




Prediction of global spread of COVID-19 pandemic: a review and research challenges

Saloni Shah¹ · Aos Mulahuwaish¹ · Kayhan Zrar Ghafoor^{2,3}  · Halgurd S. Maghdid⁴

Accepted: 13 March 2021 / Published online: 16 July 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract

Since the initial reports of the Coronavirus surfacing in Wuhan, China, the novel virus currently without a cure has spread like wildfire across the globe, the virus spread exponentially across all inhabited continent, catching local governments by surprise in many cases and bringing the world economy to a standstill. As local authorities work on a response to deal with the virus, the scientific community has stepped in to help analyze and predict the pattern and conditions that would influence the spread of this unforgiving virus. Using existing statistical modeling tools to the latest artificial intelligence technology, the scientific community has used public and privately available data to help with predictions. A lot of this data research has enabled local authorities to plan their response—whether that is to deploy tightly available medical resources like ventilators or how and when to enforce policies to social distance, including lockdowns. On the one hand, this paper shows what accuracy of research brings to enable fighting this disease; while on the other hand, it also shows what lack of response from local authorities can do in spreading this virus. This is our attempt to compile different research methods and comparing their accuracy in predicting the spread of COVID-19.

Keywords COVID-19 · Machine learning · Deep learning · Prediction methods

✉ Kayhan Zrar Ghafoor
kayhan.ghafoor@knu.edu.iq

Saloni Shah
skshah@svsu.edu

Aos Mulahuwaish
amulahuw@svsu.edu

Halgurd S. Maghdid
halgurd.maghdid@koyauniversity.org

¹ Department of Computer Science and Information Systems, Saginaw Valley State University, 7400 Bay Rd, University Center, MI, USA

² Department of Computer Science, Knowledge University, University Park, Kirkuk Road, Erbil 446015, Iraq

³ Department of Software Engineering, Salahaddin University, Erbil, Iraq

⁴ Department of Software Engineering, Faculty of Engineering, Koya University, Koysinjaq, Kurdistan Region-F.R., Iraq

1 Introduction

The global spread of the Coronavirus disease has become a healthcare concern, but the rapid mutation of the strain and high infectious rate of the disease has made it a socio-economic issue for countries worldwide. The World Health Organization (WHO) was alerted by Chinese officials about dozens of pneumonia-like diseases in Wuhan's city as a new celebration was taking place in the country. The centers for disease control and prevention identified a sea market in Wuhan suspected to be a center of the outbreak. By March of 2020, the disease had spread to major countries across the globe. WHO officially announced the COVID-19 disease outbreak as a pandemic on 11 March 2020. Figure 1 shows the rapid increases in cases globally (Johns Hopkins University and Medicine 2020).

The infectious nature of the disease and lack of vaccinations created a restriction on social interactions and economic collapse, overwhelming economic and healthcare systems everywhere. This created an urgent need to study the virus to curb the spread, find a cure, and help local authorities all over the world decide on measures to prevent the spread of the virus. The need is more pronounced to help countries decide how best to open back their economies and manage healthcare logistics.

It is important to predict—with accuracy and specificity—the spread of COVID-19. Using existing data, there is a need to forecast trends on the spread of the virus. Many governments rely heavily on such predictions to plan their next actions, whether it is to allocate medical resources or to ease or increase the level of lockdowns.

Considering that this virus has made it really hard to focus on a single mode of transmission, it has become fundamentally important for scientific communities to focus on various factors that can affect the spread of the disease. One such field of study that could contribute to the cause and research for the spread of COVID-19 is social media or social connectivity (Kuchler et al. 2020) and (Guardian 2020). The social connectedness index discussed in Kuchler et al. (2020) can significantly impact identifying the links between mass populations from one geographical area to the other. This connectedness identified by

