Check for updates

The role of machine learning in health policies during the COVID-19 pandemic and in long COVID management

Lindybeth Sarmiento Varón[1†], Jorge González-Puelma[1,2†],
David Medina-Ortiz[3], Jacqueline Aldridge[3],
Diego Alvarez-Saravia[1,2], Roberto Uribe-Paredes[3] and
Marcelo A. Navarrete[1,2]*

[1]Centro Asistencial Docente y de Investigación, Universidad de Magallanes, Punta Arenas, Chile,
[2]Escuela de Medicina, Universidad de Magallanes, Punta Arenas, Chile, [3]Departamento de Ingeniería en
Computación, Facultad de Ingeniería, Universidad de Magallanes, Punta Arenas, Chile

The ongoing COVID-19 pandemic is arguably one of the most challenging health crises in modern times. The development of effective strategies to control the spread of SARS-CoV-2 were major goals for governments and policy makers. Mathematical modeling and machine learning emerged as potent tools to guide and optimize the different control measures. This review briefly summarizes the SARS-CoV-2 pandemic evolution during the first 3 years. It details the main public health challenges focusing on the contribution of mathematical modeling to design and guide government action plans and spread mitigation interventions of SARS-CoV-2. Next describes the application of machine learning methods in a series of study cases, including COVID-19 clinical diagnosis, the analysis of epidemiological variables, and drug discovery by protein engineering techniques. Lastly, it explores the use of machine learning tools for investigating long COVID, by identifying patterns and relationships of symptoms, predicting risk indicators, and enabling early evaluation of COVID-19 sequelae.

KEYWORDS

COVID-19, public health policies, mathematical models, machine learning, long COVID, SARS-CoV-2

1. Introduction

Mathematical models help to understand the functioning and dynamics of a given system trough equations and rules, as such, can simulate conditions and scenarios associated with multiple public policies, non-pharmaceutical interventions (NPI), and vaccine performance (1). Therefore, mathematical models became major tools for guiding the decision-making of governments and health systems during the pandemic (2). This section briefly introduces SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) and describes relevant events during the progress pandemic. We then summarize the main applications of mathematical models and the various uses to describe the transmission behaviors of SARS-CoV-2.

1.1. What is the SARS-CoV-2?

SARS-CoV-2 is the pathogen causing the 2019 coronavirus disease (COVID-19). COVID-19 manifestations range from mild flu symptoms to severe acute respiratory syndrome. SARS-CoV-2 virion contains a 29Kb RNA genome wrapped in a capsid covered by the Spike, the main protein responsible for the high infection rate (3). During the transmission between humans, the genome accumulates mutations, generating variants with selective advantages that predominate in different countries (4). Intrinsic factors like transmissibility and natural mutation rate, host factors such as age, risk group, immunity, and socio-cultural factors like economy, culture, and current levels of globalization have determined the coronavirus evolution. Integrating SARS-CoV-2 data is essential to predict its behavior, prevent its continuous expansion, and understand this disease. Three years after the pandemic, the scientific community has generated an unprecedented amount of data, now facing the challenge of translating this data into knowledge.

1.2. The first 3 years of the pandemic

The first reported case of COVID-19 was in December 2019 (5). With the exponential increase in infections worldwide, the World Health Organization (WHO) declared the disease a pandemic in March 2020. Governments adopted different NPI to mitigate the virus's high reproduction rates. These measures included face masks, social distancing, and lockdowns. While these measures were implemented worldwide, just a few countries, such as Vietnam and New Zealand, demonstrated the complete -although transitory- elimination of the transmission (6). In April and May 2020, the first predictions of the pandemic course were based on statistical models performed by the Institute for Health Metrics and Evaluation and provided a reasonable projection in the short-time (7). During the first wave, it was also possible to establish that 10% of the cases were responsible for 80% of the secondary infections, indicating a high heterogeneity in transmission spread as compared to other pathogens (8).

In the first pandemic year, it was identified that social contact in public transport or closed areas allowed high transmission rates (9, 10). In turn, it was determined that face masks reduce droplet particle transmission (11). Furthermore, NPI was essential to flatten the spread curve in the first year of the pandemic preventing new waves of cases after curves pick, limiting overcrowding of hospital beds, and giving time to improve treatment strategies (12). Adaptations of the Susceptible-Infected-Recovered models helped to demonstrate the NPI effectiveness in preventing the transmission of the virus. Besides, these same models allowed the detection of an increase in virus circulation with the relaxation of the measures (13). Other models, facilitated the test-track-isolation developing strategies to prevent the spread, demonstrating that efficient track strategies help to reduce the number of new cases (14). At the same time, the first signs of SARS-CoV-2 genetic adaptation arose between March and May 2020, with the emerging D614G variant, which showed clear worldwide transmissibility advan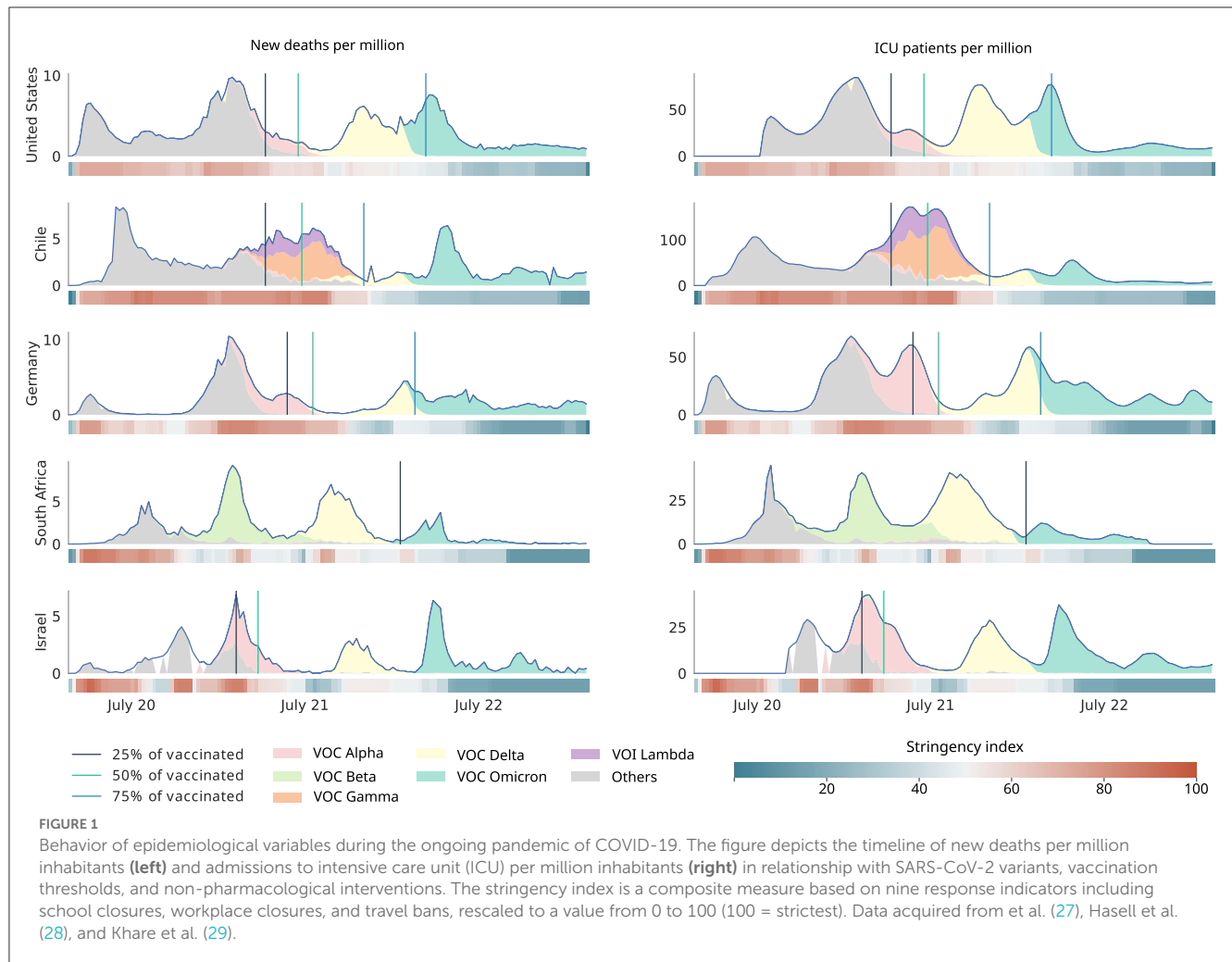tages (15). The control of the pandemic at that time relied on the development of herd immunity, being established that the necessary protection of the population is approximately $1 - \frac{1}{R_0}$, being estimated at 67% of the people (16). In August 2020, reinfection cases demonstrated that natural immunity only provides temporary protection (17). In December 2020, the first clinical trials of vaccines were developed, leading to the emergency approval of traditional and novel vaccine formulations -such as mRNA vaccines-. These studies quickly established that immunity begins between 10 and 14 days after the first dose (18). A second dose shows protection over 90%, preventing hospitalizations and deaths (19). The vaccines can block propagation, making cases less infectious, with a 92% reduction in transmission rates (20). At the same time, quantitative models pointed to the possibility of immune escape when complete schemes are not generated. At the end of 2020, the Alpha variant (B.1.1.7), according to the WHO terminology, was the variant responsible for the significant increase in cases in the United Kingdom. This variant was characterized by presenting spike mutations with binding advantages to the ACE-2 receptor (21), showing clear selection advantages, a phenomenon observed simultaneously in different parts of the world (22, 23).

