainly on a body of arguments and evidence for it to be considered established. It relies heavily on the experimental device, of the randomized controlled trial type, or any other quasi-experimental device in the population. These are obviously systems that are either impossible to set up for ethical or practical reasons in certain cases, or too difficult to set up, in particular in the context of an emerging, lethal and rapidly progressing disease. Moreover, the concept of single, direct and fully deterministic causality is difficult to hold in medicine and epidemiology, where it is more considered that there is a network of causalities, direct and indirect, mediated or not, that the phenomena are multifactorial.

Obviously, the more certain we are of the causes and the precise chains of causes of a phenomenon, the more legitimacy the decisions and public health measures to be put in place will have. Our current capacities to establish the causes of a phenomenon being limited and poorly adapted to crisis situations, it is logical to wonder about the capacities of AI to help us identify such causes or even the networks of causes to from available data, particularly routinely.

Evaluation Methods Based on Observation or Quasi-Experimental, as for Virtual Clinical Trials

Any health intervention, including population health, requires an evaluation of its impact in order, on the one hand, to validate its effectiveness, if applicable, and, on the other hand, to be able to adjust its methods as needed. This good practice present

in any good epidemiology and public health textbook is certainly one of the least followed, and by far. This type of evaluation is certainly costly and difficult to implement, and can have the side effect, which is sometimes embarrassing for decision-makers, of showing that despite the resources committed, the effectiveness of the intervention is very modest. The economic and image profitability of this type of evaluation is therefore assumed to be very low.

Methods bypassing the need for experimental settings that is expensive and complex to set up, based on the use of data already available, for example observational data, have been proposed. For example, we are talking about virtual clinical trials [30]: the main idea is that there already exists, without necessarily having to carry out new ad hoc studies, a large number of configurations recorded in databases, whether separate or not. It is then a question of bringing together all these necessary configurations, organizing them in such a way that one can, for example, reconstitute the equivalent of control and intervention groups, then analyzing the whole as in the context of a clinical trial. Obviously, as it stands, even if such groups can be reconstituted, at least one main quality specific to controlled trials is lacking, namely the control of a set of individual or contextual factors capable of biasing the results obtained.

By analogy, we can estimate that there are different types of interventions attempted in different configurations, in the "natural" state, without intervention in this sense, and which can be exploited in order to "virtually" evaluate the impact of these interventions in the population on the pathology studied.

The Augmented Expert in Public Health

In connection with the question of causality and our ability to base decisions and interventions on the identification of chains of transmission, as in medicine and surgery, we can ask ourselves the question of an "augmented expert" in public health. This is the theme of human-machine hybridization, one complementing the other in such a way as to improve the performance of the first, or in a virtuous and circular way, one reciprocally.

Also, all the techniques considered in the previous sections can contribute to the development of this augmented expert, by providing it with resources largely based on data and AI: assistance in the analysis of large volumes of data and of various natures, frequent updating of data and indicators, taking into account sentiment analyzes such as opinion polls, carrying out numerous quasi-experimental studies or virtual clinical trials to develop and evaluate in silico the estimated impact of different public health measures …

In the end, it would be just a very technologist update of the concept of evidence-based policy.

Some Specificities and Examples in the Context of the Pandemic: Epidemic Modeling, Public Health and Counting Deaths

Claire Morgand and Thomas Lefèvre

Epidemic Modeling: How to Retrieve Accurate and Timely Data to Feed a Model

If we want to use the modeling of epidemics as decision support tools in public health, then a certain number of constraints appear and weigh on these models, as they weigh on any AI algorithm which would pass from the prototype stage to the stage of routine use.

First, the model must be able to update and adapt to new data, specific to the current phenomenon, therefore here, the emerging epidemic. There is therefore always this compromise between a model that is very specific and adapted to a particular epidemic, and a model that is general and generalizable enough not to be used only once, or only retrospectively.

The second problem relates to the nature of the data needed to adapt and "feed" the model, so that it is useful in anticipating the evolution of the epidemic and testing different intervention hypotheses. It is therefore necessary to access the data relevant for the estimation of the model, and this in a repeated way and in the most reactive and least biased way possible. However, in infectious disease, a fortiori emerging, we are almost always a step behind, because we can only measure what has already happened. A difficulty is added to this already strong constraint: the observability of the cases. Indeed, in any infectious disease, there is a variable proportion of contagious, but asymptomatic cases, which will not be counted correctly, since precisely not observed.

Thus, our capacities for modeling and evaluating and updating various indicators specific to the surveillance of the epidemic are thereby limited. But these considerations even affect a measure that nevertheless seems simpler and more accessible, while being fundamental in this context: the census of the dead.

Counting Deaths in the Context of a Health Crisis Based on Information Systems and AI

Claire Morgand

Challenges of Counting Death, the Case of the French System

Historically, counting deaths and births has been the birthmark of statistics, in the sense of State accounts, and as tools for the government of populations [31]. In our time, one can imagine that the identification of the number of deaths in a given

territory is a settled, well-ordered affair and that beyond that, the systematic search for medical causes of death responds to a well-established system. We often forget that the things that seem the most natural, acquired or simple, are in the end not as robust as one might expect. At all times of an epidemic of an emerging disease, a fortiori seeming to present a significant lethality, being able to correctly count the deaths attributable to this disease, is a major issue. It is also quickly one of the only indicators around which public intuitions will communicate to give a measure of the epidemic.

We present here the case of the French system for registering deaths and identifying the medical causes of death, in particular in the light of a pathology and an epidemic such as that at Covid.

In France, when a death occurs, it must be declared by a doctor. The doctor completes a death certificate that contains an administrative component and a medical component. The medical component informs a morbid process and the possible associated causes of death. From this information, coding rules defined and maintained by the WHO (vol 2 of the ICD 10 version 2016) are applied to retain the underlying cause of death (the one appearing in the national statistics). The models of death certificates and the coding rules are used by 43 countries, the objective being to have a temporal and spatial comparability of causes of death. In France, the coding of medical causes is centralized at Inserm and an expert system associated with coders encode all the death certificates occurring in the territory. Overall, the production stages are illustrated in Fig. 13.1.

The deadline for finalizing an annual database now exceeds 3 years. Indeed, between the collection of death certificates, which can take several months (between the transmission of certificates and the digitization of them) and the coding now carried by a mixed system, the delays in making the finalized database available are important.

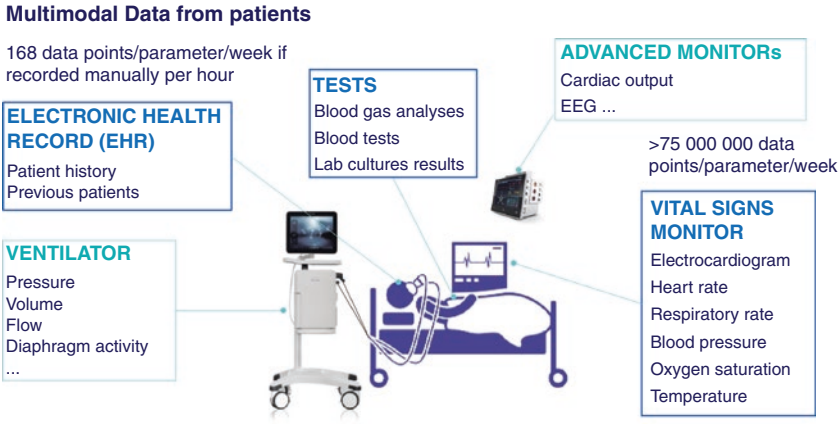


Fig. 13.1 Different stages for the annual production of the medical causes of death statistics in France

To define the role that AI can play in coding, it is first necessary to have a good understanding of the difficulties related to certification and the difficulties related to the causes of death proper (WHO rules not always medical, rare causes, new causes, etc.). Secondly, it is also a question of defining well the cases of uses where AI could have an interest (frequent, recurrent causes and annual monitoring of the main major families of causes of death) and those where AI will not necessarily have a great added value (rare causes, new causes type emerging disease and therefore health watch). The current data production process does not allow for a weak signal analysis.

Certification of Deaths and Collection of Death Certificates, Specificity to Emerging Diseases Such as COVID

The physician reporting a death can do so in two formats: a paper format that is still mostly used (Nearly 80% at the beginning of the COVID crisis versus 70% in January 2022) or an electronic format that is not widely used in France. The death certificate, whatever the modality of the certificate has an international format and the doctor must be able to freely enter the text of the medical component and not be helped or influenced (WHO instructions). As a result, two issues emerge during a health crisis:

1. The delays of transmission from the medical component to the coding centre are on average of 4 months for the paper format which implicitly makes it difficult to envisage a health watch from this certification modality. They are shortened to a few minutes in case of electronic certification.
2. Since the physician must freely certify, the variability of the medical terms used by physicians to certify is very important (for example: to certify a death by myocardial infarction the dictionary of the coding aid software contains more than 400 different entries). This type of variability makes the supervised approach of Natural language processing (NLP) models difficult. Moreover, the majority of certificates being handwritten on paper certificates, important readability problems prevent correct coding of all terms present on the certificate.

So if the goal was to be able to track COVID-related mortality in a very short period of time, it is only possible at the outset to rely on electronic certificates which are not distributed across all sectors of care in a balanced manner since the vast majority of them are issued by health institutions and the only deaths occurring in the hospital are then detained. In France 60% of annual deaths occur in hospitals, so that 40% occur in cities (medico-social structures or home of the patient, public roads etc.).

However, for another objective such as the assessment of annual excess mortality due to COVID or the impact on annual mortality of COVID on other causes of death, the certification modality should not have an impact on the production of this data (subject to legibility).

Issues in Coding Causes of Death, Especially in Emerging Diseases

To see the potential interest of AI in coding medical causes, it is important to understand how death certificates are now coded.

Figure 13.2 illustrates the current strategy.

After the causes of death present in the certificates are entered by an operator in the coding and scanning software (for paper certificates), an expert system applies the decision rules imposed by the WHO to identify the initial cause of death. To do this, the software must first assign to each certified medical term an ICD code and then in a second time apply the decision rules (from thousands of tables of probable causal links or some between the different codes) and define the underlying cause of death. If the tool cannot define the cause of death or cannot assign an ICD code, then a human coder must intervene on the certificate. This can simply be a medical term not present in the tool dictionary (see important certification variability) that the coder must implement in the dictionary with the correct ICD code(s). It may also be difficult for the tool to establish causality between the different causes of the certificate, as this does not allow it to definitively define the cause of death. It is then up to the coder to make the decision. To date, the tool codes about 50% of death certificates in autonomy, so there are still 50% manually coded each year.

In view of these explanations, it is then simple to understand that when a new pathology emerges, the WHO must integrate one or more ICD codes to specify the pathology, that it also define the decision rules that will make it possible to retain pathology as a cause of death or not, and that the software be updated with these elements to be able to code and allow the identification of deaths related to covid.

To date, the coding process is therefore a mixed system, consisting of an expert system associated with human intervention and dependent on decisions taken by the WHO.

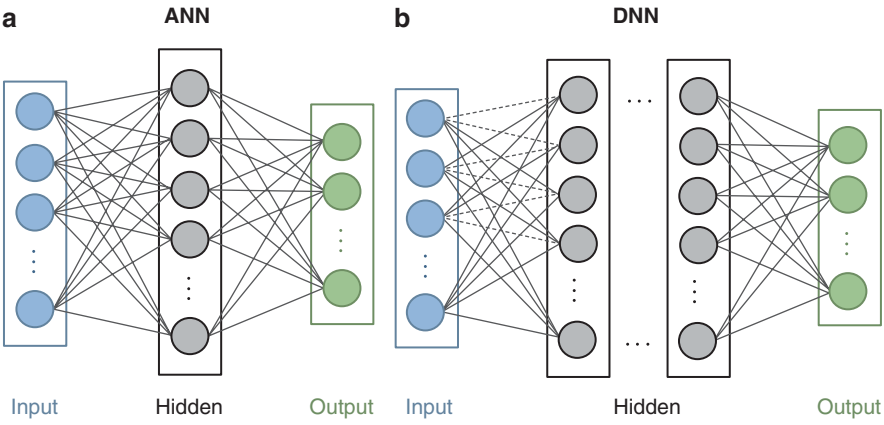


Fig. 13.2 The different stages of coding medical causes of death (French system)

AI's Place in Coding Deaths on Emerging Causes/Diseases, COVID Contextual Example

With the advent of COVID, three strategies to improve coding automation were developed:

1. Improving the standardization of terms through the coding tool to be as comprehensive as possible on physician-certified terms to best complete the first step of coding: the assignment of an ICD code per term placed on the death certificate. The expert system has this ability to standardize (step of NLP) but it is implemented by humans.
2. The application of supervised NLP methods to search for defined words (COVID, Coronavirus, SARS COV, etc.) in certificates to identify those mentioning these terms. Several limitations appear immediately with this approach: the only identification of these terms on electronic certificates (and therefore the only identification of certificates mentioning covid in hospital); the only mention of Covid that does not prejudge the direct link between mention of terms and death of Covid in a certain way and finally associations of "negative" terms such as "negative Covid test" which require adjusting the search algorithm as needed to avoid this type of false positive. For example, the first tests using this strategy overestimated by about 20% the COVID-related deaths occurred directly. So in order to support decisions during a health crisis, an overestimation may not appear as a disaster, nevertheless it is impossible to evaluate the impact of the measures taken on this mortality with this strategy.
3. The use of deep learning methods to identify these deaths. Independently of Covid, these methods have been tested on a body of death certificates for the two steps previously described: assigning an ICD code by term and identifying the original cause of death. Figure 13.3 illustrates the evolution of coding strategies.

Deep learning methods have obviously shown an interest in reproducing the coding produced by the reference tool when the causes are frequent, known and simple rules. Indeed, this type of approach requires a learning corpus and reproduces what he has learned from the existing corpus. The tool's performance in assigning the correct ICD code to each term (NLP-MCD) was very good (95.2% vs. 82.5% for the usual coding tool [32]) and for the identification of the initial cause (AI-UCD) also: 97.8% versus 90% for the usual system (once entered correctly [33]).

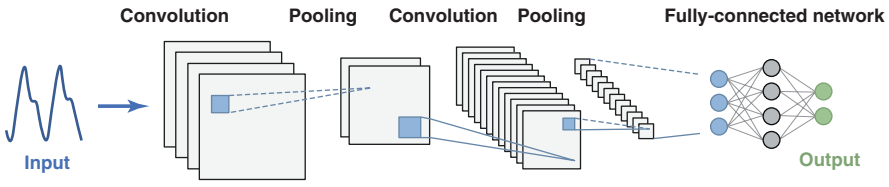


Fig. 13.3 Use of artificial intelligence in the evolution of the coding of medical causes of death (French system)

These performances are obviously very good but to be tempered because they vary according to the frequency of the pathologies (less good for rare pathologies) and especially are not good for new pathologies or for deaths responding to complex coding rules such as external causes (accidents, homicides, suicides, etc.). As a result, when an emerging pathology such as Covid occurs, the tool does not have a learning base and it is not possible to reliably identify the deaths involved. Coding rules for Covid cases being transposable to influenza-related deaths, tests have been done in this direction at the beginning of the pandemic with controls by an expert coder of the outputs proposed by the tool but the sensitivity and specificity of the methods being unconvincing, these were abandoned.