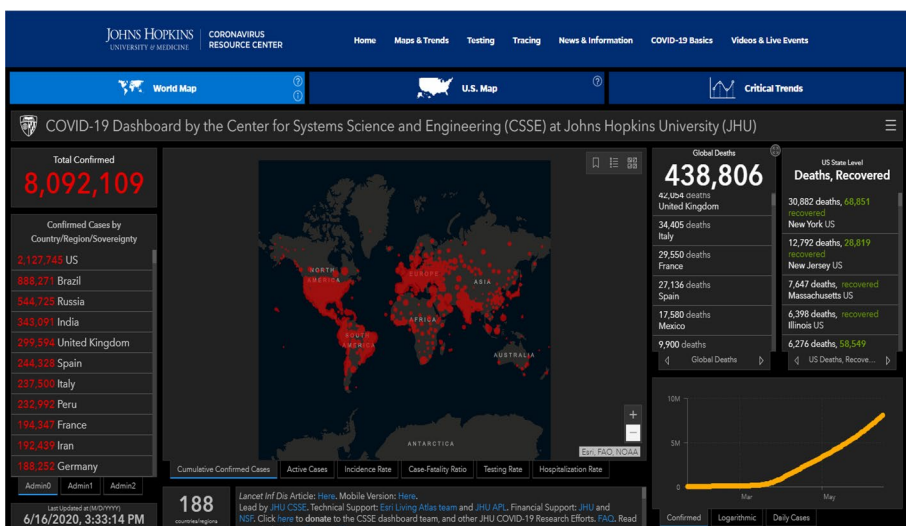


Fig. 1 Snapshot of the Johns Hopkins University, Coronavirus Resource Centre (June 16, 2020)

social media can open possibilities to correlate the spread of the disease that depends on different demographics or geographies.

The remainder of this paper is structured as follows: Sect. 2 discusses the related work and the contribution of the paper. Section 3 summarizes different research studies identified to understand the spread of COVID-19, and a tabular form explains the key features of each paper referred to. Section 4 presents AI and technologies that are contributed to combating the COVID-19 spreading. Section 5, we highlighted all lessons learned in this area of study. Section 6 discusses the open research challenges. Finally, Sect. 7 concludes the paper.

2 Related work and contributions

In the existing state-of-the-arts, there are many surveys and tutorials that are tailored to the prediction of COVID-19 spreading. These publications vary from more general works to particular technical reports and explications in the area. For instance, Wynants et al. (2020) have tried to review and appraise the validity and usefulness of published prediction models for diagnosing coronavirus disease, the prognosis for infected population, and detecting increased risk criteria in the general population of becoming infected. This kind of effort helps future researchers to focus on an area not explored or an area with promising results to establish a sustainable forecast model.

In essence, in this paper, we have tried to put together an ensemble of studies that take into account different variables for the spread of the disease. Table 1 illustrates published papers in the domain of prediction models for COVID-19 spreading. We summarize the main contributions of this survey as follows:

- This paper reviews recent literature on the prediction of COVID-19 spreading, what kind of dataset was used for accurate prediction, and then highlighting the research gaps. This study's findings present that several parameters will contribute to the spreading of the virus, such as social interactions measured by social networks.
- We have identified crucial applications of AI, and different technologies are being used to combat COVID-19.
- We have identified research challenges, directions on prediction methods, and AI-enabled technologies to combat the COVID-19 spreading.

This survey can be used as a useful resource for future research directions in prediction models for COVID-19 spreading. Our intent is to help the scientists and researchers to get a clear idea of research already done and possibly guide them on how to proceed with their own studies.

3 Prediction models

Prediction models can be categorized based on machine learning models and mathematical models, as depicted in Fig. 2. We discuss and detail this categorization here.

In the existing literature, the authors have used different statistical and mathematical calculations, data models, and AI [as illustrated in Fig. 3 (Pham et al. 2020)] to try

Table 1 Comparison of existing prediction models for COVID-19 spreading

Authors	Objective	Method/Model used	Dataset used	Output and accuracy	Weakness
Li Yan et al. (2020)	Focused on using biomarkers, obtained via blood samples, to be able to predict severe COVID-19 cases that result in higher risk of mortality	Supervised XGBoost classifier machine learning-based model (decision-tree-based)	Blood samples from 485 infected patients in the region of Wuhan, China (Jan 10–Feb 18, 2020)	The model can predict the mortality rate for patients more than 10 days in advance with more than 90% accuracy	Since the method is dependent on data, the model will vary when using different datasets. Single-centered, retrospective study lacking large-sample, multi-centered study
Elmousalami and Hassani (2003)	Use time series model to analyze and predict spread of COVID-19. Using existing datasets from renowned sources as John Hopkins university	Time series models (moving average (MA), weighted moving average (WMA), and single exponential smoothing (SES)) and mathematical formulations	WHO, the national health commission of China and Johns Hopkins University developed open database for the COVID-19 cases	Day-level forecasting models on COVID-19 using time series models and mathematical forecasting	Depends upon data available. Forecasting may miss underreporting of data
Tomar and Gupta (2020)	Using data-driven estimation models, predict rate of infection of COVID-19 in India 30 days ahead. Also predict impact of preventive measures like social distancing on the infection rate	LSTM based technique used with the MATLAB environment	Indian Govt. COVID-19 Dashboard database (April 30–Jan 4, 2020) https://www.mygov.in/covid-19?cbps=1	Number of recovery days, effect of transmission rate on the number of cases, effect of transmission rate with social distancing observed	Models are based on limited data availability impacting the accuracy
Zhao et al. (2020)	Use the aforementioned models to analyze spread of COVID-19 in regions, depending up how local authorities intervene and what policies they adopt to curb the spread of this pandemic	Maximum-Hasting (MH) parameter estimation method and the modified Susceptible Exposed Infected Recovered (SEIR) model	Data released by the Johns Hopkins University	Classify six studied African nations into three categories: suppression, mitigation, or mildness. Pretty accurate categorization of nations	Assumes intervention intensity of studied nations at a fraction of comparison model (China). Model may not be able to predict rate of growth, in case suggested interventions are not carried out (in time predictions)