The subsequent variant of similar global relevance was Delta (B.1.617.2), characterized by its high replicative capacity. Vaccine effectiveness studies showed protection against Alpha and Delta variants (24). Vaccination programs were effective reducing deaths, hospitalization admission, and intensive care unit (ICU) occupancy (see Figure 1). In November 2021, a new outbreak was reported in South Africa, caused by a new circulating variant presenting a 60–70 spike gene deletion. This variant was called Omicron (B1.1.529) and expanded rapidly throughout the world, replacing the Delta variant. Omicron carries more than 30 spike mutations (25), being responsible for high worldwide reinfection rates (26). Vaccines have also shown a protective effect against this variant, although deaths were reported among unvaccinated individuals. Omicron subvariant (XBB1.5) has been described as responsible for 40.5% confirmed cases in the EE.UU. as of late December 2022. It has also been observed that recombinant XBB and BA.2 Omicron subvariant strains, widely spread in Asia, do not show different symptoms than the previous variants, nor do they show signs of being more severe than their predecessors.

Figure 1 summarizes the key variables depicting the pandemic evolution in five exemplary cases. Each country showed different spread behaviors of SARS-CoV-2. The measures showed variable effectiveness. In most countries, other public health policies and government plans were applied to mitigate the effects of the spread. However, in most cases, the fatalities decreased after implementation.

1.3. Applications of mathematical modeling during the pandemic

The SIR models (Susceptible, Infected, and Recovered) are spread dynamics analysis models used during the early days of the pandemic (30). SEIR models (Susceptible-Exposed-Infected-Recovered) correspond to an adapted SIR model to understand propagation mechanisms (31). These models do not account for heterogeneity within the population, thus novel strategies

FIGURE 1
Behavior of epidemiological variables during the ongoing pandemic of COVID-19. The figure depicts the timeline of new deaths per million inhabitants **(left)** and admissions to intensive care unit (ICU) per million inhabitants **(right)** in relationship with SARS-CoV-2 variants, vaccination thresholds, and non-pharmacological interventions. The stringency index is a composite measure based on nine response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (100 = strictest). Data acquired from et al. (27), Hasell et al. (28), and Khare et al. (29).

incorporated a component of population subdivision into multiple groups and interconnected systems, allowing the representation of several mechanisms of interaction between different sub-populations by a multi-group SEIRA (Susceptible-Exposed-Infected-Recovered and Asymptomatic Model) (32). Another interesting development was the statistically-based temporal reclassification of cases. This approach allowed more precise modeling of SARS-CoV-2 propagation dynamics, by correcting errors in diagnostic test reporting times and infection time registries (33, 34).

With the application of NPI strategies to prevent the spread of SARS-CoV-2, the mathematical models were adapted to incorporate this new knowledge. This adaptations enabled the anticipation of the effect of NPI relaxation measures in function of epidemiological variables, such as levels of hospitalization, use of ICU, and lethality (35). SEIRA models also helped to asses the effect of vaccines and pharmaceutical interventions (36).

With the first vaccination plans and high immunization rates started the relaxation of public policies (37). However, the ability of the virus to mutate and generate variants was associated with new peaks in cases incidence. Mathematical models were adapted to this scenario by incorporating information on genomic surveillance

programs, spread of variants, and the effects of immunization (38–40).

Altogether, mathematical tools proved its relevance in modeling the behavior of propagation systems and their effect on populations. The SIR classical model as well as different adaptations such as SEIR, SEIRA, and others, contributed significantly to the development of government plans and public health policies. Nevertheless, traditional mathematical modeling strategies rely on existing knowledge and cannot account for dynamics not explicitly incorporated during modeling. Methods based on machine learning (ML) and artificial intelligence (AI) can overcome these intrinsic limitations by generating autonomous systems that learn from the modeled dynamics to predict new behaviors and adapt to unknown scenarios.

1.4. Vaccines developments, efficacy, and adverse effects

Population immunity is considered a landmark for epidemic control. Since immunity through natural infection might result in unacceptable morbidity and mortality, the development of efficient

COVID-19 vaccination programs was a prioritary public policy for most countries (41, 42). The race to develop highly effective and safe vaccines resulted in various platforms allowing their implementation at unprecedented speed (43–45).

Due the modest response of traditional vaccines against other coronaviruses such as Middle East Respiratory Syndrome Coronavirus (MERS) and Severe Acute Respiratory Syndrome (SARS), the development of novel formulations was a major scientific goal (42, 46). A new vaccine technology based on mRNA technology emerged as candidates in late December 2020 and two formulations granted emergency approval BNT162b2 (Pfizer-BioNTech), and mRNA-1273 (Moderna) (47). The developed vaccines showed promising results in reducing transmissibility and the probability of death, reaching an efficacy > 90% in phase III clinical trials (48).

The widespread immunization poses the challenge of quantifying and understanding short- and long-term toxicity for novel vaccine formulations. Most studies have shown short-term safety in the general population. However, in certain groups, severe adverse events were reported i) anaphylaxis (2.5–4.8 cases per million adult vaccine doses administered) (49, 50), ii) myocarditis (52.4 cases and 56.3 cases per million doses) (51), iii) thrombosis with thrombocytopenia syndrome (2-4 cases per one million doses administered) (52), and iv) Guillain-Barré syndrome (7.8 cases per million) (53), as well as an association with multisystemic inflammatory syndrome (54).

A major challenge is to reliably detect long-term effects that might occur at different rates in different patients subgroups (55). Causal association becomes difficult due to the high immunization rates achieved in most countries. In this complex scenario mathematical models, ML, and AI, could provide powerful tools provided that public policies focus on collection of sufficient high-quality data.

1.5. What is long COVID?

1.5.1. Characteristics and definitions of long COVID

Long COVID (LC) is a novel multi-systemic disease defined by the persistence or appearance of a wide variety of symptoms with variable intensity, regardless of the initial disease severity by probable or confirmed SARS-CoV-2 infection (56). In response to the absence of a consensus definition, the WHO proposed using the term Post-COVID-19 listed in the ICD-10 classification based on the Delphi consensus (57). This condition usually manifests 3 months after the SARS-CoV-2 infection, the symptoms last for at least 2 months in the absence of alternative diagnosis (58).

The National Institute for Health Research, classifies LC into i) post-intensive care syndrome (post-ICU syndrome), ii) post-viral fatigue syndrome, iii) permanent organ damage, iv) decompensation of previous chronic diseases, v) the onset of a new disease triggered by COVID-19, and vi) pharmacological toxicity from COVID-19 treatment (59).

Other authors had suggested six post-COVID syndrome subsets, including i) non-severe COVID-19 multiorgan sequelae, ii)

pulmonary fibrosis sequelae, iii) myalgic encephalomyelitis/chronic fatigue syndrome, iv) postural orthostatic tachycardia syndrome, v) post-intensive care syndrome, and vi) medical or clinical sequelae (60).
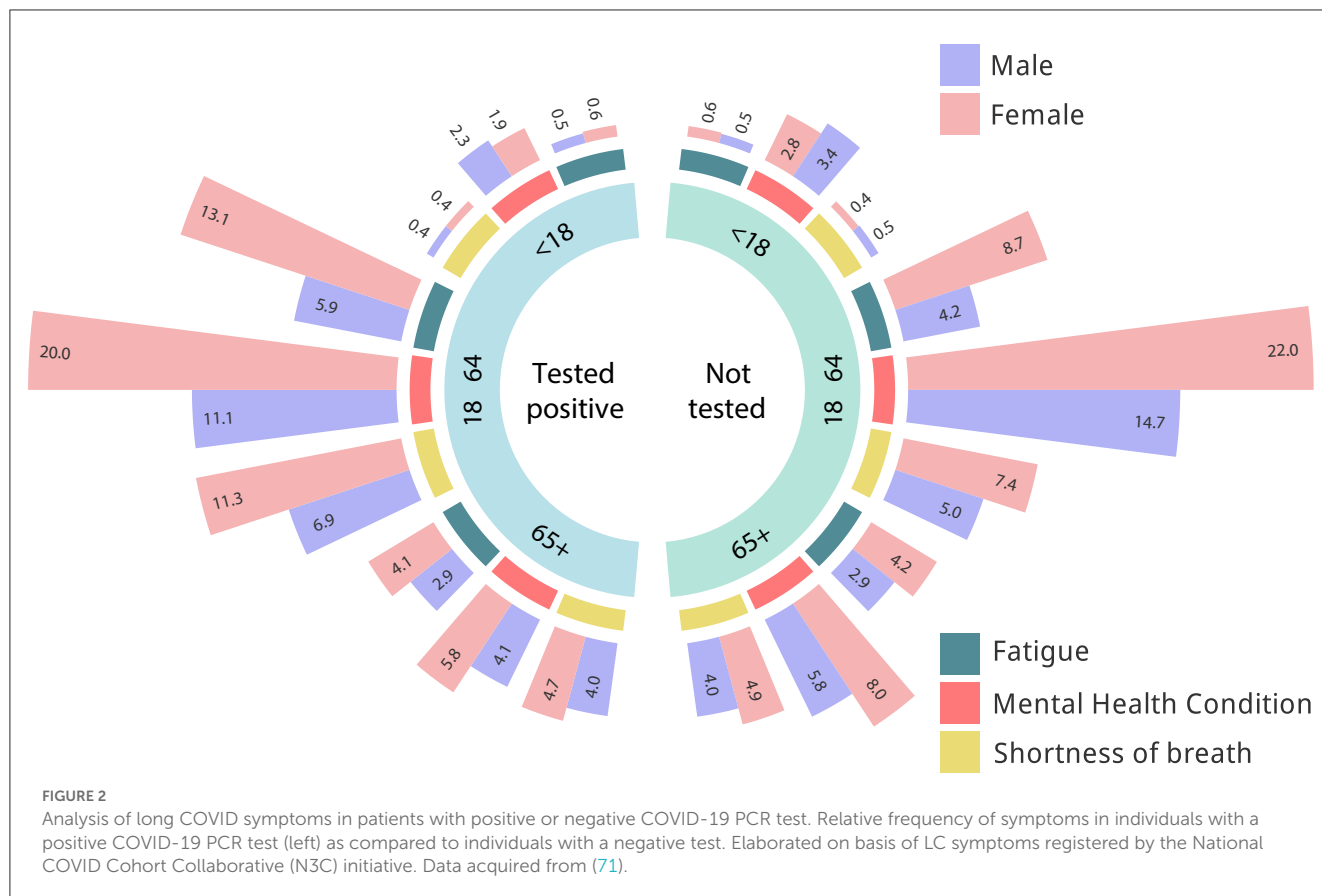
1.5.2. Symptoms and incidence of long COVID