To Take Away

The use of NLP methods for the identification of certificates indicating a new cause of death (without certainty that this cause will be properly retained as a cause of death) could improve health monitoring but by overestimating deaths related to this cause and especially by overrepresenting intrahospital deaths since almost the only ones certified electronically. Moreover, the application of methods of standardisation of terms would improve the share of certificates automatically coded by the expert system, when it has integrated the ICD codes and newly created decision rules for an emerging pathology. Nevertheless, despite this possibility of improvement, the production delays inherent in paper certification do not allow to date to envisage a monitoring of death causes on this method alone.

Deep learning methods have shown a very clear interest not only in assigning ICD codes to each medical term present on the certificate but also in identifying the underlying cause of death: However, performance is good for common causes, already existing and affected by simple rules of decision, so that considering this type of approach alone to achieve the monitoring of causes of death on an emerging disease is today inappropriate.

To consider monitoring pandemic mortality, it is therefore:

- To improve the electronic certification rate by expanding it to all health structures and death occurring outside hospitals—To improve the automatic coding part of the expert system (standardisation of the terms of the expert system dictionary) at least on common causes of death to leave a smaller share of non-coded certificates to coders, including those related to rare or emerging causes.
- To improve the automatic coding part of the expert system (standardisation of the terms of the expert system dictionary) at least on common causes of death to leave a smaller share of non-coded certificates to coders, including those related to rare or emerging causes.
- To consider coding using a mixed method extended to deep learning, these methods having shown their interest in the usual and frequent causes of death: Thus, the deep learning tool could code certificates displaying the classic causes (especially since the tool can give a validity score of the proposed underlying cause); thus reducing the share of certificates requiring coding by the method applied today (expert system/human coders).

Future strategies to improve the production of medical cause of death data will need to take into account the arrival of ICD 11, which should combine a new coding aid tool using these three approaches (deep learning + expert system + human coder).

AI at the Service of Epidemiology and Public Health in the Context of the Covid19 Pandemic

Hélène Colineaux, Laurent Tournois, and Thomas Lefèvre

In this part, we discuss the uses of AI that have been effectively proposed in the service of epidemiology and public health, in the context of the Covid19 pandemic. To do this, we relied on a targeted bibliographic search: we decided to essentially explore the scientific articles indexed in PubMed, and of the following three types: reviews, systematic reviews or meta-analyses. Of course, this type of review has limitations, related to the type of articles searched, as well as the database consulted. However, our goal here was not so much to identify a very specific use of AI in the context of the pandemic, but more to identify a typology of the proposed uses.

Based on the results of this research and the identified typology of published uses of AI, we then describe few more specific examples, in order to present their performance as reported in the studies, and to measure their compliance with the TRIPOD/CHARMS guidelines. We finally estimate their degree of maturity, through an assessment of their Technology Readiness Level (TRL).

Bibliographic Search

Hélène Colineaux and Thomas Lefèvre

We carried out a bibliographic search on February 28, 2022, using PubMed. The search equation used was as follows:

((“epidemiology”[All Fields] OR “public health”[All Fields] OR “prevalence”[All Fields] OR “incidence”[All Fields] OR “risk factor”[All Fields] OR “modeling”[All Fields])
AND (“covid”[Title] OR “sars cov*”[Title] OR “pandemi*”[Title])*
AND (“artificial intelligence”[Title/Abstract] OR “machine learning”[Title/Abstract] OR “deep learning”[Title/Abstract] OR “prediction”[Title/Abstract]))
AND ((meta-analysis[Filter] OR review[Filter] OR systematicreview[Filter])
AND (2020/1/1:2022/2/22[pdat])).

It’s a relatively simple equation, but one that has become more relevant now than 5–10 years ago, when machine learning, for example, was not necessarily a term appearing in the title of articles or in keywords. Previously, it would have been necessary to include a whole list of algorithm names to hope to cover the maximum possible cases.

This equation returned about 200 articles. We excluded articles that were clearly off-topic and those that were exclusively related to biomedical research (image processing, prognosis, diagnosis or therapeutic study, etc.). In general, in paper reviewing the use of artificial intelligence in the context of the COVID-19 pandemic, the assistance of clinical practice takes a large place: automatic analysis of radiological images, diagnostic or prognostic algorithms, drug discovery algorithms, etc. Besides, the main reason for exclusion of articles in our review was that these papers exclusively reviewed uses for clinical practice. These uses allow to optimize human and material resources in this context of important demand of care, so they could also be seen as kinds of public health tools, but we focused here only on direct uses for population health.

We finally identified almost 60 articles that really relate wholly or partly to epidemiology or public health and AI. We explored in more detail 23 of these articles [34–56], which we report in Table 13.1, as the other being only partially dealing with these aspects or with a too large scope (uses of "digital technologies").

From these selected articles, we have identified a proposal for a typology of these uses of AI.

A Typology of AI Use at the Service of Epidemiology and Public Health in the Context of the Pandemic

Hélène Colineaux and Thomas Lefèvre

From the different applications described by the authors in these articles and the classifications they used, we proposed a typology regrouping all the identified uses of AI in epidemiology and public health. We describe a typology in six categories, plus an emerging seventh (see Fig. 13.4). These categories can be more oriented or exclusively epidemiological, or more relevant to public health—the border can be porous, these two disciplines relying on each other by history and construction.

Outbreak Monitoring

The AI has been applied on detected positive cases [57], daily hospitalizations [57], social media activity [58–62], wastewater analysis [63], etc., to visualize, predict, characterize the spreading of the virus, in a short time, or even in real time [64]. It was besides mentioned that the emergence of the pandemic was early detected by a Canadian AI-based surveillance system [65]. The detailed uses of this category, with examples, are:

- Documentation of the number of cases, deaths or vaccinations, in real time, like in the "global map" platform produce by the Johns Hopkins university and medicine [66].
- Prediction of Covid-19 evolution: size, lengths, starting and ending time of peaks, total number of cases, of hospitalizations, of deaths, etc. [67–80]. It might

Table 13.1 Twenty-three articles retrieved and analyzed thanks a bibliographic search to explore the variety of AI use in epidemiology and public health

	Authors	Title	Revue	Year
[34]	Abd-Alrazaq et al.	Artificial Intelligence in the Fight Against COVID-19: Scoping Review	J Med Internet Res	2020
[35]	Abdulkareem et al.	The Promise of AI in Detection, Diagnosis, and Epidemiology for Combating COVID-19: Beyond the Hype	Front Artif Intell	2021
[36]	Adly et al.	Approaches Based on Artificial Intelligence and the Internet of Intelligent Things to Prevent the Spread of COVID-19: Scoping Review	J Med Internet Res	2020
[37]	Asadzadeh et al.	Information technology in emergency management of COVID-19 outbreak	Inform Med Unlocked	2020
[38]	Barbieri et al.	How Artificial Intelligence and New Technologies Can Help the Management of the COVID-19 Pandemic	Int J Environ Res Public Health	2021
[39]	Bragazzi et al.	How Big Data and Artificial Intelligence Can Help Better Manage the COVID-19 Pandemic	Int J Environ Res Public Health	2020
[40]	Chang et al.	Application of artificial intelligence in COVID-19 medical area: a systematic review	J Thorac Dis	2021
[41]	Chen et al.	Artificial Intelligence for COVID-19: Rapid Review	J Med Internet Res	2020
[42]	Dogan et al.	A systematic review on AI/ML approaches against COVID-19 outbreak	Complex Intell Syst	2021
[43]	Guo et al.	The application of artificial intelligence and data integration in COVID-19 studies: a scoping review	J Am Med Inform Assoc JAMIA	2021
[44]	Malik et al.	How artificial intelligence may help the Covid-19 pandemic: Pitfalls and lessons for the future	Rev Med Virol	2020
[45]	Naseem et al.	Exploring the Potential of Artificial Intelligence and Machine Learning to Combat COVID-19 and Existing Opportunities for LMIC: A Scoping Review	J Prim Care Community Health	2020
[46]	Napolitano et al.	Impact of computational approaches in the fight against COVID-19: an AI guided review of 17,000 studies	Brief Bioinform	2021
[47]	Shah et al.	Fusion of AI techniques to tackle COVID-19 pandemic: models, incidence rates, and future trends	Multimed Syst	2021
[48]	Syeda et al.	Role of Machine Learning Techniques to Tackle the COVID-19 Crisis: Systematic Review	JMIR Med Inform	2021
[49]	Syrowatka et al.	Leveraging artificial intelligence for pandemic preparedness and response: a scoping review to identify key use cases	NPJ Digit Med	2021

(continued)

Table 13.1 (continued)

	Authors	Title	Revue	Year
[50]	Swayamsiddha et al.	The prospective of Artificial Intelligence in COVID-19 Pandemic	Health Technol	2021
[51]	Rodríguez-Rodríguez et al.	Applications of Artificial Intelligence, Machine Learning, Big Data and the Internet of Things to the COVID-19 Pandemic: A Scientometric Review Using Text Mining	Int J Environ Res Public Health	2021
[52]	Tayarani et al.	Applications of artificial intelligence in battling against covid-19: A literature review	Chaos Solitons Fractals	2021
[53]	Uohara et al.	The Essential Role of Technology in the Public Health Battle Against COVID-19	Popul Health Manag	2020
[54]	Vaishya et al.	Artificial Intelligence (AI) applications for COVID-19 pandemic	Diabetes Metab Syndr	2020
[55]	Xu et al.	Artificial intelligence for COVID-19: battling the pandemic with computational intelligence	Intell Med	2022
[56]	Zhang et al.	Data science approaches to confronting the COVID-19 pandemic: a narrative review	Philos Transact A Math Phys Eng Sci	2022

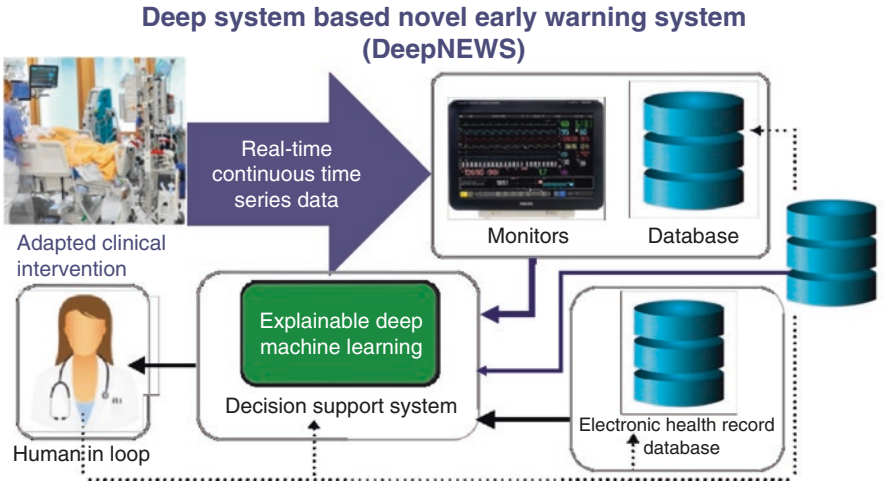


Fig. 13.4 Exploring the typology of uses of AI in epidemiology and public health: seven categories retrieved from bibliographic search

be the application which produce the more literature. Several types of methods can be used for forecasting as data-driven statistical model [81, 82], epidemiology-based compartment models [83, 84] or individual-based agent model [85]. For example, Car et al. [86] produced a worldwide model of the maximal number of patients across all locations in each time unit, using multilayer perceptrons. And Huang et al. [87] used Convolutional Neural Network to predicted number of confirmed case in several cities in China.

- Identification of more affected areas, at different levels (country, region, city) to geographically adapt measures or behavior to the risk [88, 89]. For example, Misra et al. [88] produced a navigation tool for avoiding COVID-19 hotspots.

Epidemiologic Outcomes and Characteristics Discovering

To informed the public health decisions and to parameterized some forecasting models [84], accurate and reliable epidemic knowledge had to be rapidly produced. For this aim, AI had also been used:

- Description of the characteristics of the virus: R0, incubation period, asymptomatic rate, mortality rate, etc. [90–92]. For example, Tessmer et al. [92] used multi-layer perceptron, convolutional neural network, and long-short term memory, to learn and estimate the R0.
- Description of these characteristics among groups, to identify vulnerable ones, improve spread predictions and identification of individual factor risks. For example, Li et al. [93] used Ridge model to identifying factor associated with cases and mortality rate, and Jehi et al. [94] identifying factors of COVID-19 susceptibility.
- Identification of environmental factor: for example, Pramanik et al. [95] measured the effect of climatic factors on the magnitude of outbreak and Travaglio et al. [96] estimated the effect or air pollution on infectivity and mortality rate.

Social Control and Monitoring

The fight against the pandemic has heavily relied on social measures. IA was used as a tool to implement and monitor these measures. We can separate two main types of applications:

- Contact tracing, i.e. contact identification, contact listing, risk assessment of contacts and contact follow-up, e.g. based on the analyze of tracked movements and interactions of the infected person [18, 97, 98]. Maghdid et al. [97] developed, for example, a smartphone-based approach to automatically and widely trace the contacts for confirmed COVID-19 cases.
- Controlling the norms application, i.e. the respect of social distancing norms [99–102], the respect of quarantine by contact or infected people [103, 104], the respect of face masks wearing [105]. For example, Ahmed et al. [102] used a technology identifying human in video to track social distance.

About social control, some authors have even suggested the use a network of tracking systems, technologies such as infrared cameras, data sharing systems, and IA to optimize city-wide monitoring and control (“smart city”) [89, 106–108].

Social activity was also used, not only to be controlled, but also to produce information for other purposes [109]. Internet searches or social media like twitter [110]

have been mining and learned to predict evolution of spread and its characteristics [58–62], to evaluate the perception and reaction of the population [89, 111–118], or to monitoring the mental health [119]. For example, Zhang et al. [118] used ML classification models based on deep learning language model to identify positive, negative and complicated emotions in social media conversations about COVID-19 in five languages. Koh et al. [120] have also mining Twitter to evaluate the expression of loneliness across diverse communities.