Table 1 (continued)

Authors	Objective	Method/Model used	Dataset used	Output and accuracy	Weakness
Yang et al. (2020)	To predict the probability of epidemic, its peak and more importantly what would be impact of intervention measures in China. Also attempt to predict the impact of delaying intervention leading into second outbreak / peak	Susceptible-Exposed-Infectious-Removed (SEIR) and Long Short-Term Memory (LSTM) models	Integrated population migration data before and after January 23 (inbound and outbound events by rail, air and road traffic, were sourced from a web-based program) and most updated COVID-19 epidemiological data (National Health Commission of China)	The models used, predict the trend for spread of COVID-19 with reasonable confidence in mainland China and also show promise for future prediction of the epidemic	The accuracy of the models will depend on the implementations of control measures
Zou et al. (2020)	Propose a new model that takes into account untested or unreported cases while predicting rate of cases (active or deaths) of COVID-19 infection	UCLA-SuEIR (Susceptible, unreported, Exposed, Infectious and Recovered)	New York Times COVID-19-data and Johns Hopkins University Center for Systems Science and Engineering data	Provide projections of the number of infections and deaths, and predict peak dates of active cases	The biggest challenge to substantiate the findings of this new model will be data, as it is not reported
Hamzaha, et al. (2020)	Predictive analysis using the SEIR model and Sentiment analysis of verified news into positive and negative news	Susceptible-Exposed-Infectious-Removed (SEIR) and Bidirectional Encoder Representations from Transformers (BERT)	John Hopkins, WHO, local Chinese website—DingXiangYuan	Reflected data on a website, using standard technique. Good from a visual standpoint	No new approach and sentiment analysis may not produce accurate results all the time

Table 1 (continued)

Authors	Objective	Method/Model used	Dataset used	Output and accuracy	Weakness
Fanelli and Piazza (2020)	A prediction model for maximum number of infected individuals along with timing of the peak. Using simple quantitative models, show how containment efforts can help in reducing the spread	Mean-field approximation in modified Susceptible-Infectious-Recovered-Deceased (SIRD) model	GitHub repository associated with the interactive dashboard hosted by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, Baltimore, USA	Provide estimates for the time and magnitude of the epidemic peak	The key drawback of the model is that it assumes standard conditions and fails to track with rapid recovery (decrease in number of infected cases) and overestimates the number of deaths when extreme measures like social distancing are used
Sajadi et al. (Mohammad 2020)	To study climate data with the intent to establish correlation in regions that have similar climate setup and come up with a model to predict possible new locations based on similarity of climate with the current COVID-19 hotspots	ERA-5 reanalysis of the data—then compared to areas that are either not affected, or do not have significant community spread—Eventual statistical analysis done with produced maps using Graph Pad Prism	Examined climate data from cities (globally) with significant community spread of COVID-19	The analysis shows a statistically significant association between temperature and specific humidity for areas that are significantly and less significantly affected by the pandemic	Lack of human factors like intervention, climatic factors like cloud cover, and viral factor like mutation of the virus which can lead to unpredictability of the model

Table 1 (continued)

Authors	Objective	Method/Model used	Dataset used	Output and accuracy	Weakness
Mollalo et al. (2020)	Using spatial models coupled with multiple socio-economic data factors to try explaining variation of the spread of COVID-19 in the United States (USA) based on geographic modeling	5 Models based on spatial analysis technique (Global—OLS, SLM, SEM and Local—GWR, MGWR)	County-level counts of COVID-19 cases retrieved from USA Facts (Jan 22–April 9, 2020). Crude incidence rates were computed for the counties and joined to the administrative boundary shapefile of counties obtained from the TIGER/Line database	GIS models showing the spread of COVID-19	Model couldn't include county level data, impacting accuracy of predictions. Model also cannot show the impact of lockdown procedures on spread
Kuchler et al. (2020)	Using social connectedness index, establish correlation and predict the spread of COVID-19 between socially connected people in such areas	Analysis of Connectedness Index (SCI) introduced by Bailey et al. (2018)	Aggregated (anonymized) data from Facebook	Heatmap of socially connected folks, in selected hotspots and prediction on spread of virus based on social network	Model may use inaccurate or incomplete data since people do tend to input incorrect information or may have opted out thus reducing the accuracy of the aforementioned index