Between 2.3 and 60% of COVID-19 survivors could experience LC symptoms during the first year, and up to 42% 2 years after the infection (61–63). Patients with LC present variable symptoms, including fatigue (29%), muscle pain, palpitations, cognitive impairment (28%), dyspnea (21%), anxiety (27%), chest pain, and arthralgia (18%) (see Figure 2) (64). Other patients report respiratory system dysfunction (26%), or cardiovascular complications (32–89%) 3 months after the onset of infection (65–67). Gastrointestinal symptoms have been associated with an imbalance of gut microbiota, as well as psychological and central nervous system effects (68, 69). Most of these symptoms are associated with a reduction in the quality of life. However, the distinction between SARS-CoV-2-related symptoms to those linked to other, often pre-existing conditions remains extremely challenging. As clinical studies addressing this issues take a long time to develop the NIH launched the Rapid Acceleration of Diagnostics initiative, and the NIH LC Computational Challenge (70). This initiative aims to use AI and ML to predict which patients with SARS-CoV-2 infections are most likely to develop LC. Figure 2 depicts the relative frequency of LC symptoms registered by the National COVID Cohort Collaborative (N3C) initiative. Inviduals that tested positive for SARS-CovV-2 show a higher frequency of alterations in symptoms such as fatigue and shortness of breath. The prevalence of these symptoms seems higher in women. However, the small magnitude of the differences highlights the challenge of differentiation long COVID from other conditions.

2. Machine learning application to COVID-19

During the COVID-19 pandemic, ML methods have played a relevant role in the development of diagnostic strategies (72, 73), forecasting the epidemiological behavior (74), and as a tool to support the development and monitoring of public health policies (75). Figure 3 summarizes the most relevant ML applications during the COVID-19 pandemic.

2.1. COVID-19 diagnosis

Different strategies based on ML algorithms were designed during the COVID-19 pandemic to elaborate predictive models of efficient clinical diagnosis (76). The main inputs used to build the models are based on images, sounds, respiratory information, symptoms, and mixed data (77). Convolutional neural networks (CNN) architectures are commonly employed to develop classification models via image inputs (e.g., x-ray, CT-chest, and ultrasounds) (78). Sounds from respiratory information, such as cough and breath, were common inputs for the development of

FIGURE 2
Analysis of long COVID symptoms in patients with positive or negative COVID-19 PCR test. Relative frequency of symptoms in individuals with a positive COVID-19 PCR test (left) as compared to individuals with a negative test. Elaborated on basis of LC symptoms registered by the National COVID Cohort Collaborative (N3C) initiative. Data acquired from (71).
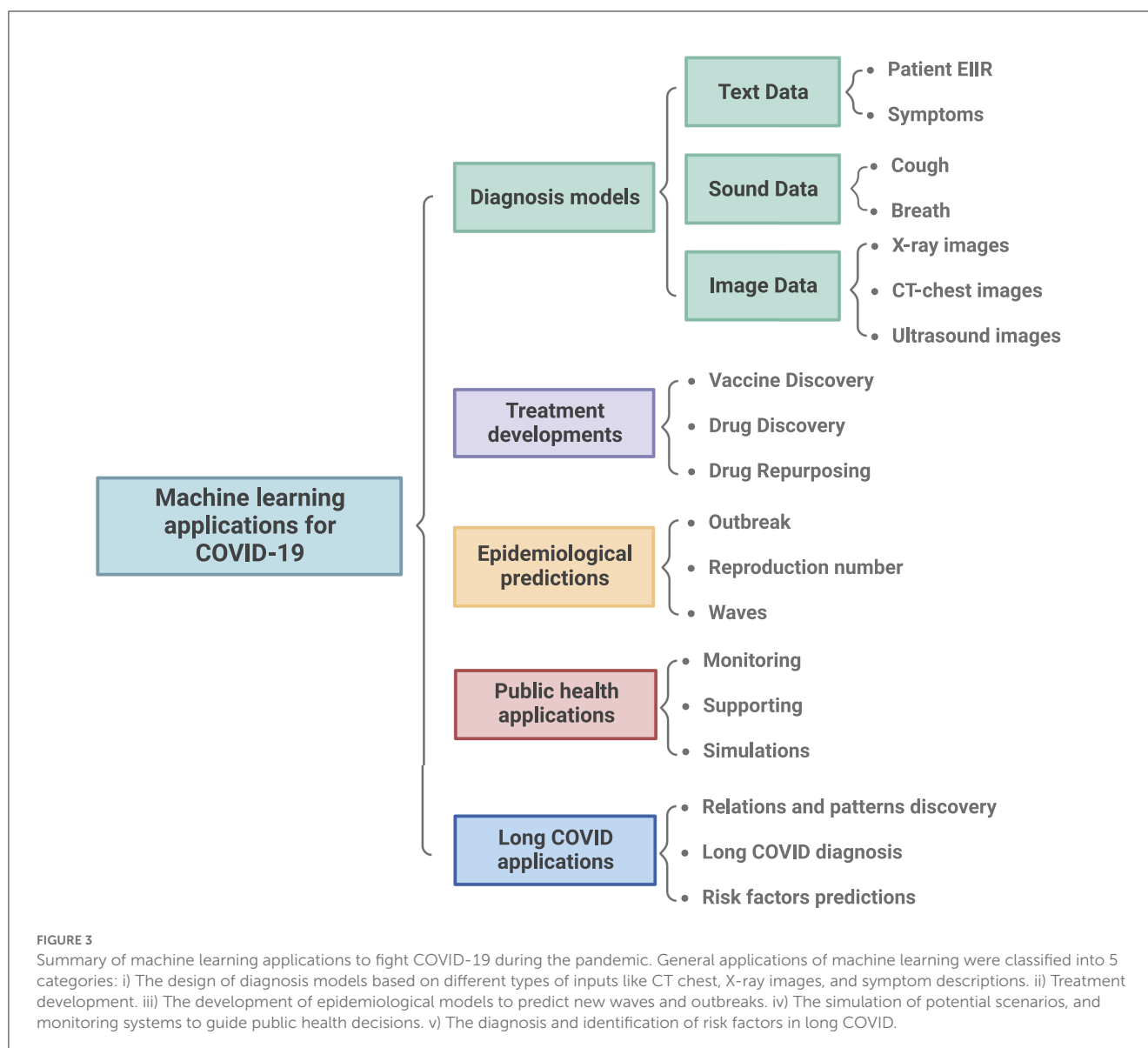
predictive models employing recurrent neural network (RNN) or long short-term memory architectures (LSTM), since this type of architectures have the advantage to maintain the information on signal frequencies (79). Hybrid methods that combine symptoms and clinical diagnostic tests as inputs facilitate the development of more complex predictions models or classifications systems. The hybrid methods include not only vector information or matrix spaces, but also data on disease's propagation. The incorporation of virus characteristics, close contacts, and contagion networks using graph neural networks results in highly efficient prediction systems (80).

To demonstrate the usability of classification models based on ML techniques, a clinical diagnostic model using CT chest images was developed following the architecture proposed in Figure 4 and updating our previously reported method for CT chest images classification (34). Generally, models based on CNN architectures can be divided into three large blocks: i) pattern processing and extraction, ii) learning, and iii) classification blocks. To extract patterns, a set of three layers composed of CNN, batch normalization, max pooling, and dropout, was developed. Then, a flattened layer is used to prepare the inputs to the fully connected or dense layers, which are part of the learning block, composed of dense layers interspersed with batch normalization, ending with a dropout layer. Finally, a last layer of classification is added to develop the outputs. As activation functions, ReLU and SoftMax

were used. In addition, binary cross entropy associated with an Adam optimizer was used as a cost function. A total of 2,482 images were used to train the diagnostic model extracted from (81). For the training process, a classic validation approach was followed by segmentation of the training and validation data set (80:20), and the TensorFlow framework was employed for its implementation (82). Model training was followed for a total of 10 epochs. The proposed architecture achieved a precision of 99.81% and 0.027 loss function, demonstrating the high performance obtained by the proposed architecture. The implemented model can be used as a support strategy for clinical diagnosis in patients with COVID-19. Besides, it is possible to apply transfer learning techniques to use the same images and the same architecture proposed to estimate the probability that patients present sequelae, one of the most recent areas of study associated with the concept of LC.