Assisted/Augmented Scientific Research and Knowledge Sharing

Knowledge had to be produce rapidly, and it had also been produced massively since the beginning of the pandemic. To really help the clinical and public health decisions, the imperative need to evidence synthesis has emerge [121, 122]. Relevant reviews are very time-consuming, and AI has been used to assist them, e.g. with mining the information [123] and automatically sorting articles [124], to rapidly characterize them and their complex relationships, e.g. with graphical representations [125], to automatically update the review and to identify priority topics for further research, addressing important knowledge gaps [126–129]. In addition to the analysis of the scientific literature, AI has been also used to evaluate of the quality and readability of online information about the coronavirus disease [130, 131].

Healthcare Resources Adaptation and Optimization

During the pandemic, the functioning of the health system was challenged for several reasons [47]: by the epidemic itself, due to the explosion in demand for care, which was out of all proportion to the usual health needs of the population and required rapid and extreme adaptation of healthcare system; by the reductions in the resources for care for reasons other than COVID-19 (closure of operating theatres, cancellation of "non-urgent" care, increasing demand of equipment, etc.), which also required adapting the supply of care, for example in psychiatry, cardiology, surgery, etc.; by the lockdown and isolation of population which created a disruption in usual treatments and follow-ups.

AI, and technologies more broadly, were seen as an opportunity to reduce the burden on health care system during the pandemic times, especially in Low-Middle Income Countries even more challenged by limited resource issues [45]. We can regroup these uses in three strategies:

- Reducing the demand of care, using general-public diagnosis software, informative or supportive "chatbots" [132–136]. For example, Battineni et al. [135] proposed an AI-based "chatbot" to evaluate diagnostic and recommend immediate measures for patients exposed to nCOV-19, in order to reduce need for medical specialist consultation.

- Reducing the need of material and, above all, human resources in practice, by using AI-assisted telemedicine, diagnostic assistance software, automated decision-making processes, patient monitoring framework, etc. [137–141]. Hallak et al. [141], for example, reviewed the application of AI in ophthalmology, especially to patient triage and telemedicine care, during COVID-19 era.
- Adapting the healthcare system functioning, e.g. by developing prioritization strategies for resource allocation, such as vaccines [142, 143] or by predicting the needs [144, 145]. For example, Cheng et al. [144] used random forest to predict the ICU transfer of a patient within 24 h and Shashikumar et al. [145] used neural network to identify patients who could need mechanical ventilation. AI also permit the optimization of other care, for example Rozario [146] used ML to optimize the operating room allocation in the actual context of extended waist list of surgical procedures due to the cancellations.

Social, Economic and Governmental Measures Assessment

AI had been used to evaluate the effect of public health interventions [57, 147–152]. For example, Porcher [153] propose to document all governments' response to pandemic around the world. Mégarbane et al. [57] produce a tool to monitoring lockdown efficacy and assessing its adequate duration. Dandekar et al. [152] used mixed first-principles epidemiological equations and data-driven neural network model to estimate the role played by the quarantine and isolation in controlling the reproduction number Rt of the virus.

AI had also been used to assess the "side effects" of pandemic, media coverage and governmental decisions, i.e.:

- Assessing the effects on the economy, e.g. by identifying actions minimizing the number of infection and the effect on economy [154, 155] or by measuring the impacts on the electricity and petroleum demand [156] or on supply-chain [157].
- Effects on the environment, e.g. with the measure of lockdown effects on pollution [158].
- Effects on the society, e.g. by describing the effects of mobility restrictions and their heterogeneity between socio-economic or racial groups [159], by evaluating the perceptions and reactions of the population [89, 111–117, 160], or by describing the effect on vulnerable populations [161].
- Effects on the whole health, especially on mental health [162–164].

Infodemics

The pandemic has been characterized by the over-abundance of information and the important diffusion of rumors and fake news, which are not without social and consequences [165, 166]. AI has been proposed to automatically detecting misinformation at their early stage of diffusion [167–174]. For example, Uyheng et al. [173]

used random forest to detect hate speech in Twitter. Paka et al. [174] also developed a "cross-stitch based semi-supervised end-to-end neural attention model" to identify fake-news in Twitter.

What Performance of AI in the Uses Identified in Epidemiology and Public Health?

Laurent Tournois

The evaluation of the performance of AI algorithms for a given use case is a necessity, and in the majority of cases, it should be done by comparison with the historical technique, the gold standard. It is often implicit that AI will improve the performance expected in a use case, compared to the historical reference technique. This is not systematic; it is even quite often false. Also, as for other types of study or techniques, guidelines have been established for the reporting of studies presenting an AI algorithm applied to a use case: the TRIPOD [175] and CHARMS guidelines [176]. It is possible to rely on these guidelines in order to assess, on the one hand, the quality of the reporting of studies on the use of AI in epidemiology and public health, and on the other hand, to give a first estimation of the performance of these algorithms.

We present here some examples and results that we considered illustrative regarding the study of the pandemic. A summary of AI-based applications and their performances is provided in Table 13.2.

Outbreak Monitoring

AI may be used as a modeling tool of the spread of COVID-19 to monitor disease outbreaks around the world. For instance, the authors in [86] used AI-based techniques to model the worldwide spread of COVID-19. They designed three multilayer perceptrons (MLPs) to estimate the maximal number of infected, recovered and deceased people due to COVID-19 per location and unit time respectively. The models were trained from data of 402 locations, mainly in the northern hemisphere, and collected from 22nd January 2020 to 12th March 2020, consisting in 20,706 data points in total. The data are available at https://github.com/CSSEGISandData/COVID-19 (accessed on 7th March 2022). The features of each MLP included the latitude and the longitude of the location and the number of days since the first case appeared, relative to the date of data collection. The authors used fivefold cross validation to select the best models and R^2 (coefficient of correlation) between the real and predicted values to assess the models performance. The values of R^2 were 0.98599, 0.97941 and 0.99429 with standard deviation 0.037, 0.072, 0.021 for the estimation of the maximal number of infected, recovered and deceased people due to COVID-19 respectively. Those R^2 values indicated that the models seemed to be

Table 13.2 Summary of the AI-based applications with their usage, purpose, type of AI algorithm developed in the article and model performance

Usage	Reference	Purpose	Type of AI	Model performance
Outbreak monitoring	Car et al. [87]	Estimation of the maximal number of infected, recovered and deceased people due to COVID-19 per location and unit time	MLP	R^2 between predicted and ground truth values of maximal number of infected, recovered and deceased people are respectively 0.98599, 0.97941, 0.99429 with standard deviation 0.037, 0.072, 0.021.
Outbreak monitoring	Huang et al. [87]	Estimation of the number of cumulative confirmed cases of COVID-19 in seven cities in China	CNN	MAE ranges from 3.808 to 426.179 (average is 102.643) and RMSE ranges from 4.779 to 456.910 (average is 109.439) depending on the city
Epidemiologic outcomes and characteristics discovering	Li et al. [93]	Determination of the risk factors associated with COVID-19 in terms of cases, deaths and case fatality rates	ridge logistic model	For the COVID-19 cases, deaths and case fatality rates, the AUC is respectively 0.8, 0.83, 0.81 in the 154 countries and 0.95, 0.94, 0.89 in the 50 US states.
Epidemiologic outcomes and characteristics discovering	Jehi et al. [94]	Development of a nomogram that predicts the risk of a positive COVID-19 case	multivariable logistic regression model	With a cutoff of 0.13 (positive being detected equal to negative being avoided), sensitivity, specificity, negative predictive value and positive predicted value in terms of classification between positive and negative COVID-19 case are respectively 0.762, 0.765, 0.957, 0.319
Social control and monitoring	Zhang et al. [118]	Identification of sentiments in social media conversations about COVID-19 in Arabic, French, Spanish, English and Chinese	Deep learning model	Accuracy of the models are greater than 0.82. The accuracies range from 0.824 ± 0.004 (French tweets) to 0.905 ± 0.002 (Arabic tweets)

(continued)

Table 13.2 (continued)

Usage	Reference	Purpose	Type of AI	Model performance
Social control and monitoring	Ahmed et al. [102]	Identification of humans for social distance tracking using an overhead perspective	Object detection deep learning model	Precision, recall and accuracy for human detection are respectively 0.86, 0.83, 0.95. The tracking accuracy is 0.95.
Assisted/ augmented scientific research and knowledge sharing	Joshi et al. [123]	Literature mining for research purposes and summarization of articles	NLP techniques with deep learning	No information
Assisted/ augmented scientific research and knowledge sharing	Wise et al. [125]	Visualization of complex relationships between COVID-19 and scientific articles and topic modeling	Knowledge graph, Z-LDA	Average F1 score of 0.9192.
Healthcare resources adaptation and optimization	Cheng et al. [144]	Prediction of ICU transfer within 24 h for a patient	Random forest	Sensitivity, specificity, accuracy and AUC (threshold at 0.5) are respectively 0.728, 0.763, 0.762, 0.799.
Healthcare resources adaptation and optimization	Shashikumar et al. [145]	Identification of COVID-19 patients in need of mechanical ventilation	Feed-forward neural network	AUC is 0.882 at 0.8 sensitivity
Social, economic and government assessment	Dandekar et al. [152]	Estimation of quarantine effect on the COVID-19 infection spread	MLP	No performance metric but comparison graphs between predicted and real number of COVID-19 cases showing promising results.
Infodemics	Uyheng et al. [177]	Identification of hate speech, offensive speech and normal speech from tweets	Random forest	Micro-averaged F1 score is 0.8417 and weighted F1 is 0.8293

CNN convolutional neural network, *MLP* multilayer perceptron, *NLP* natural language processing, *Z-LDA* Z-latent Dirichlet Allocation

performing enough to estimate the number of infected, recovered and deceased individuals in the early stages of the COVID-19 pandemic. The authors declared that their models were the most performing compared with the models performance in the literature. They intended to implement their models within a web interface. They also suggested exploring other AI-based architectures, such as recurrent neural networks (RNN), increasing the datasets size to improve the model performances and

comparing models for different infective diseases. The code, models and inputs are available at COVID-19 MLP, Riteh AI and Robotics Group https://github.com/RitehAIandRobot/COVID-19-MLP (accessed on 7th March 2022).

The spread of the pandemic may also be monitored by the number of cumulated confirmed COVID-19 cases. The authors in [87] estimated this number for seven cities in China (Wuha, Huanggang, Xiaogan, Ezhou, Yichang, Wenzhou, Shenzhen) by using convolutional neural networks (CNNs). The models were trained from cases confirmed from January 23rd, 2020 to February 17th, 2020 and they were validated from a test set which consisted in cases confirmed from 18th February 2020 to 2nd March 2020. The features of the CNN were defined as a 6×5 input vector encompassing the number of cumulated confirmed cases, total confirmed new cases, cumulated cured cases, total cured new cases, cumulated deaths and total new death per day since the past 5 days before the day to predict. The authors compared the CNN model performances with all those features and with only one feature, the number of cumulative confirmed per day since the past 5 days before the day to predict, as input of the model. They also compared the performance of the CNN model with other AI models such as long short-term memory (LSTM), gate recurrent units (GRU) and MLP networks. The best performance on the test set was obtained with the CNN with all features as input compared with the other AI models. For the CNN model, the mean absolute error (MAE) ranges from 3.808 to 426.179 with an average of 102.643 and the root mean square error (RMSE) ranges from 4.779 to 456.910 with an average of 109.439 depending on the city for which the model was applied. Those performance values seemed relevant with a usage in the context of the pandemic. However, the authors pointed out that they used small datasets and suggested that the model performance may be improved by the establishment of more complete databases and the development of hybrid AI models, that is to say AI models with mixed structures.

Epidemiologic Outcomes and Characteristics Discovering

During the COVID-19 pandemic, researchers used AI to identify risk factors associated with COVID-19 to understand the characteristics of the disease and provide relevant information to governments. The authors in [93] developed ridge logistic models to identify novel critical factors associated with COVID-19 cases, deaths and case fatality rates in 154 countries and in the 50 US states respectively. The models estimated the number of COVID-19 confirmed cases (per one million) and deaths (per one million). The case fatality rates were calculated on a given day by dividing the number of COVID-19 deaths that day by number of cases 14 days before. The model was trained on 75% of the available data and evaluated on the 25% remaining data. However, the dataset size was not provided by the authors. The features used as inputs in the model encompassed 77 variables for the 154 countries dataset and 55 variables for the US states dataset. After model training, for the COVID-19 cases, deaths and case fatality rates, the area under the receiver operating characteristic curve (AUC) was respectively 0.8, 0.83, 0.81 in the 154 countries dataset and 0.95, 0.94, 0.89 in the 50 US states dataset. Relevant features regarding

the COVID-19 cases, deaths and case fatality rates were extracted from the model by using a least absolute shrinkage and selection operator (LASSO) algorithm with fivefold cross validation. Only the variables with a nonzero β-coefficient were kept. A total of 57 variables that could affect COVID-19 transmission were selected over the initial 77 variables for the 154 countries dataset. Similarly, 27 variables were selected over the initial 55 variables for the US states dataset. The authors found that the three best predictors for the 154 countries dataset were unimproved water sources, internet usage and urbanization levels for COVID-19 cases, a body mass index lower than 18, unimproved sources of water and HIV for deaths and latitude, health expenditure and vehicle usage for case fatality rates. The top three predictors for the US states were Black people, sports events and inland for COVID-19 cases, international tourism revenue, Black people and community hospitals for deaths and private transport in urban environments, sex ratio and birthrate for case fatality rate. The authors declared that some factors needed to be further validated in large-scale clinical data and that the sample sizes were not sufficient in terms of numbers of predictors used (since β-coefficients of some variable were low).

In [94], the authors developed a multivariable logistic regression model to estimate the number of COVID-19 infections to build a nomogram that predicts the risk of a positive COVID-19 case. The patients came from Cleveland clinic locations in Ohio and Florida and presented symptoms of a respiratory tract infection or with other risk factors for COVID-19. The model was trained with a set of 11,672 patients including 818 positive cases for COVID-19 and was evaluated with an external validation set of 2295 patients including 290 positive cases for COVID-19 and coming from the Cleveland Clinic in Florida from 2nd to 6th April 2020. With a cutoff of 0.13 (corresponding to positive cases being detected equal to negative cases being avoided), sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV) on the external validation set are respectively 0.762, 0.765, 0.957 and 0.319. The authors found that the model overestimated its predictions for a probability of positive case greater than 40% (in the external validation set). They also highlighted that being Male, African American, older patients, and people with known COVID-19 exposure were at higher risk of being positive for COVID-19 whereas the risk was reduced for people who had pneumococcal polysaccharide or influenza vaccine or who were on melatonin, paroxetine, or carvedilol treatment. A higher risk of COVID-19 was also noticed for people with a poor socioeconomic status. However, the authors also pointed out that reverse transcription polymerase chain reaction (RT-PCR) was not enough sensitive for nasal swabs (sensitivity of 0.63) and pharyngeal swabs (sensitivity of 0.32). They mentioned that their findings must still be reproduced and validated. Therefore, the model performance values should be interpreted with carefulness.