Table 1 (continued)

Authors	Objective	Method/Model used	Dataset used	Output and accuracy	Weakness
Maghdid et al. (2004)	A smartphone-based contact tracing approach. It focuses on notifying people who may have been in contact of a positive COVID-19 person. It helps local authorities on lockdown management (area and duration to lockdown)	Unsupervised Machine Learning (UML)—K means clustering algorithm	Customized dataset part of the developed solution storing following information: name, zip code, age, phone number, MAC address of smartphone, gender, COVID-19 status	Smartphone app dashboard with individual alerts and GIS enabled map of possible COVID hotspots. A web portal system dashboard tracing possible impacted user	Accuracy of the tool depends upon registration of users. The more users register for the tool, the higher the accuracy. Research uses Android based application. Not sure if one is developed for iOS given the popularity of Apple smartphones. Privacy concerns given the type of information the app stores
Kufel et al. (2020)	Auto-regressive Integrated Moving Average (ARIMA) model for predicting the dynamics of COVID-19 incidence at different stages of the epidemic	Auto-regressive Integrated Moving Average	Johns Hopkins University Center for Systems Science and Engineering database	Forecast growth during 6 selected sub-periods, of epidemic in 32 European countries	Model is most probably beneficial for short-term forecasts. Didn't factor or address the role of non-pharma interventions and population testing policies
Tuli et al. (2020a)	Compare data forecast between the Generalized Inverse Weibull and Gaussian Distribution	Generalized Inverse Weibull (GIW) distribution	WHO database	An iteratively weighted curve fitting model using the GIW distribution (called "Robust Weibull"). Was able to improve the accuracy of predictions over the Gaussian model	Didn't factor in population density, age, intervention methods by government in the regression model

Table 1 (continued)

Authors	Objective	Method/Model used	Dataset used	Output and accuracy	Weakness
Tuli et al. (2020b)	Developing and using to forecast a LSTM model based on Generalized Inverse Weibull (GIW) distribution	LSTM-based Robust Weibull approach (W-LSTM)	Epidemic data from European Centre for Disease Prevention and Control (ECDC). Socio-economic data from IndexMundi and World Bank. Virus data from Biorxiv	Forecast model integrating LSTM and Generalized Inverse Weibull (GIW) distribution. Demonstrated better results over ARIMA and other traditional ML based models	No obvious weaknesses suggested by the authors
Zheng et al. (2020)	To forecast the Inflection point of new cases in select countries using existing data	State transition matrix (STM) modeling	National Health Commission of China	A scenario-based forecast model to forecast different inflection points during the spread of the virus globally	Accuracy of forecast was limited due to not enough data
Shahid et al. (2020)	To compare and assess best predictive model amongst autoregressive integrated moving average (ARIMA), support vector regression (SVR), long short term memory (LSTM), Gated recurrent units (GRU), bidirectional long short term memory (Bi-LSTM)	ARIMA, SVR, LSTM, Bi-LSTM, GRU	Harvard dataset taken from the link: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/L201OT	Bi-LSTM showed the best results using performance measures like MAE, RMSE, R2_score	No obvious weaknesses suggested by the authors

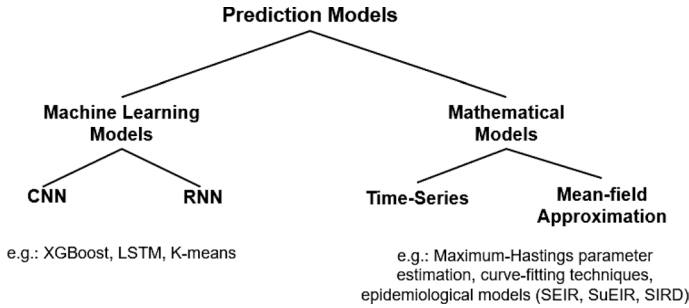


Fig. 2 Categorization of prediction models based on machine learning and mathematical models