2.2. COVID-19 treatments and strategies to prevent adverse effects

ML applications related to the design of treatment strategies have focused on drug discovery, drug repurposing, and vaccine discovery methods (83). For drug repurposing, algorithms are usually based on networks of knowledge graphs including virus and host interactions (84). These strategies have used

FIGURE 3
Summary of machine learning applications to fight COVID-19 during the pandemic. General applications of machine learning were classified into 5 categories: i) The design of diagnosis models based on different types of inputs like CT chest, X-ray images, and symptom descriptions. ii) Treatment development. iii) The development of epidemiological models to predict new waves and outbreaks. iv) The simulation of potential scenarios, and monitoring systems to guide public health decisions. v) The diagnosis and identification of risk factors in long COVID.
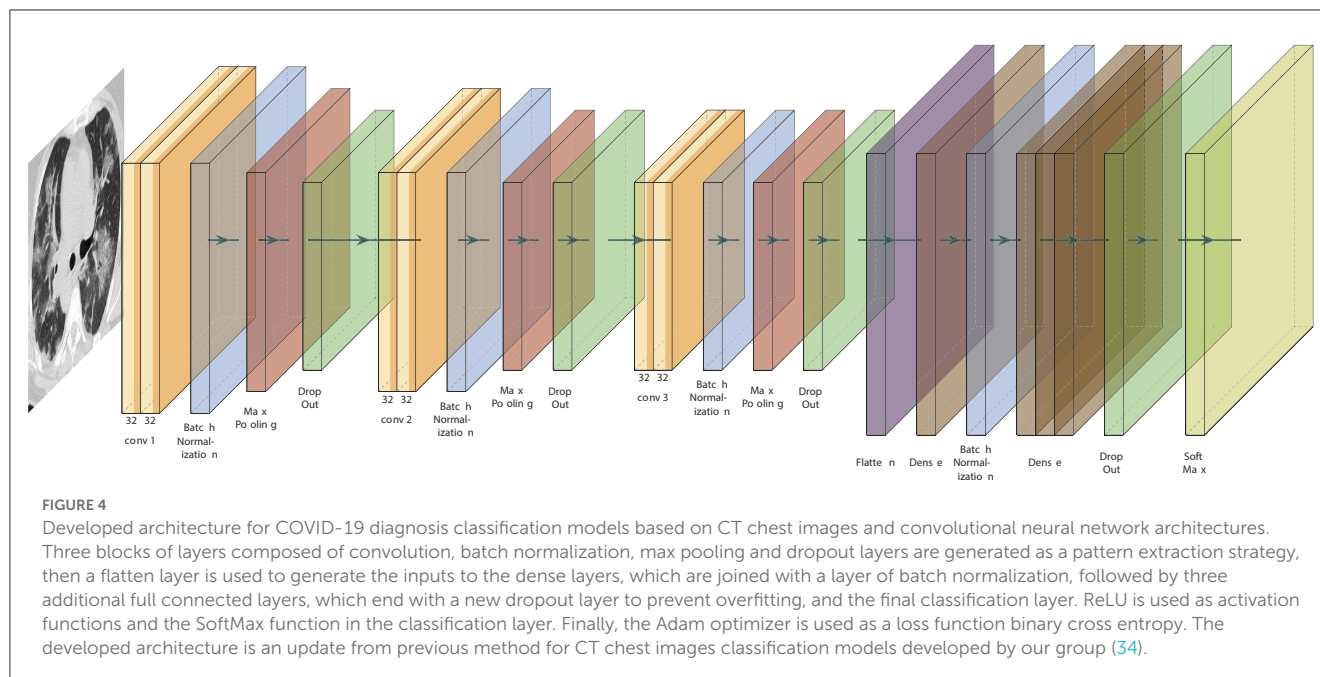
particular network label propagation combined with semi-supervised learning method based on regularized Laplacian to identify interactors of SARS-CoV-2 (85). Another example is the elaboration of predictive systems based on protein-protein interaction to estimate affinity between two elements (86). This issue has been addressed by either CNN or graph convolutional neural networks (GCNN) architectures. Protein complexes are typically represented using strategies based on topological information (87), solvent accessible surface (SAS) (88), voxel-based molecular surface representation (89), and various molecular descriptors (90).

Another of the traditional drug repurposing methods are the gene expression based algorithms (83). The changes in the expression levels of defensive genes in disease states can be used as phenotypic descriptors or quantifiers of the transcriptomic effects of the explored drugs. Besides, methods based on integrated docking simulation algorithms have made it possible to optimize drug repurposing systems (91).

Different computational tools have been developed for drug and vaccine discovery. Zhavoronkov et al. (92) developed a generative chemistry pipeline based on the knowledge of protein, molecule structures, and homology models strategies to identify new drugs related to SARS-CoV-2. Tang et al. (93) have built processes based on deep learning (DL) algorithms to design new antivirus drugs of a chemical or peptide nature based on the information available in the literature and different chemical rules.

Molecular simulations using docking techniques allowed the development of virtual screening methodologies and iterative searches to discover new drugs of interest. The discovery of new chemical compounds with desirable activities is possible by combining the structural information with strategies of deep generative models (94, 95).

Predictive models using the linear protein sequences and the chemical compounds represented as SMILES have been proposed to predict affinity between proteins and chemical compounds (96). Different numerical representations strategies have been

FIGURE 4
Developed architecture for COVID-19 diagnosis classification models based on CT chest images and convolutional neural network architectures. Three blocks of layers composed of convolution, batch normalization, max pooling and dropout layers are generated as a pattern extraction strategy, then a flatten layer is used to generate the inputs to the dense layers, which are joined with a layer of batch normalization, followed by three additional full connected layers, which end with a new dropout layer to prevent overfitting, and the final classification layer. ReLU is used as activation functions and the SoftMax function in the classification layer. Finally, the Adam optimizer is used as a loss function binary cross entropy. The developed architecture is an update from previous method for CT chest images classification models developed by our group (34).

implemented to encode the protein sequences, such as binarization coding, physicochemical properties, and Fourier transforms to represent protein sequences in spaces of signals (97). Alternatively, methods based on natural language processing (protein language models) have been developed (98). In the case of SMILES, different autoencoders and transformers strategies have been created, including variational autoencoders and graph junction trees (99).

Performance between methods based on linear sequences information and those that only incorporate structural details are similar. However, the processes that use representations based on NLP seem to present a higher performance because the autoencoders manage to learn the structural relationships that guide the function (100). Nevertheless, the learning strategies and the abilities to extract complex patterns from the information used for the development of predictive models are properties of DL methods that, to date, have not been fully understood due to their functioning as black boxes. The incorporation of techniques based on explainable AI, is under development to understand the underlying functions and mechanics of the ML algorithms (101).

Concerning the strategies to prevent the adverse effect provoked by the vaccination programs, ML analyzes revealed distinct arterial pulse variability according to side effects of mRNA vaccine. This can facilitate a time-saving and easy-to-use method for detecting changes in the vascular properties associated with cardiovascular side effects following vaccination (102).

The application of explainable ML techniques has allowed to detect relevant variables to perform predictive models with hight performances. Abbaspour et al. (103) applied SHAP strategies combined with XGB model to identify important predictors (e.g., demographics, any history of allergy, any prior COVID diagnosis or positive test, vaccine manufacturer, and time-of-day-of-vaccination) associated to COVID-19 vaccine-related side effects.

Analyzes of the Vaccine Adverse Event Reporting System datasets with ML and a statistical approaches identified and classified pre-existing factors as having an impact on post-vaccination morbidity and reactogenicity (104). Nevertheless, this information is limited because the main databases do not have a larger record size and do not cover all types of vaccines, provoking problems in the generalization of the identified behaivors.

2.3. COVID-19 epidemiology

The design and implementation of ML models used for predicting epidemiological variables was a significant challenge. The need of high volumes of data to generalize the behavior of the predictive models (105), made necessary to develop methods for optimizing the representation of the inputs by autoencoders or embedding (106). The developed models were generated to promote the implementation of computer systems for the simulation of scenarios (107) and to facilitate the elaboration of government public policies focused on preventing the increase in the number of contagious or the outbreak of new waves (108).

Depending on the input type, the construction of predictive models can be based on forecasting methods using strategies such as ARIMA (109, 110) or LSTM architectures (111). Other strategies were based on logistic regression methods (112), nonlinear regressions (113), autoregressive models (114), and Gaussian Process Regression (115). The inputs used to develop the predictive models contemplate information based on time series and consider contagion spread records, NPI, scenarios, and different types of crucial information related to epidemiological variables. Mathematical methods based on linear algebra and kernel applications were used to combine the different kinds of data in hybrid systems elaborated with RNN and CNN architectures (116).

2.4. COVID-19 public health

One essential use of ML strategies was combining mathematical models to develop hybrid knowledge systems to support decisions in public health. These systems can be classified mainly into monitoring applications and simulation systems (116). Concerning monitoring tools, predictive models allow the generation of early alerts of behaviors during a pandemic. These alerts were usually related to predicting waves and new contagion outbreaks. More limited strategies but with significant impact were the methodologies to forecast the level of ICU occupancy in hospitals and health systems and their correlation with increases in contagion rates and mutational variants since it allowed early warning of the occupancy level and facilitated decision-making to prevent a whole occupancy level (117).