Social Control and Monitoring

AI can leverage social media for surveillance and monitoring of the population during the COVID-19 pandemic. For instance, the authors in [118] used deep learning models to identify sentiments in social media conversations about COVID-19 in six

different languages (English, French, Arabic, Spanish, Italian and Chinese). One model was trained for each language and each model was able to identify if a social media post conveyed positive expressions (such as being thankful, optimistic, empathetic), negative expressions (anxious, sad, annoyed, pessimistic and denial) or a complicated expression (joking). The models were trained from 10,000 annotated posts published from 1st March 2020 to 15th May 2020 in each defined language, except Chinese posts which consist in 21,173 Sina Weibo messages. Each post underwent sentiment annotation by human experts and was converted into a 768-dimensional vector used as input in the models. The data are available at https://github.com/gitdevqiang/SenWave (accessed on 7th March 2022). After training with fivefold cross validation, the performance of the models was assessed in term of classification accuracy, which ranged from 0.824 ± 0.004 (obtained for French tweets) to 0.905 ± 0.002 (obtained for Arabic tweets). The models were then applied to million tweets and Sina Weibo messages (for Chinese posts). The authors found a rapid increase of conversations about COVID-19 in all languages followed by a slow decline. The tweets that conveyed joking and negative emotions appeared to turn into positive emotions, especially in Arabic posts as the pandemic came under control). The authors also pointed out that the increase in tweet volume was driven by the government enforcement of confinement measures and the collapse of stock market and the global economy.

In [102], the authors used object detection algorithms to detect humans and track social distancing from video sequences with an overhead perspective. The video frames, consisting in people moving freely, were collected from the indoor of the Institute of Management Sciences, Hayatabad, Peshawar, Pakistan. The dataset of video sequences was split into a training set and a test set with training-to-test size ratio of 70/30. The object detection model was then trained from a YOLOv3 architecture to localize humans in the videos as bounding boxes. In other words, the model predicted the 2D spatial coordinates, the width and the height of a bounding box delimiting a person. Afterwards, the Euclidean distance between each pair of box centroids was calculated to estimate the distance between people. Those who do not respect social distance were highlighted in red on the video and tracked. The precision, recall and accuracy of the model for human detection were respectively 0.86, 0.83, 0.95. The tracking accuracy was 0.95. The model was compared with other object detection architectures, such as Mask-RCNN, Faster-RCNN and Fast-RCNN. According to the authors, the YOLOv3-based model was the most performing one compared with those model architectures.

Assisted/Augmented Scientific Research and Knowledge Sharing

AI models were developed for data mining for research purposes regarding the COVID-19 pandemic. For instance, the authors in [123] created a literature mining system from biomedical literature databases and a research summarization tool for early-stage comprehension of the pandemic with natural language processing (NLP) techniques. The system, called deepMINE, was developed from the CORD-19

database [ref] containing 29,315 articles representing a total of 146,115,136 words. However, the authors did not provide any information about the type of AI used in the NLP techniques used to create the system or any performance metrics. Despite the lack of information regarding the development and evaluation of the system, the tool was supposed to be available at https://deepmine.in/ according to the authors. However, this link was not accessible (tried to access on 7th March 2022). Moreover, since no information was provided about the performance of the model, it is currently not possible to analyze the performance of the AI system like this is done for the previous examples of AI applications.

The authors in [125] designed a COVID-19 knowledge graph by extracting and visualizing complex relationships between COVID-19 and scientific articles from the CORD-19 dataset. They defined the knowledge graph from five types of entities: paper, author, concept, institution, and topic. The authors used a multi-label classifier based on Z-latent Dirichlet Allocation (Z-LDA) to predict the topic of articles among ten defined topics (vaccine/immunology, genomics, public health policies, epidemiology, clinical treatment, virology, influenza, healthcare industry, lab trials and pulmonary infections). The model was trained from titles, abstracts and text bodies from less than 59,000 articles. The model reached an average F1-score of 0.9192, which suggested that the classifier was able to attribute topics to given articles. The model was then validated by verifying that the classifier predicted well the topic of articles from the Journal of Virology and the Journal of Vaccine. The validation results were rather promising for a future use by researchers. The COVID-19 knowledge graph is now used to power an archive search engine available on https://www.cord19.aws/ (accessed on 7th March 2020).

Healthcare Resources Adaptation and Optimization

In the context of the COVID-19 pandemic, the intensive care units (ICU) were rapidly overwhelmed by COVID-19 patients who needed ICU care level. Therefore, AI applications were developed to help adapting the resources and optimizing the flow of patients who needed intensive care.

In [144], a random forest algorithm was developed to predict if a patient needed an ICU transfer within 24 h. The model was trained from 401 patients and validated from 521 patients. All patients were older than 18 years old and admitted to Mount Sinai Hospital from 26th February 2020 to 18th April 2020. The features used to develop the model include data about vital signs, nursing assessment, laboratory data and electrocardiograms. For each patient, the last three observations were used as input. The missing data were replaced with the median value calculated across the cohort. The model performance was assessed with the sensitivity, specificity, accuracy and AUC with a classification threshold of 0.5 to determine if a patient needed intensive care within 24 h or not. With a 95% confidence interval, the model achieved [0.632, 0.811] sensitivity, [0.747, 0.779] specificity, [0.746, 0.777] accuracy and [0.752, 0.846] AUC. However, the authors pointed out that the random forest model was built from a low-size dataset with class imbalance. Therefore, they

recommended to use the resulted model for resources prioritization rather than clinical decision support. Moreover, they did not guarantee the generalization capacity of the model to other data than those which were used to train and evaluate the model since the data came from a sole hospital. To improve their model, they proposed to add other features such as the fraction of inhaled O_2, the level of respiratory support and the percutaneous oxygen saturation.

In another AI-based application, the authors in [145] developed a feed-forward neural network, called VentNet, to identify COVID-19 patients in need of mechanical ventilation. They trained the network from 18,528 patients admitted to ICU at the University of California San Diego Health between 1st January 2016 to 15th January 2020 and validated the model from 3888 patients admitted to ICU at the Massachusetts General Hospital between 1st February 2020 to 4th May 2020. All the patients were adults and those with ICU-like level of care are considered since ICUs expanded to non-traditional floors due to the high rate of COVID-19 patients who required ICU transfer. The patients for whom the length of stay in ICU lower than 4 h or higher than 20 days, or with a need of non-invasive mechanical ventilation only are excluded from the datasets. The features used in the model included 40 clinical variables organized in 1 h non-overlapping time intervals. The missing data were imputed by the mean of the values for respective features. The model achieved an AUC of 0.882 for the test set. This performance value was the best compared to a logistic regression model (AUC was 0.773) and a model based on ROX score (AUC was 0.782). For the COVID-19 ICU population, the AUC reached 0.919. The authors pointed out that the models were developed for patients with a need of invasive mechanical ventilation only, and that the resulted model was trained with historical data instead of data collected during the COVID-19 pandemic. Therefore, even though the model achieved good performance on the COVID-19 populations in the test set, one must be careful to use the model for data acquired during the pandemic since it was not validated with such data.

Social, Economic and Government Measures Assessment

Government responses regarding the COVID-19 pandemic may be assessed by AI models. In [152], the authors developed a susceptible-infected-recovered (SIR) model augmented with a MLP to quantify the effect of the quarantine in the COVID-19 infection spread. The MLP model predicted a quarantine strength function, corresponding to the strength of the quarantine over time. The model was trained from available data of Wuhan (collected from 24th January 2020 to 3rd March 2020), Italy (collected from 27th January 2020 to 23rd March 2020), South Korea (collected from 2nd February 2020 to 17th March 2020) and predictions were made from available data of USA (collected from 8th March 2020 to 1st April 2020). The model inputs included the number of susceptible, infected, recovered and quarantined people. No performance metric was precisely defined. However, the authors compared graphs representing the number of infected and recovered cases for each region over days post 500 infected and with or without the use of the

MLP to assess the performance of the AI model over the traditional SIR model. At first glance, those graphs highlighted that the SIR model augmented with MLP better fitted the data than the traditional SIR model in the Wuhan dataset. The MLP model fitted relatively well the data for the Italy, the USA dataset and the number of recovered cases for the Wuhan and South Korea datasets. However, the model did not manage to fit well the number of infected people in the South Korea for data after 15 days post 500 infected people and Wuhan datasets for data after 25 days post 500 infected people. Therefore, one should be careful when using the model. Despite this defect, the authors found a strong correlation between the strengthening of the quarantine, the actions taken by governments and a decrease in the effective reproductive number (representing the average number of secondary cases per infectious case in a population). Further study seemed to be required to validate the MLP-augmented SIR model.

Infodemics

Nowadays, with the widely use of social media, the effects of the COVID-19 pandemic on the population emotional state as it is reported by people using social network is relatively easy to access. For instance, the authors in [118] developed an AI-based model to classify social media conversations into specific emotion categories conveyed by those conversations. The authors in [177] used a random forest algorithm to identify hate speech, offensive speech and regular speech from tweets and assessed links between hate speech and bot-driven activity. The model was built from 12 million tweets about the COVID-19 pandemic in the USA and in the Philippines from 5th March 2020 to 19th May 2020. The features of the random forest model included predictors defined by the NetMapper software, such as the lexical counts of pronouns, abusive terms, exclusive terms, absolutist terms and the identity terms. The authors added the counts of English words, and words from common languages in the Philippines in the features. The model achieved a micro-averaged F1-score of 0.8417 and weighted F1-score is 0.8293. The random forest model was the best one compared with other models such as a random-based model, a heuristic-based model and a logistic regression model. Those results suggested that the application of the model may be relevant for other locations than the USA or the Philippines. However, the authors did not provide information about any train and test set. Therefore, it is not possible to assess the generalization capacity of the model on other data. The authors also performed an analysis of links between bot-driven activities and hate speech with a random forest-based algorithm called BotHunter. In the USA dataset, they found that low-hate humans and high-hate bots had significant influence over network information. In the Philippines dataset, the analysis suggested that the information flow on Twitter tended to be more controlled by bots. The authors also pointed out that bots control major aspects of information flow in the conversations about the COVID-19 pandemic.

In the next section, the maturity of the AI applications discussed in this part is assessed.

Beyond Performance, What About the Degree of Maturity of Published AI Algorithms?

AI algorithms are expected to be developed and validated, so that they can be used in a specific use case. Also, after having ensured a good share of acceptable and relevant performance of an algorithm in a given use case, as well as the quality of the reporting of the studies presenting it, one may be interested in its degree of maturity, that is to say, the level of integration of a given algorithm from the observation of the basic principles to the moment when it is actually put into production, used and usable in routine. To do so, we can use the concept of technology readiness level—TRL. Indeed, this concept is currently used by the National Aeronautics and Space Administration (NASA) and the Department of Defense of the USA to assess the maturity level of technologies to be implemented in those structures. The TRL scale is based on nine values which describe the integration level of a technology from the observation of the basic principles of that technology to its successful use in real operational environments [178]. This scale is mainly based on research, development and validation levels of a technology. The maturity of the AI-based applications described in the previous section ("What Performance of AI in the Uses Identified in Epidemiology and Public Health?") may be assessed using an adapted TRL scale inspired from the NASA TRL scale and defined by six levels of integration (Table 13.3). The first level corresponds to the formulation of the idea to develop the AI model. In the second level, the model is developed but is not validated on a well-defined dataset. In the third level, the AI algorithm is validated from one or several internal datasets but requires further validations on one or several external validation datasets. In the fourth level, the AI model is validated on one or several external datasets with small representative data. In the fifth level, it is validated on a large-scale external dataset but may still require further validations. In the sixth level, the AI model is completely validated and is reported to be successfully used in real and operational environments with new data.

We take here some examples and results used in the previous section, and study them from the angle of their estimated TRL. Table 13.4 summarizes the estimated TRL of each AI application and the key points used to estimated that TRL. Details about how the TRL was estimated for each example are described below.

Table 13.3 Definition of the levels of maturity for the AI algorithms designed for epidemiology and public health purposes during the COVID-19 pandemic, adapted from Mankins, 1995 [178]. *eTRL* estimated technology readiness level

eTRL	Characteristic
1	Formulation of the AI model concept
2	Proof of study without clear validation
3	Internal validation of the model performance
4	External validation of the model performance with small datasets
5	External validation of the model performance with large datasets
6	Reports of successful usages in real operational environments with new data

Table 13.4 Summary of estimated TRL per AI applications described in section "What Performance of AI in the Uses Identified in Epidemiology and Public Health?". *eTRL* estimated TRL

Reference	Publication status	Key points to assess the TRL of the model	eTRL
Car et al. [86]	Peer-reviewed	Internal validation	3
Huang et al. [87]	Preprint	Datasets (train and test sets) are small but not detailed (internal or external dataset not precised)	2
Li et al. [93]	Peer-reviewed	Small datasets, distribution of data in the datasets is not detailed, features need to be validated at a large scale	4
Jehi et al. [94]	Peer-reviewed	Overestimation of predictions but overfitting is not studied, data imbalance, further validations are required	4
Zhang et al. [118]	Peer-reviewed	Internal validation	3
Ahmed et al. [102]	Peer-reviewed	Train and test datasets are not detailed in terms of distribution or size.	2–4
Joshi et al. [123]	Preprint	No information about the natural language processing algorithm used or the dataset distribution or the performance metrics.	–
Wise et al. [125]	Preprint	Size and distribution of the train and validation datasets are not detailed but the final knowledge graph is applied to power a search engine.	2–4
Cheng et al. [144]	Peer-reviewed	Small dataset, external validation, need further validation studies and model performance improvement.	3
Shashikumar et al. [145]	Peer-reviewed	External validation set, imbalanced dataset towards the outcome to predict, relatively small dataset towards given outcomes.	4
Dandekar et al. [152]	Preprint	The datasets were not detailed. No performance metric but graph analysis between estimated and real values.	2
Uyheng et al. [177]	Peer-reviewed	The type, size and distribution of datasets are not detailed. The initial dataset is relatively large (12 million instances).	2–4

Outbreak Monitoring

In Car et al. [86], even though the performance of the AI model seemed good and relevant for a use in real an operational environment, the authors used K-fold cross validation without external validation set to select the best model among models with different hyperparameters and assess its performance. Therefore, the performance of the model may be biased towards the validation sets. The model should be trained with an external validation set to prevent the overestimation of the model performance. Moreover, the data seemed very sparse (mainly consisting in values of 0 for several features), which could lead to overfit the model towards the predictions made from instances with features equal to 0. However, the problem of overfitting is not clearly treated. Furthermore, the high regularization parameter of the model

suggested that some features were not relevant for the predictions of the model. Therefore, further studies are required to validate the performance of the model.