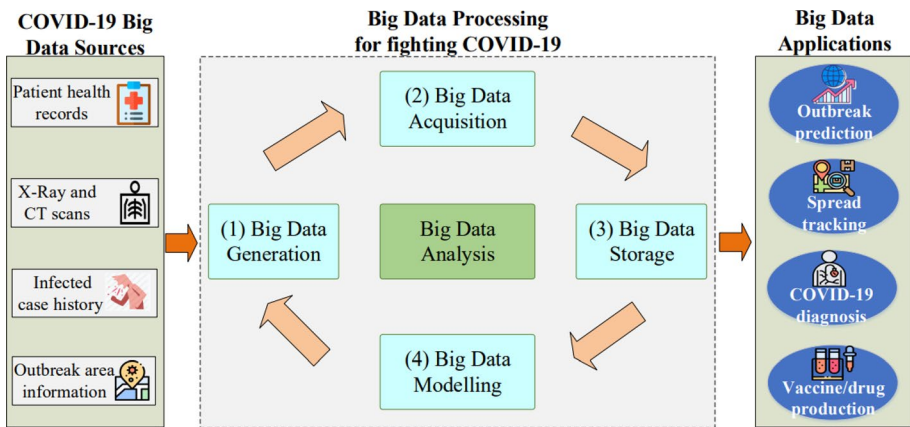


Fig. 3 Big data and its applications for fighting COVID-19 pandemic

and help and expedite decision-making and logistical planning in healthcare systems. Whether it is using existing clinical data, biomarkers, traditional medical knowledge, or weather patterns, they have tried to come up with a way to help with proper arrangement and utilization of available, and in many cases, limited healthcare resources around the world. In some cases, they have come up with prediction models that can help clinicians get an early warning for patient care by a few days to help reduce mortality in COVID-19 patients. In some other cases, researchers have used AI tools coupled with datasets covering social behavior or responsiveness of local bodies to the COVID-19 outbreak to help predict the rate of infections. Below, Fig. 3 (Pham et al. 2020) shows the use of Big Data in different applications for fighting COVID-19.

Li Yan et al. (2020) focused on using biomarkers obtained via blood samples to be able to predict severe COVID-19 cases that result in a higher risk of mortality. They selected the following three biomarkers: lactic dehydrogenase (LDH), lymphocyte, and high-sensitivity C-reactive protein (hs-CRP) to train machine learning tools in forecasting potential mortality in patients with high accuracy. Furthermore, their trained models were able to predict worsening cases well in advance, up to 10 days in some cases, giving medical professionals a fighting chance to change a patient’s treatment. This

research showcases the use of existing medical knowledge of biomarkers coupled with AI tools to help forecast severe cases in advance.

Elmousalami and Hassanien (2003) use the time series model to analyze and predict the spread of COVID-19. Using existing datasets from renowned sources as Johns Hopkins University, they were able to forecast and, in some cases, validate assumptions related to the spread of this virus. Key takeaways from their research are:

- In absence of strict lockdown and social distancing policies, rate of spread is exponential.
- They are able to validate the hypothesis of person-to-person transmission as driver of exponential spread.
- Compounding of COVID-19 cases is more than 25% in absence of social distancing practices.
- Lastly, the most important inference is that exponential growth is more because of virus transmission rather than increased testing rates. This is critical analysis as many countries have been complacent with enforcing strict guidelines due to the belief that more tests performed will result in higher number of positive COVID-19 patient identification.

Tomar and Gupta (2020) used Long Short-Term Memory (LSTM)—a type of Recurrent Neural Networks (RNN) method to predict a number of COVID-19 cases in India. Their analysis showed infection rate in case preventive measures like social distancing and lockdown were practiced versus the rate if no such measures were in place. Their trained model [shown in Fig. 4 (Tomar and Neeraj Gupta 2020)] was also able to predict positive and recovered cases within a certain accuracy range. They deployed a curve fitting technique to assess the accuracy of their prediction model come up with close results. This tool has the