The simulation of scenarios by ML allowed the evaluation of public policy effect on populations of interest (118). Despite the versatility of ML, dynamic changes in the knowledge embedded in the system—NPI modifications, the application of vaccine programs, emergence of SARS-CoV-2 variants, etc- makes necessary a constant adaptation of ML based models. Incorporation of reinforced learning might help to facilitate this process.

2.5. Application to long COVID

With the emergence of LC, ML methods have been employed for the development of predictive tools, the construction of statistical systems for relating patient phenotypes, and the elaboration of rules and complex patterns to understand the interactions between systems and types of sequelae. The application of unsupervised learning algorithms like k-means and kernel representations strategies enabled to correlate symptoms and different classifications of LC (119).

Based on data from the N3C electronic health record repository, Pfaff et al. (119) have developed an ML model to classify the likelihood of LC diagnosis. Using XGBoost machine learning algorithm this study identified a series of features, including the healthcare utilization rate, patient age, dyspnea or respiratory symptoms, other pre-existing risk factors (diabetes, kidney disease, congestive heart failure, or pulmonary disease), and treatment medication information to predict LC.

Binka et al. (120) proposed a classification model based on elastic net penalized logistic regression algorithms for classifying patients as positive or negative for LC. The model proposed by Binka et al. (120) employed as descriptors demographic characteristics, pre-existing conditions, COVID-19 related data, and all symptoms/conditions recorded >28–183 days after the COVID-19 symptom onset/reported.

Fritsche et al. (121) described associations from the previous and acute medical phenomena of COVID-19 as predisposing diagnoses for LC employing statistical and relation features models. Performed phenomenon-wide association studies (PheWa) and Phenotype Risk Scores (PheRS) have uncovered a plethora of diagnoses associated with LC. These studies associated seven phenotypes with the pre-COVID-19 period (e.g., irritable bowel syndrome, concussion, nausea/vomiting, and shortness of breath)

and 69 acute-COVID-19 phenotypes (predominantly respiratory and circulatory phenotypes) significantly associated with LC. Using PheRS, a quarter of the COVID-19 positive cohort was identified with a 3.5-fold increased risk of LC compared to the bottom 50% of their distributions (121).

Sengupta et al. (122) proposed an interpretable DL approach based on Gradient-weighted Class Activation Mapping using N3C and RECOVER data to predict risk factors contributing to the development of LC. This model used a temporally ordered list of diagnostic codes six weeks post-COVID-19 infection for each patient, with an accuracy of 70.48%. Gupta et al. (123) proposed a stacking ensemble learning technique based on deep neural networks for early predicting cardiovascular disease risk in recovered SARS-CoV-2 patients with LC symptoms, achieving an accuracy of 93.23%.

The here reviewed studies highlight the versatility of ML methods to study LC, facilitating not only the implementation of predictive diagnostic tools but also encouraging the integration of clinical data with, social, demographic and other information, for the development of robust systems. Despite the versatility of ML techniques, there are still enormous challenges for their application in LC analysis, in particular the collection of meaningful data sets for the development of predictive systems.

3. Discussion

Mathematical models have helped to understand the dynamics of the spread of SARS-CoV-2 and helped to predict different scenarios during the COVID-19 pandemic, becoming one of the most relevant tools for developing public health policies. Correlating sanitary measures with virus variants and the effects on the reproduction rate enabled the assessment of government policies that will help to face new outbreaks of SARS-CoV-2 or future pandemics. The development of reliable mathematical models, statistical techniques for test correction, and methods of analysis of heterogeneous populations, together with the value of testing strategies and traceability of close contacts, has been remarkable achievements. Combining these systems with ML and AI methods increased the predictive power of the models and facilitated the simulation of scenarios.

Developing predictive systems for COVID-19 was one of the significant challenges assumed by thousands of scientists during the pandemic. The main achievements were developing models for clinical diagnostic systems, ML for drug and vaccine discovery, and forecasting models for epidemiological variables to support public health policies and monitoring systems. In turn, the development of predictive systems coupled with techniques such as protein language models and molecular techniques facilitated the study of variants at the genomic level. Such models helped to understand how mutations affected critical viral proteins, helping drug and vaccine designs.

The ongoing pandemic has introduced a complete set of challenges, and currently, a novel multisystem disease defined by the persistence or appearance of new symptoms after SARS-CoV-2 infection has emerged. This complex entity-denominated LC has yet to be fully elucidated, mainly because

it is characterized by a wide range of clinical manifestations, methodological limitations, and heterogeneous definitions that make clinical and computational analysis difficult. Despite rapidly emerging studies and growing evidence, current data needs to be improved. A primary task is to establish an approach to identify natural language data associated with potential LC patients. This task will likely require well-designed prospective studies, unified definitions of LC, an accurate distinction of SARS-CoV-2-related symptoms, and adequate follow-up times that include current patients, underrepresented groups, children, and minority populations. It is granted that ML strategies will play a critical role in the understanding of LC and other upcoming challenges of the ongoing SARS-CoV-2 pandemic.

Author contributions

LS, JG-P, and DM-O: conceptualization. DM-O, DA-S, and JA: methodology. DM-O and MN: validation. LS, JG-P, DM-O, JA, and DA-S: investigation. LS, DM-O, JG-P, and MN: writing, review, and editing. MN and RU-P: supervision, funding resources, and project administration. All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Shankar S, Mohakuda SS, Kumar A, Nazneen P, Yadav AK, Chatterjee K, et al. Systematic review of predictive mathematical models of COVID-19 epidemic. *Med J Armed Forces India*. (2021) 77:S385–392. doi: 10.1016/j.mjafi.2021.05.005

2. Contreras S, Medina-Ortiz D, Conca C, Olivera-Nappa Á. A novel synthetic model of the glucose-insulin system for patient-wise inference of physiological parameters from small-size OGTT data. *Front Bioeng Biotechnol*. (2020) 8:195. doi: 10.3389/fbioe.2020.00195

3. Huang Y, Yang C, Xu XF, Xu W, Liu SW. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol Sin*. (2020) 41:1141–9. doi: 10.1038/s41401-020-0485-4

4. Magazine N, Zhang T, Wu Y, McGee MC, Veggiani G, Huang W. Mutations and evolution of the SARS-CoV-2 spike protein. *Viruses*. (2022) 14:640. doi: 10.3390/v14030640

5. Velavan TP, Meyer CG. The COVID-19 epidemic. *Trop Med Int Health*. (2020) 25:278. doi: 10.1111/tmi.13383

6. Baker MG, Wilson N. The covid-19 elimination debate needs correct data. *BMJ*. (2020) 371. doi: 10.1136/bmj.m3883

7. Holmdahl I, Buckee C. Wrong but useful–what covid-19 epidemiologic models can and cannot tell us. *N Engl J Med*. (2020) 383:303–5. doi: 10.1056/NEJMp2016822

8. Wu Z, Harrich D, Li Z, Hu D, Li D. The unique features of SARS-CoV-2 transmission: comparison with SARS-CoV, MERS-CoV and 2009 H1N1 pandemic influenza virus. *Rev Med Virol*. (2021) 31:e2171. doi: 10.1002/rmv.2171

9. Li Y, Campbell H, Kulkarni D, Harpur A, Nundy M, Wang X, et al. The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries. *Lancet Infect Dis*. (2021) 21:193–202. doi: 10.1016/S1473-3099(20)30785-4

10. Brauner JM, Mindermann S, Sharma M, Johnston D, Salvatier J, Gavenčiak T, et al. Inferring the effectiveness of government interventions against COVID-19. *Science*. (2021) 371:eabd9338. doi: 10.1126/science.abd9338

11. Howard J, Huang A, Li Z, Tufekci Z, Zdimal V, van der Westhuizen HM, et al. An evidence review of face masks against COVID-19. *Proc Natl Acad Sci USA*. (2021) 118:e2014564118. doi: 10.1073/pnas.2014564118

12. Cobey S. Modeling infectious disease dynamics. *Science*. (2020) 368:713–714. doi: 10.1126/science.abb5659

13. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*. (2020) 584:257–61. doi: 10.1038/s41586-020-2405-7

14. Larremore DB, Wilder B, Lester E, Shehata S, Burke JM, Hay JA, et al. Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening. *Sci Adv*. (2021) 7:eabd5393. doi: 10.1126/sciadv.abd5393

15. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell*. (2020) 182:812–27. doi: 10.1016/j.cell.2020.06.043

16. Fine PE. Herd immunity: history, theory, practice. *Epidemiol Rev*. (1993) 15:265–302. doi: 10.1093/oxfordjournals.epirev.a036121

17. To KKW, Hung IFN, Ip JD, Chu AWH, Chan WM, Tam AR, et al. Coronavirus disease 2019 (COVID-19) re-infection by a phylogenetically distinct severe acute respiratory syndrome coronavirus 2 strain confirmed by whole genome sequencing. *Clin Infect Dis*. (2020) 73:e2946–51. doi: 10.1093/cid/ciaa1275

18. Saad-Roy CM, Morris SE, Metcalf CJE, Mina MJ, Baker RE, Farrar J, et al. Epidemiological and evolutionary considerations of SARS-CoV-2 vaccine dosing regimes. *Science*. (2021) 372:363–70. doi: 10.1126/science.abg8663