In Huang et al. [87], the performance of the model also looked promising. However, the datasets size were not given but were considered small by the authors and the possible presence of overfitting was not mentioned. Moreover, the model developed in this article was built from data before the COVID-19 disease was considered as a pandemic (before 11th March 2020). Therefore, the model may not take into account pandemic settings and may not be applicable in the context of the COVID-19 pandemic. In addition, this model was trained from data from seven cities in China, no validation was performed for other locations around the world, probably because China was the first country to be hit by COVID-19 and most of available data were related to China cities. Therefore, further validations seemed to be required to apply the AI model in the current context of the pandemic.

Epidemiologic Outcomes and Characteristics Discovering

In Li et al. [93], the AI model was developed with fivefold cross validation on the train set and assessed with an external validation set. The robustness of the model was tested by variation of training and test set proportions. The authors found that some features needed to be further validated for large-scale clinical data. Moreover, the model may not be applicable to the current situation since the testing capacity against COVID-19 is not similar between countries. The authors also highlighted that the sample sizes were not sufficient in terms of number of predictors used in the model, which may lead to ambiguous associations of factors with the risk of COVID-19.

In Jehi et al. [94], the datasets were specific to Cleveland clinic locations. Therefore, the resulting model was rather specific to Cleveland COVID-19 patients. The authors pointed out that the predictions of the model were overestimated for probabilities of positive case greater than 40% in the external validation set. This may be the result of model overfitting. However, this issue was not studied in this article. In addition, the datasets were imbalanced in terms of positive and negative patients towards COVID-19, which may bias the model. Moreover, the authors declared that further validations were needed since the pandemic has evolved so that the data distribution may currently be different from the data which were used.

Social Control and Monitoring

In Zhang et al. [118], the model was trained with fivefold cross validation and without external validation dataset. The datasets were imbalanced towards English posts (64% of posts in the datasets). Moreover, the model performance was not assessed with an external validation dataset. Therefore, further validations are required.

In Ahmed et al. [102], the URL given by the authors to access the datasets was not reachable. Therefore, no information was provided about the size or the

distribution of the dataset. The reader only knows that two datasets were used, a training and a test set with instance proportions of 70/30 respectively. The performance metrics were defined but the intersection-over-union threshold for the localization of humans was not detailed. Moreover, some people were not detected by the AI model (see figures 12c, h, g in their article), which suggested that the model performance may be improved.

Assisted/Augmented Scientific Research and Knowledge Sharing

In Joshi et al. [123], the authors provided the size of the global dataset. However, they did not provide any information about the type of natural language processing algorithm they used to develop the model or any metric used to assess the model performance. The authors provided a link to access the AI model called deepMINE. However, this link was not reachable (try to access on 7th March 2022). Therefore this application is not really assessable in terms of technology readiness level.

The knowledge graph developed in Wise et al. [125] is currently available on the Internet. However, the size and the distribution of the datasets used to train or validate the topic modeling task was not detailed. Only the size of the initial dataset split into a train and validation set was given. Therefore, it is rather hard to estimate precisely the TRL. One may only give an interval of TRLs (from 2 to 4) that may apply to the AI model.

Healthcare Resources Adaptation and Optimization

The authors in [144] developed a random forest algorithm from data which were imbalanced in gender and in age for people older than 80 years old, which may lead to bias the model towards the over-represented categories of people. No study of bias was performed. Nevertheless, the authors detected data imbalance towards people who do not require ICU transfer (the rate of ICU transfer was 3.7%). To overcome this issue, they performed under-sampling on the training dataset to balance the data and get equally distributed cases in terms of ICU transfer. However, the test (or external validation) set was not modified, so the patients that do not required ICU transfer were five times more represented than people who require ICU transfer. Moreover, compared with the number of predictors, the number of instances may be too small. In other words, the datasets were relatively small for the number of variables used for each instance.

In [145], the authors developed a model from historical data of 18,528 ICU patients before the COVID-19 pandemic. Therefore, the model may not consider the current COVID-19 pandemic characteristics. They used tenfold cross validation

on the train set and assessed the performance of the model on an external validation set containing data acquired during the COVID-19 pandemic. The datasets were highly imbalanced in terms of ventilated and non-ventilated patients. To prevent overfitting, the authors used L1-L2 regularization. Further validation studies may be required since the data were relatively small concerning people in need of invasive mechanical ventilation and those data only came from two locations in the USA.

Social, Economic and Government Measures Assessment

The authors in [152] developed an AI model integrated with a SIR model. The size and distribution of the datasets were not clearly defined and there was no performance metric. Therefore, the performance of the AI model was not quantifiable. As a consequence, further details about the datasets and a clear performance value may help understand the contribution of the AI model in the MLP-augmented SIR model.

Infodemics

In [177], the authors developed a classifier from an original dataset of 12 million tweets. However, the type, distribution and size of the datasets were not given by the authors. Therefore, it is not possible to assess the generalization capacity of the model on other data, so to assess precisely the maturity of the classifier. Moreover, the authors declared that a qualitative assessment of the predictions in relation to empirical observations is required, which suggest that the assessment of the maturity of the model in terms of TRL is relatively difficult. Therefore, the level of maturity of the model may be assessed between 2 and 4 in terms of estimated TRL.

From all the AI models presented in this section, it was not possible to assess one model and to assess precisely three models due to a lack of information. Future model development and validation studies should follow the TRIPOD [175] guidelines for better transparency and comprehension of AI models.

Two Years of Pandemic: Lessons for Epidemiology and the Place of AI

Thomas Lefèvre

We close this chapter with this last part in the form of both an assessment and perspectives, with some general questions in terms of ethics, the need for regulation and ecology.

Many AI Applications for Epidemiology and Public Health in the Context of the Pandemic: Yet Still Evidence for Their Reliability and Usefulness to Bring

The analyzes conducted and reported in the previous section clearly show that we are still at this first moment that we know about AI applied to health: AI is above all used as a new source of data, to problems already identified and according to approaches that are mostly identical to those known for several decades. In addition, most of these new techniques have not yet proven their worth, and more embarrassingly, they are too rarely evaluated in conjunction with gold standards. Finally, the maturity of the algorithms published remains low, and the entire chain of external validation, evaluation of the service rendered and finally production, is absent. The example of the pandemic is interesting, because it has given rise to almost synchronous production never seen before everywhere in the world, with substantial dedicated resources over an ultimately short period of time. However, it is difficult to say that the results obtained were up to the means deployed in the case of AI. Thus, we are certainly not at the possible moment of breakthrough technology, or of the epistemological and industrial revolution announced in this field. Still, the efforts invested may have sown fertile seeds for the future, but it is too early to know. On the other hand, it is more worrying to note that if the technical efficiency was not at the rendezvous, the techniques deployed were accompanied or could be used as tools integrating population control strategies, sometimes poorly guarantee by law and the state.

"Lancetgate": A Lesson About Identified Risks of Massive Data Collection and Reuse and Its Consequences on Public Health Decisions

Even if it is not directly about AI in the example that we are going to briefly discuss, it is about the collection and analysis of massive data, the reuse of routine care data, all of which had a direct impact in terms of public health policy, at least in several countries. This is what has been described as a large study based on collecting data from hundreds of hospitals around the world, exploring the value of hydroxychloroquine in the treatment of patients with Covid. In a context of crisis, we realize that the criterion of volume, in data and in the centers concerned, is a major criterion that can win support in terms of reliability and generalizability of the results. This study was published in the prestigious medical scientific journal The Lancet. Quickly, suspicions appeared as to the seriousness, or the transparency of the method used by the study. Very quickly, this was relayed by the general media all over the world, becoming what is known as the "lancetgate" [179]. However, with

a view to basing health policy decisions on scientific data, or at least justifying some of them on studies chosen at a given time, and this in the most reactive way possible, this publication has been used by several governments to justify the ban on the prescription of hydroxychloroquine in patients with Covid. We are not discussing here whether or not this molecule has an interest in this context, we simply draw attention to two elements: (1) the studies which claim to mobilize very large volumes of clinical data over a geographical area which does not cannot be superimposed with a network of establishments already in collaboration, very quickly, are currently to be examined very closely and must be questioned with great rigor; (2) scientific publications can lead to, or justify, large-scale policy decisions. Thus, once again, we are still a long way from the possibly future moment when reusable data sources are equally reliable with each other, perfectly documented, interoperable, quickly linked, accessible to the research community and make it possible to guide political decision-making in real time. Moreover, even in this type of ideal situation, it should be noted that this should in no way replace a contradictory debate between peers, which is a crucial element in the consolidation of results in scientific fact.

What AI Has to Learn from Epidemiology and Public Health?

A current line of thought to evolve either the principles of AI, or its place in society in its interactions with humans, is that of intuition versus reasoning [180]. Both rely on a set of experiences in sufficient number and variety to be good enough, but in qualitatively different ways. The AI has no intuition, and can get better if we keep presenting it with new cases to consider, up to a point, but it won't think "outside the box". Thus, interesting results are beginning to emerge in certain fields, such as mathematics [181], where a human researcher can make breakthroughs by following his intuition, and supported by an AI on another side, that of the verification of hypotheses on large sets of data.

However, AI seems to be able to work in areas where a solid intuition has been put in place, nourished by sufficient experience and varied data. However, currently, neither epidemiology nor public health seem to be in a position to produce this type of expert and intuition. Intuition also seems poorly suited to situations of emerging phenomena, in particular those whose evolution is not predictable a priori in the short and medium term.

This could therefore be a source of major limitation for the development of efficient and useful AI for epidemiology and public health: indeed, for a certain number of uses where our knowledge or human performance appears to us to be limited, we expect AI that it can overcome these limitations. However, it would seem that this specific expectation is not the one to which AI is an appropriate response.

Predicting vs. Explaining: Is It Reconcilable? Is Explainable AI Necessary in Epidemiology?

Classical epidemiology is interested in identifying the causes of diseases, in particular with the aim of developing public health interventions that maintain or improve the health of populations. The causal dimension is currently central, and among the most problematic and complex to establish.

In parallel, AI has progressed in recent years precisely because it has been freed from an over-specification of reasoning rules intelligible by a human being, very close to causality. The flipside is that most successful AI models rely on making connections between many features, some if not all of these features may be obscure to the human, and the nature of the connections as well. Furthermore, the amount of links and features makes it unrealistic for humans to fully explore a given algorithm—just as computer code today cannot be fully verified in all its possible configurations of use. Part of AI research is making AI explainable. It is not certain that a better performance/explainability trade-off exists [182].

The development of AI is more oriented towards "prediction", which may have an interest in epidemiology and public health to deploy measures, according to the degree of risk and expected imminence. However, part of AI research is oriented towards visualization and the search for causality or quasi-causality in data sets. Certainly, both approaches can serve epidemiology and public health.

The Status of the Whistleblower in the Case of Emerging Diseases: The Hybrids of Simondon and Latour

"Weak signals" only become signals when meaning has been given to a set of data collected independently or not of each other, in a given context. There is no scientific rationale to justify the establishment of a system of generalized collection of massive data with no more specific purpose than monitoring "weak signals", unspecified by definition.

The place of whistleblowers remains essential to this day, that is to say that the human is not replaceable in this role. But this position is highly risky and unstable and requires protection.

Several countries and institutions are calling for or are in the process of adopting measures to protect whistleblowers. Such laws now seem necessary [183].

In the context of the pandemic, we can thus remember the ophthalmologist who had exchanged a document with his colleagues, in Wuhan, alerting to the emergence of cases grouped geographically and presenting a SARS-type syndrome [184]. He was arrested by the police shortly before dying of the illness he complained of. A little later, during the establishment of the quarantine in Wuhan, a journalist regularly alerted to a different vision of the field compared to the official version and broadcast around the world. She is currently in prison [185].

The place of AI in relation to humans in this role of detection and alert therefore remains to be specified. However, it would probably be a mistake to always want to separate human and AI, to consider that they are two different "species".

A mistake would probably consist in wanting to consider that AI is purely "artificial" and separated from human activity. The tools shaped by humans are for the needs he identifies, and for specific purposes that can evolve. Tools are not purely artificial but have a part of existence linked to our humanity, to typically human activities [186, 187]. It would no doubt be more relevant to seek how our two species can best articulate.

We live and contribute to shaping a world through our activities, and creating tools is part of these actions. Their use is inseparable from our way of being, and therefore of acting in the world [186].

Towards a Potentially More Actionable and Precise Public Health: The Challenges of Regulation, Ethics and Ecology on an International Level

Benefice/Risk Balance of the Use of AI in the Context of Population Management

What have been the uses of AI in public health, in the context of the management of the Covid pandemic? We can retain the use of stammering precision public health, in particular with the use of smartphone applications for contact tracing and health passes, the analysis of social networks and mobility data to estimate the feelings, opinions and movements of populations, and finally, the possibility of using AI for social control more generally, and monitoring compliance with public health measures. All this can be based on noble inspirations and in the interest of the health of populations. However, as said, none of these components has on the one hand been the subject of specific regulations, in particular under a state of health emergency, nor on the other hand been the subject of systematic assessments of impact and effectiveness. In such a context, the risks to privacy and human rights may appear disproportionate to the expected, if not measured, benefits. The most dramatic thing would be that this first example on a global scale sets a precedent justifying repeating the same weaknesses, without drawing from it the good practices to be respected which are essential for the future.

Another striking point is precisely the aspects of regulations both for personal data and for the use of AI in health. Indeed, if the regulatory framework has recently evolved, for example at the European level, a major weakness is that the regulations are rarely, if ever, of international application. However, by nature, a pandemic is international in scale, and knows no borders—even when they are closed. There is a real problem of adequacy between the phenomenon to be stemmed and the scope of the regulatory measures.

Finally, an additional risk weighing on privacy, individual freedoms and human rights, in addition to the risks expressed just before, is due to the generalization of the adoption of a state of health emergency in many countries, making it possible to derogate from a number of usual regulations and rights, to impose restrictions that are normally impossible or exceptional. This risk is all the more tangible as this state of health emergency is currently being written into the law of various countries, de facto erasing its "exceptional" character and legitimizing the possibilities violation of human rights in the case of nations whose government control mechanisms may be weakened or failing.

The Need for a Cautious Analysis of the Real Cost of AI Use in Health

Finally, we would like to insist on two fundamental aspects concerning AI and the Covid pandemic.

The first may be due to a purely transitory phenomenon but is nonetheless a fact for this pandemic: as the MIT review reports, AI has been of no use in managing the pandemic [188]. However, both AI and the pandemic have drained and still drain many human, technical and financial investments. There is therefore a source to question.