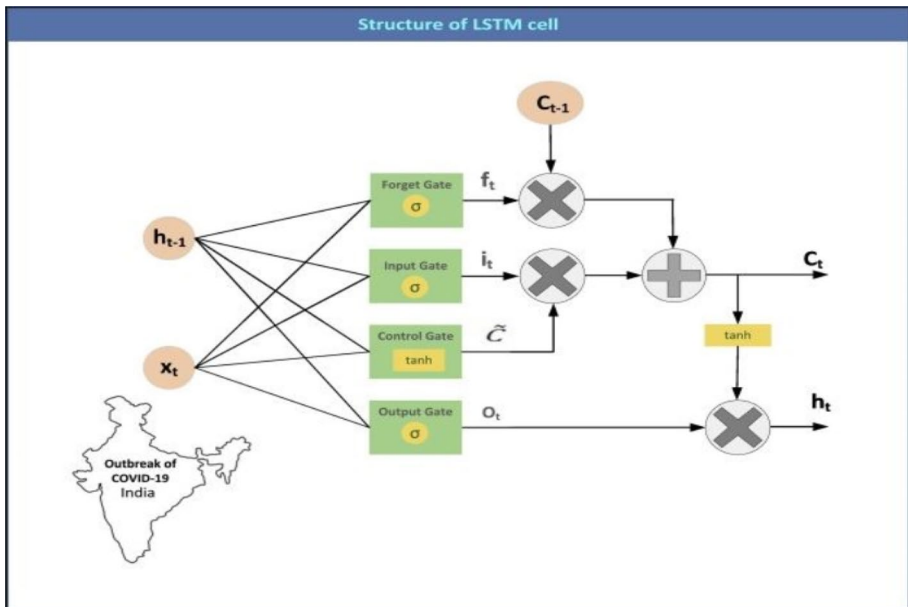


Fig. 4 Basic structure of LSTM

potential to help local authorities with forecasting infection and recovery rates, giving them needed data to prepare for outbreaks and plan the deployment of limited medical resources.

Zhao et al. (2020) use the Maximum-Hasting (MH) parameter estimation method and the modified Susceptible Exposed Infectious Recovered (SEIR) model to analyze the spread of COVID-19 in regions depending on how local authorities intervene and what policies they adopt to curb the spread of this pandemic. The authors study three possible scenarios to deploy: suppression, mitigation, or mildness in six African countries. In case of suppression, local authorities maintain tight control and deploy all possible policies to curb the spread of the virus. South Africa and Senegal fall under this scenario and seem to be tracking with a controlled infection rate. In the second intervention method, the response is focused on mitigation of spread rather than curb it. This policy results, as shown by the models, in a control time that lags behind suppression policy by at least 10 days. African nations of Algeria, Nigeria, and Kenya show an infection curve aligned with this intervention policy. The last intervention scenario lacks proper mitigation and will result in a doubling of infection rates at a rapid pace. Egypt seems to be on track for an infection rate as predicted by this policy.

Another study by (Yang et al. 2020), Yang et al. use Susceptible-Exposed-Infectious-Removed (SEIR) and LSTM models to predict infection rates in China. The SEIR model predicts the probability of epidemic, its peak, and, more importantly, what would be the impact of intervention measures. With a certain success rate, they attempt to predict the impact of delaying intervention leading to a second outbreak/peak. They work on the Long Short-Term Memory (LSTM) model to predict a new number of infections. Key areas of improvement in the two models were picking up parameters like correct incubation period, diagnostics capacity impacting total infected numbers, and seasonal influences like temperatures. Yet, the LSTM method showed quite a similarity in data trends with actual reported data—thus providing a strong prediction model.

Hamzaha et al. (2020) conducted a predictive analysis using the SEIR model and sentiment analysis of verified news into positive and negative news. Even though this is not a new approach, the research study is able to illustrate the findings in a very crisp manner using data visualizations.

Zou et al. (2020) build upon exiting epidemic models like SIR and SEIR and propose a new model that takes into account the lack of reporting of COVID-19 cases, resulting in the inaccuracy of existing models. Their model SuEIR takes into account untested or unreported cases while predicting the rate of cases (active or deaths) of COVID-19 infection. They used machine learning methods to train their model. The model (as shown in Fig. 5 (Zou et al. 2020)) is unique because it doesn't simply fit the current curve, which is based only on reported cases. Rather, it infers the number of untested and unreported cases from the model's data analysis and uses the inferences to predict how quickly the disease

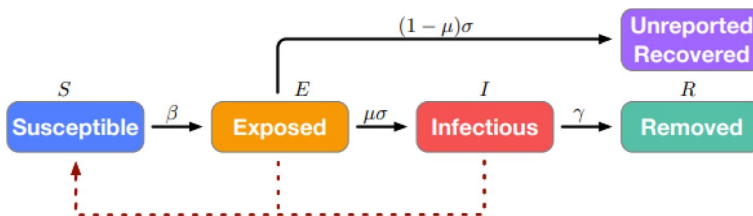


Fig. 5 Illustration of the SuEIR model. Solid lines represent the transitions of individuals and dashed lines represent the routes of infection

will spread. Another key inference from this model is that many folks who are exposed to COVID-19 may recover or, unfortunately, die without being tested and/or reported. The biggest challenge to substantiate the findings of this new model will be data. If data is indeed under—or not reported, this model's predictions may not be able to be verified.