19. Tartof SY, Slezak JM, Fischer H, Hong V, Ackerson BK, Ranasinghe ON, et al. Effectiveness of mRNA BNT162b2 COVID-19 vaccine up to 6 months in a large integrated health system in the USA: a retrospective cohort study. *Lancet*. (2021) 398:1407–16. doi: 10.1016/S0140-6736(21)02183-8

20. Tang P, Hasan MR, Chemaitelly H, Yassine HM, Benslimane FM, Al Khatib HA, et al. BNT162b2 and mRNA-1273 COVID-19 vaccine effectiveness against the SARS-CoV-2 Delta variant in Qatar. *Nat Med*. (2021) 27:2136–43. doi: 10.1038/s41591-021-01583-4

21. Prunas O, Warren JL, Crawford FW, Gazit S, Patalon T, Weinberger DM, et al. Vaccination with BNT162b2 reduces transmission of SARS-CoV-2 to household contacts in Israel. *Science*. (2022) 375:1151–4. doi: 10.1126/science.abl4292

22. González-Puelma J, Aldridge J, Montes de Oca M, Pinto M, Uribe-Paredes R, Fernández-Goycoolea J, et al. Mutation in a SARS-CoV-2 haplotype from sub-Antarctic Chile reveals new insights into the Spikeâ ĂŹs dynamics. *Viruses*. (2021) 13:883. doi: 10.3390/v13050883

23. Acevedo ML, Gaete-Argel A, Alonso-Palomares L, de Oca MM, Bustamante A, Gaggero A, et al. Differential neutralizing antibody responses elicited by CoronaVac

and BNT162b2 against SARS-CoV-2 Lambda in Chile. *Nat Microbiol*. (2022) 7:524–9. doi: 10.1038/s41564-022-01092-1

24. Lopez Bernal J, Andrews N, Gower C, Gallagher E, Simmons R, Thelwall S, et al. Effectiveness of Covid-19 vaccines against the B. 1.617. 2 (Delta) variant. *N Engl J Med*. (2021) 385:585–94. doi: 10.1056/NEJMoa2108891

25. Kristiansen H, Gad HH, Eskildsen-Larsen S, Despres P, Hartmann R. The oligoadenylate synthetase family: an ancient protein family with multiple antiviral activities. *J Interferon Cytokine Res*. (2011) 31:41–7. doi: 10.1089/jir.2010.0107

26. Pulliam JR, van Schalkwyk C, Govender N, von Gottberg A, Cohen C, Groome MJ, et al. Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa. *Science*. (2022) 376:eabn4947.

27. Mathieu E, Ritchie H, Ortiz-Ospina E, Roser M, Hasell J, Appel C, et al. A global database of COVID-19 vaccinations. *Nat Hum Behav*. (2021) 5:947–53. doi: 10.1038/s41562-021-01122-8

28. Hasell J. A cross-country database of COVID-19 testing. *Sci Data*. (2020) 7:345. doi: 10.1038/s41597-020-00688-8

29. Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, et al. GISAIDâĂŹs role in pandemic response. *China CDC Weekly*. (2021) 3:1049. doi: 10.46234/ccdcw2021.255

30. Calafiore GC, Novara C, Possieri C. A modified SIR model for the COVID-19 contagion in Italy. In: *2020 59th IEEE Conference on Decision and Control (CDC)*. Jeju: IEEE (2020). p. 3889–94.

31. Annas S, Pratama MI, Rifandi M, Sanusi W, Side S. Stability analysis and numerical simulation of SEIR model for pandemic COVID-19 spread in Indonesia. *Chaos Solitons Fractals*. (2020) 139:110072. doi: 10.1016/j.chaos.2020.110072

32. Contreras S, Villavicencio HA, Medina-Ortiz D, Biron-Lattes JP, Olivera-Nappa Á. A multi-group SEIRA model for the spread of COVID-19 among heterogeneous populations. *Chaos Solitons Fractals*. (2020) 136:109925. doi: 10.1016/j.chaos.2020.109925

33. Contreras S, Biron-Lattes JP, Villavicencio HA, Medina-Ortiz D, Llanovarced-Kawles N, Olivera-Nappa Á. Statistically-based methodology for revealing real contagion trends and correcting delay-induced errors in the assessment of COVID-19 pandemic. *Chaos Solitons Fractals*. (2020) 139:110087. doi: 10.1016/j.chaos.2020.110087

34. Sanchez-Daza A, Medina-Ortiz D, Olivera-Nappa A, Contreras S. COVID-19 modeling under uncertainty: statistical data analysis for unveiling true spreading dynamics and guiding correct epidemiological management. In: *Modeling, Control and Drug Development for COVID-19 Outbreak Prevention*. Springer (2022). p. 245–82. Available online at: https://link.springer.com/chapter/10.1007/978-3-030-72834-2_9

35. Bauer S, Contreras S, Dehning J, Linden M, Iftekhar E, Mohr SB, et al. Relaxing restrictions at the pace of vaccination increases freedom and guards against further COVID-19 waves. *PLoS Comput Biol*. (2021) 17:e1009288. doi: 10.1371/journal.pcbi.1009288

36. Contreras S, Dehning J, Mohr SB, Bauer S, Spitzner FP, Priesemann V. Low case numbers enable long-term stable pandemic control without lockdowns. *Sci Adv*. (2021) 7:eabg2243. doi: 10.1126/sciadv.abg2243

37. Contreras S, Priesemann V. Risking further COVID-19 waves despite vaccination. *Lancet Infect Dis*. (2021) 21:745–6. doi: 10.1016/S1473-3099(21)00167-5

38. Oróstica KY, Contreras S, Mohr SB, Dehning J, Bauer S, Medina-Ortiz D, et al. Mutational signatures and transmissibility of SARS-CoV-2 Gamma and Lambda variants. *arXiv preprint* arXiv:210810018. (2021). doi: 10.48550/arXiv.2108.10018

39. Contreras S, Oróstica KY, Daza-Sanchez A, Wagner J, Dönges P, Medina-Ortiz D, et al. Model-based assessment of sampling protocols for infectious disease genomic surveillance. *Chaos Solitons Fractals*. (2023) 167:113093. doi: 10.1016/j.chaos.2022.113093

40. Oróstica KY, Contreras S, Sanchez-Daza A, Fernandez J, Priesemann V, Olivera-Nappa Á. New year, new new SARS-CoV-2 variant: resolutions on genomic surveillance protocols to face Omicron. *Lancet Regional Health Am*. (2022) 7:100203. doi: 10.1016/j.lana.2022.100203

41. Hall VJ, Foulkes S, Saei A, Andrews N, Oguti B, Charlett A, et al. COVID-19 vaccine coverage in health-care workers in England and effectiveness of BNT162b2 mRNA vaccine against infection (SIREN): a prospective, multicentre, cohort study. *Lancet*. (2021) 397:1725–735. doi: 10.1016/S0140-6736(21)00790-X

42. Joshi G, Borah P, Thakur S, Sharma P, Mayank, Poduri R. Exploring the COVID-19 vaccine candidates against SARS-CoV-2 and its variants: where do we stand and where do we go? *Hum Vaccines Immunotherapeut*. (2021) 17:4714–40. doi: 10.1080/21645515.2021.1995283

43. Forni G, Mantovani A. COVID-19 vaccines: where we stand and challenges ahead. *Cell Death Diff*. (2021) 28:626–39. doi: 10.1038/s41418-020-00720-9

44. Dai L, Gao GF. Viral targets for vaccines against COVID-19. *Nat Rev Immunol*. (2021) 21:73–82. doi: 10.1038/s41577-020-00480-0

45. Rawat K, Kumari P, Saha L. COVID-19 vaccine: a recent update in pipeline vaccines, their design and development strategies. *Eur J Pharmacol*. (2021) 892:173751. doi: 10.1016/j.ejphar.2020.173751

46. Kyriakidis NC, López-Cortés A, González EV, Grimaldos AB, Prado EO. SARS-CoV-2 vaccines strategies: a comprehensive review of phase 3 candidates. *npj Vaccines*. (2021) 6:28. doi: 10.1038/s41541-021-00292-w

47. Baden L, El Sahly H, Essink B, et al. Učinkovitost in varnost cepiva mRNA-1273 SARS-CoV-2. *N Engl J Med*. (2021) 384:403–416. doi: 10.1056/NEJMoa2035389

48. Sa M, Bukhari IA, Akram J, Meo AS, Klonoff DC. COVID-19 vaccines: comparison of biological, pharmacological characteristics and adverse effects of Pfizer/BioNTech and Moderna Vaccines. *Eur Rev Med Pharmacol Sci*. (2021) 25:1663–9. doi: 10.26355/eurrev_202102_24877

49. Klein NP, Lewis N, Goddard K, Fireman B, Zerbo O, Hanson KE, et al. Surveillance for adverse events after COVID-19 mRNA vaccination. *JAMA*. (2021) 326:1390–9. doi: 10.1001/jama.2021.15072

50. Shimabukuro T. Allergic reactions including anaphylaxis after receipt of the first dose of Pfizer-BioNTech COVID-19 vaccine–United States, December 14-23, 2020. *Am J Transpl*. (2021) 21:1332. doi: 10.1111/ajt.16516

51. Friedensohn L, Levin D, Fadlon-Derai M, Gershovitz L, Fink N, Glassberg E, et al. Myocarditis following a third BNT162b2 vaccination dose in military recruits in Israel. *JAMA*. (2022) 327:1611–2. doi: 10.1001/jama.2022.4425