The second is more general, and concerns the cost of digital technology and AI, from a systemic and ecological point of view. The collection, storage, access to data as well as the development for the use of AI algorithms have a very important footprint in terms of ecology. For a long time and probably still today, the use of data and in particular the secondary use of data, has been put forward as an inexpensive use: the data "are already there", all that remains is to use them This is obviously grossly false: this requires the development of additional infrastructures, the creation of new professions, the constant increase of storage and processing capacities … all in addition to what already exists in terms of digital technology.

This cost is already problematic. It must be linked to the consequences and actions suggested by the latest IPCC report on climate [189]. The use of AI has a cost and consequences. They need to be looked at for what AI has to contribute to healthcare. It would be highly paradoxical to resort to techniques that are harmful to individual and population health in a context where they would be used to … improve health.

How far should we go in our exploration of the contribution of AI to epidemiology and public health, given its early failures and the social cost associated with it?

References

1. Last JM. A dictionary of epidemiology. 4th ed. Oxford: Oxford University Press; 2001.
2. Valleron J. Les rôles de la modélisation en épidémiologie. Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie. 2000;323(5):429–33.

3. WHO. n.d.. https://www.euro.who.int/en/health-topics/Health-systems/public-health-services. Accessed 15 Mar 2022.
4. Fröhlich H, Balling R, Beerenwinkel N, et al. From hype to reality: data science enabling personalized medicine. BMC Med. 2018;16:150. https://doi.org/10.1186/s12916-018-1122-7.
5. Evans BR, Leighton FA. A history of one health. Rev Sci Tech. 2014;33:413–20. https://doi.org/10.20506/rst.33.2.2298.
6. Fassin D. Santé Publique. In: Lecourt D, editor. Dictionnaire de la pensée médicale. Paris: PUF; 2004. p. 1014–8.
7. Panorama de la santé. Quels sont les facteurs à l'origine des gains d'espérance de vie des dernières décennies? In: Analyse internationale des états membres de l'OCDE. Paris: OCDE; 2017.
8. Yuill C, Crinson I, Duncan E. The biomedical model of health. In: Key concepts in health studies. London: SAGE Publications Ltd; 2010. p. 7–10.
9. Engel GL. The need for a new medical model: a challenge for biomedicine. Science. 1977;196(4286):129–36. https://doi.org/10.1126/science.847460.
10. Bourjolay C. La responsabilité du fumeur informé. Les Echos. 2003. https://www.lesechos.fr/2003/11/la-responsabilite-du-fumeur-informe-678844. Accessed 15 Mar 2022.
11. Hood L, Friend SH. Predictive, personalized, preventive, participatory (P4) cancer medicine. Nat Rev Clin Oncol. 2011;8:184–7.
12. Syrett K. 'P4 medicine' and the purview of health law: the patient or the public? In: Kohl U, Eisler J, editors. Data-driven personalisation in markets, politics and law. Cambridge: Cambridge University Press; 2021. p. 159–73. https://doi.org/10.1017/9781108891325.013.
13. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med. 2019;25:30–6. https://doi.org/10.1038/s41591-018-0307-0.
14. Olstad DL, McIntyre L. Reconceptualising precision public health. BMJ Open. 2019;9:e030279. https://doi.org/10.1136/bmjopen-2019-030279.
15. Horton R. Offline: in defence of precision public health. Lancet. 2018;392:1504.
16. Mondschein CF, Monda C. Chapter 5. The EU's general data protection regulation (GDPR) in a research context. In: Kubben P, Dumontier M, Dekker A, editors. Fundamentals of clinical data science. Cham: Springer; 2019.
17. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Dig Med. 2020;3:118. https://doi.org/10.1038/s41746-020-00324-0.
18. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science. 2020;368:eabb6936. https://doi.org/10.1126/science.abb6936.
19. Statista. n.d.. https://fr.statista.com/statistiques/505110/taux-de-penetration-du-smartphone-par-age-france/. Accessed 15 Mar 2022
20. Manganello J, Gerstner G, Pergolino K, Graham Y, Falisi A, Strogatz D. The relationship of health literacy with use of digital technology for health information: implications for public health practice. J Public Health Manag Pract. 2017;23:380–7. https://doi.org/10.1097/PHH.0000000000000366.
21. Kamel Boulos MN, Peng G, VoPham T. An overview of GeoAI applications in health and healthcare. Int J Health Geogr. 2019;18:7. https://doi.org/10.1186/s12942-019-0171-2.
22. Huang P, MacKinlay A, Yepes AJ. Syndromic surveillance using generic medical entities on Twitter. In: Proceedings of Australasian language technology association workshop; 2016. p. 35–44.
23. Hamon T, Gagnayre R. Improving knowledge of patient skills thanks to automatic analysis of online discussions. Patient Educ Couns. 2013;92:197–204. https://doi.org/10.1016/j.pec.2013.05.012.
24. Chiolero A, Buckeridge D. Glossary for public health surveillance in the age of data science. J Epidemiol Community Health. 2020;74:612–6.

25. Kandula S, Shaman J. Reappraising the utility of Google Flu Trends. PLoS Comput Biol. 2019;15:e1007258. https://doi.org/10.1371/journal.pcbi.1007258.
26. Wongvibulsin S, Zeger SL. Enabling individualised health in learning healthcare systems. BMJ Evid Based Med. 2020;25(4):125–9. https://doi.org/10.1136/bmjebm-2019-111190.
27. Kröger M, Schlickeiser R. Analytical solution of the SIR-model for the temporal evolution of epidemics. Part A: Time-independent reproduction factor. J Phys A. 2020;53:505601. https://doi.org/10.1088/1751-8121/abc65d.
28. Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, Leskovec J. Mobility network models of COVID-19 explain inequities and inform reopening. Nature. 2021;589:82–7. https://doi.org/10.1038/s41586-020-2923-3.
29. Gomes DCDS, Serra GLO. Machine learning model for computational tracking and forecasting the COVID-19 dynamic propagation. IEEE J Biomed Health Inform. 2021;25:615–22. https://doi.org/10.1109/JBHI.2021.3052134.
30. Dolgin E. Industry embraces virtual trial platforms. Nat Rev Drug Discov. 2018;17:305–6. https://doi.org/10.1038/nrd.2018.66.
31. Desrosières A. La politique des grands nombres: histoire de la raison statistique. Paris: La Découverte; 2016. réed., 402 p.
32. Névéol A, Robert A, Anderson R, Cohen KB, Grouin C, Lavergne T, Rey G, Rondet C, Zweigenbaum P. Multilingual information extraction task overview: ICD10 coding of death certificates in English and French. In: CLEF eHealth; 2017. http://ceur-ws.org/Vol-1866/invited_paper_6.pdf. Accessed 15 Mar 2022.
33. Falissard L, Morgand C, Roussel S, Imbaud C, Ghosn W, Bounebache K, Rey G. A deep artificial neural network-based model for prediction of underlying cause of death from death certificates: algorithm development and validation. JMIR Med Inform. 2020;8:e17125. https://doi.org/10.2196/17125.
34. Abd-Alrazaq A, Alajlani M, Alhuwail D, Schneider J, Al-Kuwari S, Shah Z, et al. Artificial intelligence in the fight against COVID-19: scoping review. J Med Internet Res. 2020;22:e20756. https://doi.org/10.2196/20756.
35. Abdulkareem M, Petersen SE. The promise of AI in detection, diagnosis, and epidemiology for combating COVID-19: beyond the hype. Front Artif Intell. 2021;4:652669. https://doi.org/10.3389/frai.2021.652669.
36. Adly AS, Adly AS, Adly MS. Approaches based on artificial intelligence and the internet of intelligent things to prevent the spread of COVID-19: scoping review. J Med Internet Res. 2020;22:e19104. https://doi.org/10.2196/19104.
37. Asadzadeh A, Pakkhoo S, Saeidabad MM, Khezri H, Ferdousi R. Information technology in emergency management of COVID-19 outbreak. Inform Med Unlock. 2020;21:100475. https://doi.org/10.1016/j.imu.2020.100475.
38. Barbieri D, Giuliani E, Del Prete A, Losi A, Villani M, Barbieri A. How Artificial Intelligence and New Technologies Can Help the Management of the COVID-19 Pandemic. Int J Environ Res Public Health. 2021;18:7648. https://doi.org/10.3390/ijerph18147648.
39. Bragazzi NL, Dai H, Damiani G, Behzadifar M, Martini M, Wu J. How big data and artificial intelligence can help better manage the COVID-19 pandemic. Int J Environ Res Public Health. 2020;17:3176. https://doi.org/10.3390/ijerph17093176.
40. Chang Z, Zhan Z, Zhao Z, You Z, Liu Y, Yan Z, et al. Application of artificial intelligence in COVID-19 medical area: a systematic review. J Thorac Dis. 2021;13:7034–53. https://doi.org/10.21037/jtd-21-747.
41. Chen J, See KC. Artificial intelligence for COVID-19: rapid review. J Med Internet Res. 2020;22:e21476. https://doi.org/10.2196/21476.
42. Dogan O, Tiwari S, Jabbar MA, Guggari S. A systematic review on AI/ML approaches against COVID-19 outbreak. Complex Intell Syst. 2021;7:2655–78. https://doi.org/10.1007/s40747-021-00424-8.
43. Guo Y, Zhang Y, Lyu T, Prosperi M, Wang F, Xu H, et al. The application of artificial intelligence and data integration in COVID-19 studies: a scoping review. J Am Med Inform Assoc JAMIA. 2021;28:2050–67. https://doi.org/10.1093/jamia/ocab098.

44. Malik YS, Sircar S, Bhat S, Ansari MI, Pande T, Kumar P, et al. How artificial intelligence may help the COVID-19 pandemic: pitfalls and lessons for the future. Rev Med Virol. 2020;31:e2205. https://doi.org/10.1002/rmv.2205.
45. Naseem M, Akhund R, Arshad H, Ibrahim MT. Exploring the potential of artificial intelligence and machine learning to combat COVID-19 and existing opportunities for LMIC: a scoping review. J Prim Care Community Health. 2020;11:2150132720963634. https://doi.org/10.1177/2150132720963634.
46. Napolitano F, Xu X, Gao X. Impact of computational approaches in the fight against COVID-19: an AI guided review of 17 000 studies. Brief Bioinform. 2021;23:bbab456. https://doi.org/10.1093/bib/bbab456.
47. Shah H, Shah S, Tanwar S, Gupta R, Kumar N. Fusion of AI techniques to tackle COVID-19 pandemic: models, incidence rates, and future trends. Multimed Syst. 2021:1–34. https://doi.org/10.1007/s00530-021-00818-1.
48. Syeda HB, Syed M, Sexton KW, Syed S, Begum S, Syed F, et al. Role of machine learning techniques to tackle the COVID-19 crisis: systematic review. JMIR Med Inform. 2021;9:e23811. https://doi.org/10.2196/23811.
49. Syrowatka A, Kuznetsova M, Alsubai A, Beckman AL, Bain PA, Craig KJT, et al. Leveraging artificial intelligence for pandemic preparedness and response: a scoping review to identify key use cases. NPJ Digit Med. 2021;4:96. https://doi.org/10.1038/s41746-021-00459-8.
50. Swayamsiddha S, Prashant K, Shaw D, Mohanty C. The prospective of Artificial Intelligence in COVID-19 Pandemic. Heal Technol. 2021;11:1311–20. https://doi.org/10.1007/s12553-021-00601-2.
51. Rodríguez-Rodríguez I, Rodríguez J-V, Shirvanizadeh N, Ortiz A, Pardo-Quiles D-J. Applications of artificial intelligence, machine learning, big data and the internet of things to the COVID-19 pandemic: a scientometric review using text mining. Int J Environ Res Public Health. 2021;18:8578. https://doi.org/10.3390/ijerph18168578.
52. Tayarani NM-H. Applications of artificial intelligence in battling against COVID-19: a literature review. Chaos, Solitons Fractals. 2021;142:110338. https://doi.org/10.1016/j.chaos.2020.110338.
53. Uohara MY, Weinstein JN, Rhew DC. The essential role of technology in the public health battle against COVID-19. Popul Health Manag. 2020;23:361–7. https://doi.org/10.1089/pop.2020.0187.
54. Vaishya R, Javaid M, Khan IH, Haleem A. Artificial Intelligence (AI) applications for COVID-19 pandemic. Diab Metab Syndr. 2020;14:337–9. https://doi.org/10.1016/j.dsx.2020.04.012.
55. Xu Z, Su C, Xiao Y, Wang F. Artificial intelligence for COVID-19: battling the pandemic with computational intelligence. Intell Med. 2022;2:13–29. https://doi.org/10.1016/j.imed.2021.09.001.
56. Zhang Q, Gao J, Wu JT, Cao Z, Dajun ZD. Data science approaches to confronting the COVID-19 pandemic: a narrative review. Phil Trans A Math Phys Eng Sci. 2022;380:20210127. https://doi.org/10.1098/rsta.2021.0127.
57. Mégarbane B, Bourasset F, Scherrmann J-M. Epidemiokinetic tools to monitor lockdown efficacy and estimate the duration adequate to control SARS-CoV-2 spread. J Epidemiol Glob Health. 2021;11:321–5. https://doi.org/10.1007/s44197-021-00007-3.
58. Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, Ahmadi M, Kalhori SRN. Predicting COVID-19 incidence through analysis of google trends data in Iran: data mining and deep learning pilot study. JMIR Public Health Surveill. 2020;6:e18828. https://doi.org/10.2196/18828.
59. Qin L, Sun Q, Wang Y, Wu K-F, Chen M, Shia B-C, et al. Prediction of number of cases of 2019 novel coronavirus (COVID-19) using social media search index. Int J Environ Res Public Health. 2020;17:E2365. https://doi.org/10.3390/ijerph17072365.
60. Jahanbin K, Rahmanian V. Using twitter and web news mining to predict COVID-19 outbreak. Asian Pac J Trop Med. 2020;13:378. https://doi.org/10.4103/1995-7645.279651.