During the early phases of the COVID-19 outbreak, another model used by Kuefel (2020) to predict the spread of COVID-19 was Auto-regressive Integrated Moving Average (ARIMA) model. The time-series model aimed to show the different epidemic phases of the infections amongst communities, using data available at the time to possibly predict future spread. Similar to other tried models, the method was unable to compensate for interventions like non-medical interventions and the impact of testing schemes deployed.

Fanelli and Piazza (2020) use the Mean-field approximation method on COVID-19 data from three hotspots across the world. They can substantiate and establish a prediction model for a maximum number of infected individuals along with the timing of the peak. They are further able to show, using simple quantitative models, how containment efforts can help in reducing the spread. The authors break data into the following four classes: susceptible, infected, recovered, and deaths (SIRD). Based on their assessment of the dataset for one region, they are able to establish predictions for other regions as they approach the peak. The model's key drawback is that it assumes standard conditions and fails to track with rapid recovery (decrease in the number of infected cases) and overestimates the number of deaths when extreme measures like social distancing are used by local authorities to flatten the curve. The model also fails to incorporate cultural aspects of different regions, which can impact the infection rate.

Sajadi et al. (2020) worked on the premise that many diseases display seasonal patterns. They studied climate data with the intent to establish correlation in regions that have similar climate setup and come up with a model to predict possible new locations based on the similarity of climate with the current COVID-19 hotspots. Key findings from their data analysis suggest hotspots be concentrated in a 30–50° N' latitude corridor, with low average temperatures, low specific and absolute humidity. The key caution called out for the prediction model is that while it establishes a very strong correlation of COVID-19 hotspots with latitude, temperature, and humidity, there needs to be caution in establishing absolute correlation of COVID-19 spreading in areas with these climatic factors. The key reason is the lack of human factors like intervention, other climatic factors like cloud cover, and viral factors like a mutation of the virus, which can lead to the unpredictability of the model.

Mollalo et al. (2020) utilized spatial models coupled with multiple socioeconomic data factors to try to explain the variation of COVID-19 in the United States (USA); based on geographic modeling. Again, their intent was to use well-established GIS toolsets to help explain the distribution of COVID-19. They started with identifying 35 socioeconomic, behavioral, environmental, topographic, and demographic factors. They used five different models (3 global and two local) to finally zone in the following four variables: median household income, income inequality, percentage of nurse practitioners, and percentage of black female population, which provided an explanation for geographical spread and variability of COVID-19 in the USA. Unlike some other studies, their model didn't significantly impact environmental factors like temperature and air quality on the distribution of COVID-19. This is an important point, as many studies have suggested the impact of temperature conditions to contribute to the spread of COVID-19 in certain parts of the world, especially in Asia. Their findings on income equality and median income lend further credence to a higher rate of COVID-19 spread amongst folks belonging to lower-income groups. The authors acknowledged data

availability, especially county level, as a limitation that hindered further development of their analysis from an accuracy standpoint. Another limiting factor the impact of how stringently the local authorities implemented lockdown procedures. Given the lack of uniformity, it was difficult to isolate or model this particular factor. The authors also didn't focus on pre-existing conditions in their analysis, which has been considered a contributor to the spread of COVID-19. Still, the local model MGWR (multiscale geographically weighted regression) performed consistently and helped introduce GIS data/tools to help predict or explain the variation of the spread of COVID-19.

Kuchler et al. (2020) used aggregated (anonymized) data from Facebook; of two early hotspots in an attempt to establish a correlation between socially connected people on Facebook, leading to the spread of COVID-19 in such areas. They came up with a mathematical equation for social connectedness index—which establishes a relationship of spread to areas with social ties to these hotspots. The premise is that individuals connected across two regions via a social platform like Facebook can predict the potential spread of COVID-19 to new regions as people from the hotspot potentially move or are in physical contact with people from these new regions. In yet another attempt, the local government has tried to use aggregated mobile phone data to zoom in on hotspots and, in some cases, using prediction techniques to anticipate where the next hotspot may emerge or help with much-needed 'contact tracing' (Zhao et al. 2020). While the results show a possible correlation that can be used as one of many ways to predict possible patterns of the spread of COVID-19, the social connectedness index reflects relative probability and, in many cases, may not establish actual correlation. The authors themselves admit that their work is in its infancy and requires more work.