52. See I, Su JR, Lale A, Woo EJ, Guh AY, Shimabukuro TT, et al. US case reports of cerebral venous sinus thrombosis with thrombocytopenia after Ad26. COV2. S vaccination, March 2 to April 21, 2021. *JAMA*. (2021) 325:2448–56. doi: 10.1001/jama.2021.7517

53. Hanson KE, Goddard K, Lewis N, Fireman B, Myers TR, Bakshi N, et al. Incidence of Guillain-Barré syndrome after COVID-19 vaccination in the vaccine safety datalink. *JAMA Network Open*. (2022) 5:e228879-e228879. doi: 10.1001/jamanetworkopen.2022.8879

54. Grome HN, Threlkeld M, Threlkeld S, Newman C, Martines RB, Reagan-Steiner S, et al. Fatal multisystem inflammatory syndrome in adult after SARS-CoV-2 natural infection and COVID-19 vaccination. *Emerg Infect Dis*. (2021) 27:2914. doi: 10.3201/eid2711.211612

55. Miao G, Chen Z, Cao H, Wu W, Chu X, Liu H, et al. From immunogen to COVID-19 vaccines: prospects for the post-pandemic era. *Biomed Pharmacother*. (2023) 2023:114208. doi: 10.1016/j.biopha.2022.114208

56. Castanares-Zapatero D, Chalon P, Kohn L, Dauvrin M, Detollenaere J, Maertens de Noordhout C, et al. Pathophysiology and mechanism of long COVID: a comprehensive review. *Ann Med*. (2022) 54:1473–87. doi: 10.1080/07853890.2022.2076901

57. Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV, Group WCCDW, et al. A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect Dis*. (2021) 22:e102–7. doi: 10.1016/S1473-3099(21)00703-9

58. Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, et al. Attributes and predictors of long COVID. *Nat Med*. (2021) 27:626–31. doi: 10.1038/s41591-021-01292-y

59. Boix V, Merino E. Post-COVID syndrome. The never ending challenge. *Med Clin*. (2022) 158:178. doi: 10.1016/j.medcle.2021.10.005

60. Yong SJ, Liu S. Proposed subtypes of post-COVID-19 syndrome (or long-COVID) and their respective potential therapies. *Rev Med Virol*. (2022) 32:e2315. doi: 10.1002/rmv.2315

61. Fernández-de Las-Pe nas C, Notarte KI, Peligro PJ, Velasco JV, Ocampo MJ, Henry BM, et al. Long-COVID symptoms in individuals infected with different SARS-CoV-2 variants of concern: a systematic review of the literature. *Viruses*. (2022) 14:2629. doi: 10.3390/v14122629

62. Stavem K, Ghanima W, Olsen MK, Gilboe HM, Einvik G. Persistent symptoms 1.5-6 months after COVID-19 in non-hospitalised subjects: a population-based cohort study. *Thorax*. (2021) 76:405–7. doi: 10.1136/thoraxjnl-2020-216377

63. Su Y, Yuan D, Chen DG, Ng RH, Wang K, Choi J, et al. Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell*. (2022) 185:881–95. doi: 10.1016/j.cell.2022.01.014

64. Carfi A, Bernabei R, Landi F. Against COVID-19. Post-Acute Care Study Group: for the Gemelli Against CCOVID-19 Post-Acute Care Study Group. Persistent symptoms in patients after acute COVID-19. *JAMA*. (2020) 9:603. doi: 10.1001/jama.2020.12603

65. Dennis A, Wamil M, Alberts J, Oben J, Cuthbertson DJ, Wootton D, et al. Multiorgan impairment in low-risk individuals with post-COVID-19 syndrome: a prospective, community-based study. *BMJ Open*. (2021) 11:e048391. doi: 10.1136/bmjopen-2020-048391

66. Tanne JH. Covid-19: even mild infections can cause long term heart problems, large study finds. *Br Med J*. (2022) 2022:378. doi: 10.1136/bmj.o378

67. Qin W, Chen S, Zhang Y, Dong F, Zhang Z, Hu B, et al. Diffusion capacity abnormalities for carbon monoxide in patients with COVID-19 at 3-month follow-up. *Eur Respir J*. (2021) 58:2003677. doi: 10.1183/13993003.03677-2020

68. Sun B, Tang N, Peluso MJ, Iyer NS, Torres L, Donatelli JL, et al. Characterization and biomarker analyses of post-COVID-19 complications and neurological manifestations. *Cells*. (2021) 10:386. doi: 10.3390/cells10020386

69. Lamers MM, Beumer J, Van Der Vaart J, Knoops K, Puschhof J, Breugem TI, et al. SARS-CoV-2 productively infects human gut enterocytes. *Science*. (2020) 369:50–4. doi: 10.1126/science.abc1669

70. Bhattacharyya A, Seth A, Rai S. The effects of long COVID-19, its severity, and the need for immediate attention: analysis of clinical trials and Twitter data. *medRxiv*. (2022) 2022–09. doi: 10.1101/2022.09.13.22279833

71. Haendel MA, Chute CG, Bennett TD, Eichmann DA, Guinney J, Kibbe WA, et al. The National COVID Cohort Collaborative (N3C): rationale, design, infrastructure, and deployment. *J Am Med Inform Assoc*. (2021) 28:427–43. doi: 10.1093/jamia/ocaa196

72. Alyasseri ZAA, Al-Betar MA, Doush IA, Awadallah MA, Abasi AK, Makhadmeh SN, et al. Review on COVID-19 diagnosis models based on machine learning and deep learning approaches. *Expert Syst*. (2022) 39:e12759. doi: 10.1111/exsy.12759

73. de Fátima Cobre A, Surek M, Stremel DP, Fachi MM, Borba HHL, Tonin FS, et al. Diagnosis and prognosis of COVID-19 employing analysis of patients' plasma and serum via LC-MS and machine learning. *Comput Biol Med*. (2022) 146:105659. doi: 10.1016/j.compbiomed.2022.105659

74. 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 Unlocked*. (2021) 25:100691. doi: 10.1016/j.imu.2021.100691

75. Chandra R, Jain A, Singh Chauhan D. Deep learning via LSTM models for COVID-19 infection forecasting in India. *PLoS ONE*. (2022) 17:e0262708. doi: 10.1371/journal.pone.0262708

76. Lalmuanawma S, Hussain J, Chhakchhuak L. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: a review. *Chaos Solitons Fractals*. (2020) 139:110059. doi: 10.1016/j.chaos.2020.110059

77. Zoabi Y, Deri-Rozov S, Shomron N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *npj Digit Med*. (2021) 4:1–5. doi: 10.1038/s41746-020-00372-6

78. Zhao L, Lediju Bell MA. A review of deep learning applications in lung ultrasound imaging of COVID-19 patients. *BME Front*. (2022) 2022:9780173. doi: 10.34133/2022/9780173

79. Dang T, Han J, Xia T, Spathis D, Bondareva E, Siegele-Brown C, et al. Exploring longitudinal cough, breath, and voice data for COVID-19 progression prediction via sequential deep learning: model development and validation. *J Med Internet Res*. (2022) 24:e37004. doi: 10.2196/37004

80. Jin W, Dong S, Dong C, Ye X. Hybrid ensemble model for differential diagnosis between COVID-19 and common viral pneumonia by chest X-ray radiograph. *Comput Biol Med*. (2021) 131:104252. doi: 10.1016/j.compbiomed.2021.104252

81. Soares E, Angelov P, Biaso S, Froes MH, Abe DK. SARS-CoV-2 CT-scan dataset: a large dataset of real patients CT scans for SARS-CoV-2 identification. *MedRxiv*. (2020). doi: 10.1101/2020.04.24.20078584

82. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. *TensorFlow*: a system for *Large-Scale* machine learning. In: *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*. (2016). p. 265–83. Available online at: https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf

83. Alafif T, Tehame AM, Bajaba S, Barnawi A, Zia S. Machine and deep learning towards COVID-19 diagnosis and treatment: survey, challenges, and future directions. *Int J Environ Res Public Health*. (2021) 18:1117. doi: 10.3390/ijerph18031117

84. Lv H, Shi L, Berkenpas JW, Dao FY, Zulfiqar H, Ding H, et al. Application of artificial intelligence and machine learning for COVID-19 drug discovery and vaccine design. *Brief Bioinform*. (2021) 22:bbab320. doi: 10.1093/bib/bbab320

85. Law JN, Akers K, Tasnina N, Della Santina CM, Kshirsagar M, Klein-Seetharaman J, et al. Identifying human interactors of SARS-CoV-2 proteins and drug targets for COVID-19 using network-based label propagation. *arXiv preprint* arXiv:200601968. (2020).