61. Liu D, Clemente L, Poirier C, Ding X, Chinazzi M, Davis JT, et al. A machine learning methodology for real-time forecasting of the 2019-2020 COVID-19 outbreak using Internet searches, news alerts, and estimates from mechanistic models. ArXiv200404019 Cs Q-Bio Stat. 2020;
62. Walker A, Hopkins C, Surda P. The use of google trends to investigate the loss of smell related searches during COVID-19 outbreak. Int Forum Allergy Rhinol. 2020;10:839. https://doi.org/10.1002/alr.22580.
63. Alhama J, Maestre JP, Martín MÁ, Michán C. Monitoring COVID-19 through SARS-CoV-2 quantification in wastewater: progress, challenges and prospects. Microb Biotechnol. 2022;15:1719. https://doi.org/10.1111/1751-7915.13989.
64. Sun K, Chen J, Viboud C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. Lancet Digit Health. 2020;2:e201–8. https://doi.org/10.1016/S2589-7500(20)30026-1.
65. McCall B. COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread. Lancet Digit Health. 2020;2:e166–7. https://doi.org/10.1016/S2589-7500(20)30054-6.
66. Johns Hopkins Coronavirus Resource Center. COVID-19 Map. Baltimore, MD: Johns Hopkins Coronavirus Resource Center; n.d.. https://coronavirus.jhu.edu/map.html. Accessed 9 Mar 2022.
67. Yang Z, Zeng Z, Wang K, Wong S-S, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J Thorac Dis. 2020;12:165–74. https://doi.org/10.21037/jtd.2020.02.64.
68. Yue M, Clapham HE, Cook AR. Estimating the size of a COVID-19 epidemic from surveillance systems. Epidemiology. 2020;31:567–9. https://doi.org/10.1097/EDE.0000000000001202.
69. Hu Z, Ge Q, Li S, Jin L, Xiong M. Artificial intelligence forecasting of COVID-19 in China. ArXiv200207112 Q-Bio. 2020;
70. Long JB, Ehrenfeld JM. The role of augmented intelligence (AI) in detecting and preventing the spread of novel coronavirus. J Med Syst. 2020;44:59. https://doi.org/10.1007/s10916-020-1536-6.
71. Gupta R, Tanwar S, Tyagi S, Kumar N. Machine learning models for secure data analytics: a taxonomy and threat model. Comput Commun. 2020;153:406–40. https://doi.org/10.1016/j.comcom.2020.02.008.
72. Tuli S, Tuli S, Tuli R, Gill SS. Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. Internet Things. 2020;11:100222. https://doi.org/10.1016/j.iot.2020.100222.
73. Ardabili SF, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, et al. COVID-19 outbreak prediction with machine learning. Algorithms. 2020;13:249. https://doi.org/10.3390/a13100249.
74. Kumari A, Vekaria D, Gupta R, Tanwar S. Redills: Deep learning-based secure data analytic framework for smart grid systems. In: 2020 IEEE Int. Conf. Commun. Workshop ICC Workshop; 2020. p. 1–6. https://doi.org/10.1109/ICCWorkshops49005.2020.9145448.
75. Magdon-Ismail M. Machine learning the phenomenology of COVID-19 from early infection dynamics. ArXiv200307602 Cs Q-Bio Stat. 2020;
76. Ahmad A, Garhwal S, Ray SK, Kumar G, Malebary SJ, Barukab OM. The number of confirmed cases of COVID-19 by using machine learning: methods and challenges. Arch Comput Methods Eng. 2021;28:2645–53. https://doi.org/10.1007/s11831-020-09472-8.
77. Kolozsvári LR, Bérczes T, Hajdu A, Gesztelyi R, Tiba A, Varga I, et al. Predicting the epidemic curve of the coronavirus (SARS-CoV-2) disease (COVID-19) using artificial intelligence: an application on the first and second waves. Inform Med Unlock. 2021;25:100691. https://doi.org/10.1016/j.imu.2021.100691.
78. Musulin J, Baressi Šegota S, Štifanić D, Lorencin I, Anđelić N, Šušteršič T, et al. Application of artificial intelligence-based regression methods in the problem of COVID-19 spread prediction: a systematic review. Int J Environ Res Public Health. 2021;18:4287. https://doi.org/10.3390/ijerph18084287.

79. Elsheikh AH, Saba AI, Panchal H, Shanmugan S, Alsaleh NA, Ahmadein M. Artificial intelligence for forecasting the prevalence of COVID-19 pandemic: an overview. Healthcare. 2021;9:1614. https://doi.org/10.3390/healthcare9121614.
80. Aslan IH, Demir M, Wise MM, Lenhart S. Modeling COVID-19: forecasting and analyzing the dynamics of the outbreak in Hubei and Turkey. Math Meth Appl Sci. 2022;45(10):6481.
81. Parbat D, Chakraborty M. A python based support vector regression model for prediction of COVID19 cases in India. Chaos, Solitons Fractals. 2020;138:109942. https://doi.org/10.1016/j.chaos.2020.109942.
82. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. 2020;20:533–4. https://doi.org/10.1016/S1473-3099(20)30120-1.
83. Zou D, Wang L, Xu P, Chen J, Zhang W, Gu Q. Epidemic model guided machine learning for COVID-19 forecasts in the United States. medRxiv:2020:2020.05.24.20111989.
84. Zheng N, Du S, Wang J, Zhang H, Cui W, Kang Z, et al. Predicting COVID-19 in China using hybrid AI model. IEEE Trans Cybern. 2020;50:2891–904. https://doi.org/10.1109/TCYB.2020.2990162.
85. Rockett RJ, Arnott A, Lam C, Sadsad R, Timms V, Gray K-A, et al. Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. Nat Med. 2020;26:1398–404. https://doi.org/10.1038/s41591-020-1000-7.
86. Car Z, Baressi Šegota S, Anđelić N, Lorencin I, Mrzljak V. Modeling the spread of COVID-19 infection using a multilayer perceptron. Comput Math Methods Med. 2020;2020:e5714714. https://doi.org/10.1155/2020/5714714.
87. Huang C-J, Chen Y-H, Ma Y, Kuo P-H. Multiple-input deep convolutional neural network model for COVID-19 forecasting in China. medRxiv:2020:2020.03.23.20041608.
88. Misra S, Deb PK, Koppala N, Mukherjee A, Mao S. S-Nav: safety-aware IoT navigation tool for avoiding COVID-19 hotspots. IEEE Internet Things J. 2021;8:6975–82. https://doi.org/10.1109/JIOT.2020.3037641.
89. Allam Z, Jones DS. On the coronavirus (COVID-19) outbreak and the smart city network: universal data sharing standards coupled with artificial intelligence (AI) to benefit urban health monitoring and management. Healthcare. 2020;8:46. https://doi.org/10.3390/healthcare8010046.
90. Zhang J, Litvinova M, Wang W, Wang Y, Deng X, Chen X, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. Lancet Infect Dis. 2020;20:793–802. https://doi.org/10.1016/S1473-3099(20)30230-9.
91. Riou J, Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. Eurosurveillance. 2020;25:2000058. https://doi.org/10.2807/1560-7917.ES.2020.25.4.2000058.
92. Tessmer HL, Ito K, Omori R. Can machines learn respiratory virus epidemiology?: a comparative study of likelihood-free methods for the estimation of epidemiological dynamics. Front Microbiol. 2018;9:343. https://doi.org/10.3389/fmicb.2018.00343.
93. Li M, Zhang Z, Cao W, Liu Y, Du B, Chen C, et al. Identifying novel factors associated with COVID-19 transmission and fatality using the machine learning approach. Sci Total Environ. 2021;764:142810. https://doi.org/10.1016/j.scitotenv.2020.142810.
94. Jehi L, Ji X, Milinovich A, Erzurum S, Rubin BP, Gordon S, et al. Individualizing risk prediction for positive coronavirus disease 2019 testing: results from 11,672 patients. Chest. 2020;158:1364–75. https://doi.org/10.1016/j.chest.2020.05.580.
95. Pramanik M, Chowdhury K, Rana MJ, Bisht P, Pal R, Szabo S, et al. Climatic influence on the magnitude of COVID-19 outbreak: a stochastic model-based global analysis. Int J Environ Health Res. 2020;32:1095–110. https://doi.org/10.1080/09603123.2020.1831446.
96. Travaglio M, Yu Y, Popovic R, Selley L, Leal NS, Martins LM. Links between air pollution and COVID-19 in England. Environ Pollut Bark. 1987;2021(268):115859. https://doi.org/10.1016/j.envpol.2020.115859.
97. Maghdid HS, Ghafoor KZ. A smartphone enabled approach to manage COVID-19 lockdown and economic crisis. SN Comput Sci. 2020;1:271.

98. Mao Z, Yao H, Zou Q, Zhang W, Dong Y. Digital contact tracing based on a graph database algorithm for emergency management during the COVID-19 epidemic: case study. JMIR MHealth UHealth. 2021;9:e26836. https://doi.org/10.2196/26836.
99. Paul M, Haque SME, Chakraborty S. Human detection in surveillance videos and its applications - a review. EURASIP J Adv Signal Process. 2013;2013:176. https://doi.org/10.1186/1687-6180-2013-176.
100. Nguyen CT, Saputra YM, Van Huynh N, Nguyen N-T, Khoa TV, Tuan BM, et al. A comprehensive survey of enabling and emerging technologies for social distancing-Part II: Emerging technologies and open issues. IEEE Access Pract Innov Open Solut. 2020;8:154209–36. https://doi.org/10.1109/ACCESS.2020.3018124.
101. Punn NS, Sonbhadra SK, Agarwal S, Rai G. Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques. ArXiv200501385 Cs. 2021;
102. Ahmed I, Ahmad M, Rodrigues JJPC, Jeon G, Din S. A deep learning-based social distance monitoring framework for COVID-19. Sustain Cities Soc. 2021;65:102571. https://doi.org/10.1016/j.scs.2020.102571.
103. Lakshmi S, Sathiyabhama B, Revathi T, Basker N, Vinothkumar R. Tracing of vehicle region and number plate detection using deep learning. In: 2020 Int Conf Emerg Trends Inf Technol Eng Ic-ETITE; 2020. https://doi.org/10.1109/ic-ETITE47903.2020.357.
104. Srinivasa Rao ASR, Vazquez JA. Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone–based survey when cities and towns are under quarantine. Infect Control Hosp Epidemiol. 2020;41:826–30. https://doi.org/10.1017/ice.2020.61.
105. LeewayHertz - Softw Dev Co. COVID-19 technology solutions. AI enabled digital solutions. n.d.. https://www.leewayhertz.com/covid-19-technology-solutions/. Accessed 9 Mar 2022.
106. Gupta M, Abdelsalam M, Mittal S. Enabling and enforcing social distancing measures using smart city and ITS infrastructures: a COVID-19 use case. ArXiv200409246 Cs. 2020;
107. Penang Institute. Smart city technologies take on COVID-19. n.d.. https://penanginstitute.org/publications/issues/smart-city-technologies-take-on-covid-19/. Accessed 9 Mar 2022
108. Singh SK, Jeong Y-S, Park JH. A deep learning-based IoT-oriented infrastructure for secure smart City. Sustain Cities Soc. 2020;60:102252. https://doi.org/10.1016/j.scs.2020.102252.
109. Chen Q, Leaman R, Allot A, Luo L, Wei C-H, Yan S, et al. Artificial intelligence in action: addressing the COVID-19 pandemic with natural language processing. Annu Rev Biomed Data Sci. 2021;4:313–39. https://doi.org/10.1146/annurev-biodatasci-021821-061045.
110. Tsao S-F, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: a scoping review. Lancet Digit Health. 2021;3:e175–94. https://doi.org/10.1016/S2589-7500(20)30315-0.
111. Husnayain A, Fuad A, Su EC-Y. Applications of Google Search Trends for risk communication in infectious disease management: a case study of the COVID-19 outbreak in Taiwan. Int J Infect Dis. 2020;95:221–3. https://doi.org/10.1016/j.ijid.2020.03.021.
112. Chundakkadan R, Ravindran R. Information flow and COVID-19 recovery. World Dev. 2020;136:105112. https://doi.org/10.1016/j.worlddev.2020.105112.
113. Song J, Song TM, Seo D-C, Jin D-L, Kim JS. Social big data analysis of information spread and perceived infection risk during the 2015 middle east respiratory syndrome outbreak in South Korea. Cyberpsychol Behav Soc Netw. 2017;20:22–9. https://doi.org/10.1089/cyber.2016.0126.
114. Bento AI, Nguyen T, Wing C, Lozano-Rojas F, Ahn Y-Y, Simon K. Evidence from internet search data shows information-seeking responses to news of local COVID-19 cases. Proc Natl Acad Sci U S A. 2020;117:11220–2. https://doi.org/10.1073/pnas.2005335117.
115. Saha K, Torous J, Caine ED, De Choudhury M. Psychosocial effects of the COVID-19 pandemic: large-scale quasi-experimental study on social media. J Med Internet Res. 2020;22:e22600. https://doi.org/10.2196/22600.

116. Cullen W, Gulati G, Kelly BD. Mental health in the COVID-19 pandemic. QJM Int J Med. 2020;113:hcaa110. https://doi.org/10.1093/qjmed/hcaa110.
117. Usher K, Durkin J, Bhullar N. The COVID-19 pandemic and mental health impacts. Int J Ment Health Nurs. 2020;29:315–8. https://doi.org/10.1111/inm.12726.
118. Zhang X, Yang Q, Albaradei S, Lyu X, Alamro H, Salhi A, et al. Rise and fall of the global conversation and shifting sentiments during the COVID-19 pandemic. Humanit Soc Sci Commun. 2021;8:1–10. https://doi.org/10.1057/s41599-021-00798-7.
119. Zhang Y, Lyu H, Liu Y, Zhang X, Wang Y, Luo J. Monitoring depression trends on twitter during the COVID-19 pandemic: observational study. JMIR Infodemiol. 2021;1:e26769. https://doi.org/10.2196/26769.
120. Koh JX, Liew TM. How loneliness is talked about in social media during COVID-19 pandemic: text mining of 4,492 Twitter feeds. J Psychiatr Res. 2022;145:317–24. https://doi.org/10.1016/j.jpsychires.2020.11.015.
121. COVID-evidence. Database. n.d.. https://covid-evidence.org/database. Accessed 9 Mar 2022.
122. Else H. How a torrent of COVID science changed research publishing — in seven charts. Nature. 2020;588:553. https://doi.org/10.1038/d41586-020-03564-y.
123. Joshi B, Bakarola V, Shah P, Krishnamurthy R. deepMINE - natural language processing based automatic literature mining and research summarization for early-stage comprehension in pandemic situations specifically for COVID-19. bioRxiv:2020:2020.03.30.014555.
124. Roberts K, Alam T, Bedrick S, Demner-Fushman D, Lo K, Soboroff I, et al. TREC-COVID: rationale and structure of an information retrieval shared task for COVID-19. J Am Med Inform Assoc JAMIA. 2020;27:1431–6. https://doi.org/10.1093/jamia/ocaa091.
125. Wise C, Ioannidis VN, Calvo MR, Song X, Price G, Kulkarni N, et al. COVID-19 knowledge graph: accelerating information retrieval and discovery for scientific literature. ArXiv200712731 Cs. 2020;
126. Khalil H, Tamara L, Rada G, Akl EA. Challenges of evidence synthesis during the 2020 COVID pandemic: a scoping review. J Clin Epidemiol. 2022;142:10–8. https://doi.org/10.1016/j.jclinepi.2021.10.017.
127. Alper BS, Richardson JE, Lehmann HP, Subbian V. It is time for computable evidence synthesis: the COVID-19 knowledge accelerator initiative. J Am Med Inform Assoc JAMIA. 2020;27:1338–9. https://doi.org/10.1093/jamia/ocaa114.
128. Bell RJ. Evidence synthesis in the time of COVID-19. Climacteric. 2021;24:211–3. https://doi.org/10.1080/13697137.2021.1904676.
129. Rada GG, Verdugo-Paiva F, Ávila C, Morel-Marambio M, Bravo-Jeria R, Pesce F, et al. Evidence synthesis relevant to COVID-19: a protocol for multiple systematic reviews and overviews of systematic reviews. Medwave. 2020;20:e7868. https://doi.org/10.5867/medwave.2020.03.7867.
130. Cuan-Baltazar JY, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the internet: infodemiology study. JMIR Public Health Surveill. 2020;6:e18444. https://doi.org/10.2196/18444.
131. Zhao Y, Zhou CC. Applying lexical link analysis to discover insights from public information on COVID-19. bioRxiv:2020:2020.05.06.079798.
132. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. Nat Med. 2020;26:459–61. https://doi.org/10.1038/s41591-020-0824-5.
133. Mashamba-Thompson TP, Crayton ED. Blockchain and artificial intelligence technology for novel coronavirus disease 2019 self-testing. Diagnostics. 2020;10:198. https://doi.org/10.3390/diagnostics10040198.
134. Lin Z, Xu P, Winata GI, Siddique FB, Liu Z, Shin J, et al. CAiRE: an end-to-end empathetic chatbot. Proc AAAI Conf Artif Intell. 2020;34:13622–3. https://doi.org/10.1609/aaai.v34i09.7098.
135. Battineni G, Chintalapudi N, Amenta F. AI Chatbot design during an epidemic like the novel coronavirus. Healthcare. 2020;8:E154. https://doi.org/10.3390/healthcare8020154.