86. Ray S, Lall S, Bandyopadhyay S. A deep integrated framework for predicting SARS-CoV2-human protein-protein interaction. *IEEE Trans Emerg Top Comput Intell*. (2022) 6:1463–72. doi: 10.1109/TETCI.2022.3182354

87. Du BX, Qin Y, Jiang YF, Xu Y, Yiu SM, Yu H, et al. Compound-protein interaction prediction by deep learning: databases, descriptors and models. *Drug Discov Today*. (2022) 27:1350–66. doi: 10.1016/j.drudis.2022.02.023

88. Mylonas SK, Axenopoulos A, Daras P. DeepSurf: a surface-based deep learning approach for the prediction of ligand binding sites on proteins. *Bioinformatics*. (2021) 37:1681–90. doi: 10.1093/bioinformatics/btab009

89. Liu Q, Wang PS, Zhu C, Gaines BB, Zhu T, Bi J, et al. OctSurf: efficient hierarchical voxel-based molecular surface representation for protein-ligand affinity prediction. *J Mol Graphics Model*. (2021) 105:107865. doi: 10.1016/j.jmgm.2021.107865

90. Jones D, Kim H, Zhang X, Zemla A, Stevenson G, Bennett WD, et al. Improved protein-ligand binding affinity prediction with structure-based deep fusion inference. *J Chem Inf Model*. (2021) 61:1583–92. doi: 10.1021/acs.jcim.0c01306

91. Feng Z, Chen M, Liang T, Shen M, Chen H, Xie XQ. Virus-CKB: an integrated bioinformatics platform and analysis resource for COVID-19 research. *Brief Bioinform*. (2021) 22:882–95. doi: 10.1093/bib/bbaa155

92. Zhavoronkov A, Aladinskiy V, Zhebrak A, Zagribelnyy B, Terentiev V, Bezrukov D, et al. Potential COVID-2019 3C-like protease inhibitors designed using generative deep learning approaches. *ChemRxiv. Preprint*. (2020) 11:102. doi: 10.26434/chemrxiv.11829102

93. Tang B, He F, Liu D, He F, Wu T, Fang M, et al. AI-aided design of novel targeted covalent inhibitors against SARS-CoV-2. *Biomolecules*. (2022) 12:746. doi: 10.3390/biom12060746

94. Ton AT, Gentile F, Hsing M, Ban F, Cherkasov A. Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds. *Mol Inform*. (2020) 39:2000028. doi: 10.1002/minf.202000028

95. Srinivasan S, Batra R, Chan H, Kamath G, Cherukara MJ, Sankaranarayanan SK. Artificial intelligence-guided De novo molecular design targeting COVID-19. *ACS Omega*. (2021) 6:12557–66. doi: 10.1021/acsomega.1c00477

96. Wang Z, Liu M, Luo Y, Xu Z, Xie Y, Wang L, et al. Advanced graph and sequence neural networks for molecular property prediction and drug discovery. *Bioinformatics*. (2022) 38:2579–86. doi: 10.1093/bioinformatics/btac112

97. Medina-Ortiz D, Contreras S, Amado-Hinojosa J, Torres-Almonacid J, Asenjo JA, Navarrete M, et al. Generalized property-based encoders and digital signal processing facilitate predictive tasks in protein engineering. *Front Mol Biosci*. (2022) 9:898627. doi: 10.3389/fmolb.2022.898627

98. Ferruz N, Höcker B. Controllable protein design with language models. *Nat Mach Intell*. (2022) 4:521–32. doi: 10.1038/s42256-022-00499-z

99. Wigh DS, Goodman JM, Lapkin AA. A review of molecular representation in the age of machine learning. *Wiley Interdisc Rev Comput Mol Sci*. (2022) 12:e1603. doi: 10.1002/wcms.1603

100. Verkuil R, Kabeli O, Du Y, Wicky BI, Milles LF, Dauparas J, et al. Language models generalize beyond natural proteins. *bioRxiv*. (2022) doi: 10.1101/2022.12.21.521521

101. Minh D, Wang HX, Li YF, Nguyen TN. Explainable artificial intelligence: a comprehensive review. *Artif Intell Rev*. (2022) 55:3503–68. doi: 10.1007/s10462-021-10088-y

102. Chen CC, Chang CK, Chiu CC, Yang TY, Hao WR, Lin CH, et al. Machine learning analyses revealed distinct arterial pulse variability according to side effects of Pfizer-BioNTech COVID-19 vaccine (BNT162b2). *J Clin Med*. (2022) 11:6119. doi: 10.3390/jcm11206119

103. Abbaspour S, Robbins GK, Blumenthal KG, Hashimoto D, Hopcia K, Mukerji SS, et al. Identifying modifiable predictors of COVID-19 vaccine side effects: a machine learning approach. *Vaccines*. (2022) 10:1747. doi: 10.3390/vaccines10101747

104. Flora J, Khan W, Jin J, Jin D, Hussain A, Dajani K, et al. Usefulness of vaccine adverse event reporting system for machine-learning based vaccine research: A Case study for COVID-19 vaccines. *Int J Mol Sci*. (2022) 23:8235. doi: 10.3390/ijms23158235

105. Gupta M, Jain R, Taneja S, Chaudhary G, Khari M, Verdú E. Real-time measurement of the uncertain epidemiological appearances of COVID-19 infections. *Appl Soft Comput*. (2021) 101:107039. doi: 10.1016/j.asoc.2020.107039

106. Mansour RF, Escorcia-Gutierrez J, Gamarra M, Gupta D, Castillo O, Kumar S. Unsupervised deep learning based variational autoencoder model for COVID-19 diagnosis and classification. *Pattern Recogn Lett*. (2021) 151:267–74. doi: 10.1016/j.patrec.2021.08.018

107. Wang D, Zuo F, Gao J, He Y, Bian Z, Bernardes SD, et al. Agent-based simulation model and deep learning techniques to evaluate and predict transportation trends around COVID-19. *arXiv preprint* arXiv:201009648. (2020). doi: 10.48550/arXiv.2010.09648

108. Kompella V, Capobianco R, Jong S, Browne J, Fox S, Meyers L, et al. Reinforcement learning for optimization of COVID-19 mitigation policies. *arXiv preprint* arXiv:201010560. (2020). doi: 10.48550/arXiv.2010.10560

109. Medina-Ortiz D, Contreras S, Barrera-Saavedra Y, Cabas-Mora G, Olivera-Nappa Á. Country-wise forecast model for the effective reproduction number R t of coronavirus disease. *Front Phys*. (2020) 8:304. doi: 10.3389/fphy.2020.00304

110. Contreras S, Villavicencio HA, Medina-Ortiz D, Saavedra CP, Olivera-Nappa Á. Real-time estimation of R t for supporting public-health policies against COVID-19. *Front Public Health*. (2020). 8:556689. doi: 10.3389/fpubh.2020.556689

111. Polyzos S, Samitas A, Spyridou AE. Tourism demand and the COVID-19 pandemic: an LSTM approach. *Tour Recreat Res*. (2021) 46:175–87. doi: 10.1080/02508281.2020.1777053

112. Hills S, Eraso Y. Factors associated with non-adherence to social distancing rules during the COVID-19 pandemic: a logistic regression analysis. *BMC Public Health*. (2021) 21:1–25. doi: 10.1186/s12889-021-10379-7

113. Saba AI, Elsheikh AH. Forecasting the prevalence of COVID-19 outbreak in Egypt using nonlinear autoregressive artificial neural networks. *Process Safety Environ Protect*. (2020) 141:1–8. doi: 10.1016/j.psep.2020.05.029

114. Singh RK, Rani M, Bhagavathula AS, Sah R, Rodriguez-Morales AJ, Kalita H, et al. Prediction of the COVID-19 pandemic for the top 15 affected countries: Advanced autoregressive integrated moving average (ARIMA) model. *JMIR Public Health Surveill*. (2020) 6:e19115. doi: 10.2196/19115

115. Ketu S, Mishra PK. Enhanced Gaussian process regression-based forecasting model for COVID-19 outbreak and significance of IoT for its detection. *Appl Intell.* (2021) 51:1492–512. doi: 10.1007/s10489-020-01889-9

116. Castillo Ossa LF, Chamoso P, Arango-López J, Pinto-Santos F, Isaza GA, Santa-Cruz-González C, et al. A hybrid model for COVID-19 monitoring and prediction. *Electronics.* (2021) 10:799. doi: 10.3390/electronics10070799

117. Cheng FY, 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. doi: 10.3390/jcm9061668

118. Alamrouni A, Aslanova F, Mati S, Maccido HS, Jibril AA, Usman A, et al. Multi-regional modeling of cumulative COVID-19 cases integrated with environmental forest knowledge estimation: a deep learning ensemble approach. *Int J Environ Res Public Health.* (2022) 19:738. doi: 10.3390/ijerph19020738

119. Pfaff ER, Girvin AT, Bennett TD, Bhatia A, Brooks IM, Deer RR, et al. Identifying who has long COVID in the USA: a machine learning approach using N3C data. *Lancet Digit Health.* (2022) 4:e532–41. doi: 10.1016/S2589-7500(22)00048-6

120. Binka M, Klaver B, Cua G, Wong AW, Fibke C, Velásquez García HA, et al. An elastic net regression model for identifying long COVID patients using health administrative data: a population-based study. In: *Open Forum Infectious Diseases. vol. 9.* Oxford: Oxford University Press US (2022). p. ofac640.

121. Fritsche LG, Jin W, Admon AJ, Mukherjee B. Characterizing and predicting post-acute sequelae of SARS CoV-2 infection (PASC) in a large academic medical center in the US. *medRxiv.* (2022) doi: 10.1101/2022.10.21.22281356

122. Sengupta S, Loomba J, Sharma S, Brown DE, Thorpe L, Haendel MA, et al. Analyzing historical diagnosis code data from NIH N3C and RECOVER Programs using deep learning to determine risk factors for Long COVID. *arXiv preprint* arXiv:221002490. (2022) doi: 10.1109/BIBM55620.2022.9994851

123. Gupta A, Jain V, Singh A. Stacking ensemble-based intelligent machine learning model for predicting post-COVID-19 complications. *N Generat Comput.* (2022) 40:987–1007. doi: 10.1007/s00354-021-00144-0