136. Rodsawang C, Thongkliang P, Intawong T, Sonong A, Thitiwatthana Y, Chottanapund S. Designing a competent chatbot to counter the COVID-19 pandemic and empower risk communication in an emergency response system. Outb Surveill Investig Resp J. 2020;13(2):71.
137. Chatterjee P, Tesis A, Cymberknop LJ, Armentano RL. Internet of things and artificial intelligence in healthcare during COVID-19 pandemic—a south American perspective. Front Public Health. 2020;8:600213. https://doi.org/10.3389/fpubh.2020.600213.
138. Li F, Valero M, Shahriar H, Khan RA, Ahamed SI. Wi-COVID: a COVID-19 symptom detection and patient monitoring framework using WiFi. Smart Health. 2021;19:100147. https://doi.org/10.1016/j.smhl.2020.100147.
139. Taha-Mehlitz S, Hendie A, Taha A. The development of electronic health and artificial intelligence in surgery after the SARS-CoV-2 pandemic-a scoping review. J Clin Med. 2021;10:4789. https://doi.org/10.3390/jcm10204789.
140. Di Carlo F, Sociali A, Picutti E, Pettorruso M, Vellante F, Verrastro V, et al. Telepsychiatry and other cutting-edge technologies in COVID-19 pandemic: bridging the distance in mental health assistance. Int J Clin Pract. 2021;75:e13716. https://doi.org/10.1111/ijcp.13716.
141. Hallak JA, Scanzera AC, Azar DT, Chan RVP. Artificial intelligence in ophthalmology during COVID-19 and in the post COVID-19 era. Curr Opin Ophthalmol. 2020;31:447–53. https://doi.org/10.1097/ICU.0000000000000685.
142. Asgary A, Valtchev SZ, Chen M, Najafabadi MM, Wu J. Artificial Intelligence model of drive-through vaccination simulation. Int J Environ Res Public Health. 2020;18:E268. https://doi.org/10.3390/ijerph18010268.
143. Bubar KM, Reinholt K, Kissler SM, Lipsitch M, Cobey S, Grad YH, et al. Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. Science. 2021;371:916–21. https://doi.org/10.1126/science.abe6959.
144. Cheng F-Y, Joshi H, Tandon P, Freeman R, Reich DL, Mazumdar M, et al. Using machine learning to predict ICU transfer in hospitalized COVID-19 patients. J Clin Med. 2020;9:1668. https://doi.org/10.3390/jcm9061668.
145. Shashikumar SP, Wardi G, Paul P, Carlile M, Brenner LN, Hibbert KA, et al. Development and prospective validation of a deep learning algorithm for predicting need for mechanical ventilation. Chest. 2021;159:2264–73. https://doi.org/10.1016/j.chest.2020.12.009.
146. Rozario D. Can machine learning optimize the efficiency of the operating room in the era of COVID-19? Can J Surg. 2020;63:E527–9. https://doi.org/10.1503/cjs.016520.
147. Chang SL, Harding N, Zachreson C, Cliff OM, Prokopenko M. Modelling transmission and control of the COVID-19 pandemic in Australia. Nat Commun. 2020;11:5710. https://doi.org/10.1038/s41467-020-19393-6.
148. Broniec W, An S, Rugaber S, Goel AK. Using VERA to explain the impact of social distancing on the spread of COVID-19. ArXiv200313762 Cs Q-Bio. 2020;
149. Soures N, Chambers D, Carmichael Z, Daram A, Shah DP, Clark K, et al. SIRNet: understanding social distancing measures with hybrid neural network model for COVID-19 infectious spread. ArXiv200410376 Q-Bio. 2020;
150. Hoertel N, Blachier M, Blanco C, Olfson M, Massetti M, Rico MS, et al. A stochastic agent-based model of the SARS-CoV-2 epidemic in France. Nat Med. 2020;26:1417–21. https://doi.org/10.1038/s41591-020-1001-6.
151. Payedimarri AB, Concina D, Portinale L, Canonico M, Seys D, Vanhaecht K, et al. Prediction models for public health containment measures on COVID-19 using artificial intelligence and machine learning: a systematic review. Int J Environ Res Public Health. 2021;18:4499. https://doi.org/10.3390/ijerph18094499.
152. Dandekar R, Barbastathis G. Quantifying the effect of quarantine control in COVID-19 infectious spread using machine learning. medRxiv:2020:2020.04.03.20052084.
153. Porcher S. Response2covid19, a dataset of governments' responses to COVID-19 all around the world. Sci Data. 2020;7:423. https://doi.org/10.1038/s41597-020-00757-y.
154. Miralles-Pechuán L, Jiménez F, Ponce H, Martínez-Villaseñor L. A deep Q-learning/genetic algorithms based novel methodology for optimizing COVID-19 pandemic government actions. ArXiv200507656 Phys Stat. 2020;

155. Ye Y, Hou S, Fan Y, Zhang Y, Qian Y, Sun S, et al. α-Satellite: an AI-driven system and benchmark datasets for dynamic COVID-19 risk assessment in the United States. IEEE J Biomed Health Inform. 2020;24:2755–64. https://doi.org/10.1109/JBHI.2020.3009314.
156. Norouzi N, Zarazua de Rubens G, Choupanpiesheh S, Enevoldsen P. When pandemics impact economies and climate change: exploring the impacts of COVID-19 on oil and electricity demand in China. Energy Res Soc Sci. 2020;68:101654. https://doi.org/10.1016/j.erss.2020.101654.
157. Guan D, Wang D, Hallegatte S, Davis SJ, Huo J, Li S, et al. Global supply-chain effects of COVID-19 control measures. Nat Hum Behav. 2020;4:577–87. https://doi.org/10.1038/s41562-020-0896-8.
158. Lu D, Zhang J, Xue C, Zuo P, Chen Z, Zhang L, et al. COVID-19-induced lockdowns indicate the short-term control effect of air pollutant emission in 174 cities in China. Environ Sci Technol. 2021;55:4094–102. https://doi.org/10.1021/acs.est.0c07170.
159. Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, et al. Mobility network models of COVID-19 explain inequities and inform reopening. Nature. 2021;589:82–7. https://doi.org/10.1038/s41586-020-2923-3.
160. Sousa-Pinto B, Anto A, Czarlewski W, Anto JM, Fonseca JA, Bousquet J. Assessment of the impact of media coverage on COVID-19–related google trends data: infodemiology study. J Med Internet Res. 2020;22:e19611. https://doi.org/10.2196/19611.
161. Tummers J, Catal C, Tobi H, Tekinerdogan B, Leusink G. Coronaviruses and people with intellectual disability: an exploratory data analysis. J Intellect Disabil Res. 2020;64:475–81. https://doi.org/10.1111/jir.12730.
162. Wang C, Zhao H, Zhang H. Chinese college students have higher anxiety in new semester of online learning during COVID-19: a machine learning approach. Front Psychol. 2020;11:587413. https://doi.org/10.3389/fpsyg.2020.587413.
163. Ćosić K, Popović S, Šarlija M, Kesedžić I. Impact of human disasters and COVID-19 pandemic on mental health: potential of digital psychiatry. Psychiatr Danub. 2020;32:25–31. https://doi.org/10.24869/psyd.2020.25.
164. Jha IP, Awasthi R, Kumar A, Kumar V, Sethi T. Learning the mental health impact of COVID-19 in the United States with explainable artificial intelligence. JMIR Ment Health. 2021;8:e25097.
165. Bursztyn L, Rao A, Roth CP, Yanagizawa-Drott DH. Misinformation during a pandemic. Cambridge, MA: National Bureau of Economic Research; 2020. https://doi.org/10.3386/w27417.
166. Lyu H, Zheng Z, Luo J. Misinformation versus facts: understanding the influence of news regarding COVID-19 vaccines on vaccine uptake. Health Data Sci. 2022;2022:9858292.
167. Rovetta A, Bhagavathula AS. COVID-19-related web search behaviors and infodemic attitudes in Italy: infodemiological study. JMIR Public Health Surveill. 2020;6:e19374. https://doi.org/10.2196/19374.
168. Chen T, Wu L, Li X, Zhang J, Yin H, Wang Y. Call attention to rumors: deep attention based recurrent neural networks for early rumor detection. ArXiv170405973 Cs. 2017;
169. Ruchansky N, Seo S, Liu Y. CSI: a hybrid deep model for fake news detection. In: Proc 2017 ACM Conf Inf Knowl Manag; 2017. p. 797–806. https://doi.org/10.1145/3132847.3132877.
170. Depoux A, Martin S, Karafillakis E, Preet R, Wilder-Smith A, Larson H. The pandemic of social media panic travels faster than the COVID-19 outbreak. J Travel Med. 2020;27:taaa031. https://doi.org/10.1093/jtm/taaa031.
171. Tasnim S, Hossain MM, Mazumder H. Impact of rumors and misinformation on COVID-19 in social media. J Prev Med Public Health. 2020;53:171–4. https://doi.org/10.3961/jpmph.20.094.
172. Bin Naeem S, Kamel Boulos MN. COVID-19 misinformation online and health literacy: a brief overview. Int J Environ Res Public Health. 2021;18:8091. https://doi.org/10.3390/ijerph18158091.

173. Uyheng J, Carley KM. Characterizing network dynamics of online hate communities around the COVID-19 pandemic. Appl Netw Sci. 2021;6:1–21. https://doi.org/10.1007/s41109-021-00362-x.
174. Paka WS, Bansal R, Kaushik A, Sengupta S, Chakraborty T. Cross-SEAN: a cross-stitch semi-supervised neural attention model for COVID-19 fake news detection. Appl Soft Comput. 2021;107:107393. https://doi.org/10.1016/j.asoc.2021.107393.
175. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med. 2015;13:1. https://doi.org/10.1186/s12916-014-0241-z.
176. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, Reitsma JS, Collins GS. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014;11(10):e1001744. https://doi.org/10.1371/journal.pmed.1001744.
177. Uyheng J, Carley KM. Bots and online hate during the COVID-19 pandemic: case studies in the United States and the Philippines. J Comput Soc Sci. 2020;3:445–68. https://doi.org/10.1007/s42001-020-00087-4.
178. Mankins J. Technology readiness levels, a white paper. Washington, DC: NASA; 1995.
179. Santi P, Larousserie D, Cabut S, Herzberg N, Morin H. Coronavirus: le « lancetgate »révèle des failles de l'édition scientifique. Le Monde. 2020. https://www.lemonde.fr/sciences/article/2020/06/15/covid-19-le-lancetgate-revele-des-failles-de-l-edition-scientifique_6042946_1650684.html. Accessed 15 Mar 2022.
180. Kahneman D. Thinking, fast and slow. New York, NY: Penguin; 2012. 512 p.
181. Davies A, Veličković P, Buesing L, Blackwell S, Zheng D, Tomašev N, Tanburn R, Battaglia P, Blundell C, Juhász A, Lackenby M, Williamson G, Hassabis D, Kohli P. Advancing mathematics by guiding human intuition with AI. Nature. 2021;600:70–4. https://doi.org/10.1038/s41586-021-04086-x.
182. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20:310. https://doi.org/10.1186/s12911-020-01332-6.
183. OECD. Whistleblower protection. Paris: OECD; n.d.. https://www.oecd.org/corruption/ethics/whistleblower-protection/. Accessed 15 Mar 2022.
184. Leplâtre S, Lemaître F. Emotion en Chine à l'annonce de la mort du docteur Li Wenliang, lanceur d'alerte sur le coronavirus. Le Monde. 2020. https://www.lemonde.fr/international/article/2020/02/06/emotion-a-l-annonce-de-la-mort-du-docteur-li-wenliang-lanceur-d-alerte-du-coronavirus_6028699_3210.html. Accessed 15 Mar 2022.
185. Meyer-Vacherand E. Emprisonnée, la lanceuse d'alerte chinoise Zhang Zhan serait entre la vie et la mort. Le Temps. 2021. https://www.letemps.ch/monde/emprisonnee-lanceuse-dalerte-chinoise-zhang-zhan-serait-entre-vie-mort. Accessed 15 Mar 2022.
186. Latour B. Où suis-je? Paris: Empêcheurs de penser en rond; 2021. 192 p.
187. Simondon G. Du mode d'existence des objets techniques. Paris: Aubier; 2021. 368 p.
188. Heaven WD. Hundreds of AI tools have been built to catch COVID. None of them helped. MIT Technol Rev. 2021. https://www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic/. Accessed 15 Mar 2022.
189. IPPC. 5th report on climate changes. https://www.ipcc.ch/languages-2/francais/publications/. Accessed 15 Mar 2022.

Index

N. Lidströmer, Y. C. Eldar (eds.), *Artificial Intelligence in Covid-19*,
https://doi.org/10.1007/978-3-031-08506-2