Furthermore, this may still be considered as a proof of concept and not an epidemiological model. Also need to be taken into consideration is that it's a social media account, and people do tend to input incorrect information; thus, it is possible that data may be flawed if user profiles are incorrect, thus providing incorrect numbers. Another possible challenge in using data from social platforms or mobile phones could be the threat of invasion of privacy. Many users may opt-out or not be open to sharing their whereabouts.

Tuli et al. (2020a), in their paper, worked on improving the Weibull distribution model. While using ML techniques and iterative weighting strategy, they came up with a 'Robust Fitting' technique. It showed improved predictions over the Gaussian predictions. Like any other model, the key constraints were that it could not account for lockdown restrictions being lifted or virus mutations that may change the speed of the spread. Tuli et al. (2020b) further tried to improve their approach by combining the Weibull method with LSTM. They demonstrated much better results of this new approach when compared to the ARIMA, LSTM, and other variants. Their research demonstrated that hybrid approaches were best able to use epidemic data and then apply ML techniques to smooth over outliers, train to respond to shifting trends that may influence the spread, density of spread, and fluctuations in data—to come up with the best forecasts.

Zheng et al. (2020) used State Transition Matrix Model (STM) to predict the inflection point (IFP) in three countries: South Korea, Italy, and Iran. They used data available in the first quarter of 2020 to draw inferences and provide three scenarios. The researchers were able to predict, for Italy and South Korea, the total confirmed and increment of confirmed cases during the months of April and May. This kind of prediction at the time and still is a valuable tool for countries and local authorities to plan for the inflection point and help formulate containment strategies.

Shahid et al. (2020) have compared several prediction models used. They produced model rankings—from good performance to the lowest, which were: Bi-LSTM, LSTM,

GRU, SVR, and ARIMA. This helps the researchers and authorities to focus on a model that may produce the best predictive results; to help plan for containment.

Car et al. (2020) used a dataset operated by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) and supported by the ESRI Living Atlas Team and the Johns Hopkins University Applied Physics Lab (JHU APL) (Johns Hopkins 2020). Data contained information of coronavirus patients—infected, recovered, and deceased—based on their location.

This dataset though a time-series dataset, was transformed into a regression dataset and used to train an artificial neural network (ANN) (Hui et al. 2020). The aim was to achieve a worldwide model of the maximal number of patients across all locations in each time unit. Post-training and cross-validation of the model, R2 scores dropped to 0.94 for confirmed, 0.781 for recovered, and 0.986 for deceased patient models. This showed high robustness of the deceased patient model, good robustness for confirmed, and low robustness for the recovered patient model. Figure 6a and b show the performance of the model for infected patients while comparing real and predicted data.

In another vein, to predict how the coronavirus is spreading and how the lockdown area could be predicted in crowded areas, Maghdid et al. (2004) propose a new model prediction based on using K-means clustering algorithm. The model on a server has been implemented to receive the participant users' location information periodically and send back the prediction status to the users. The main aim of the model is to avoid un-necessary lockdown areas and consequently mitigate the economic crisis. Applying the lockdown area due to spreading coronavirus COVID-19 via most of the countries across the globe has a negative impact on economic issues. Several experiments, scenarios, and hypotheses have been conducted and analyzed to prove the validity of the prediction model. Figure 7 shows an example of the result model prediction for two different scenarios in the Denver area Aspen and area in Colorado-USA.

The study for the spread of COVID-19 is still in its infancy, and the above paraphernalia of studies shows that there is a multitude of factors affecting its spread and mutation. Table 1 presents a comparison of different methods and models. It offers a quick glance at salient features of each method we studied, and it can help further the process for different scientists.

4 AI and technologies to fight COVID-19 pandemic

As the world shutdown due to COVID-19, a clear need was felt on what would need to be done to safely re-open. This is where technology has stepped in with solutions to fight this pandemic. Some key areas contributing are Internet of Things (IoT), Mobile Apps, Artificial Intelligence (AI), and Autonomous technology (vehicles, robots, unmanned aerial vehicles).

AI technology has stepped-in significantly to spearhead research of COVID-19 data and help with predictions. AI tools are being used to come up with detection, diagnosis, and prediction. They have gone further to help forecast the impact of different measures that local authorities can take and how best to establish a medical response to the rising of COVID-19 infections. AI tools have also helped forecast the socio-economic impact of this virus, highlighting at-risk groups and economic impact across the globe.

Pre-COVID, IoT devices had already taken off in households. Post-COVID, they continue to showcase their value in even more ways. With global medical treatment

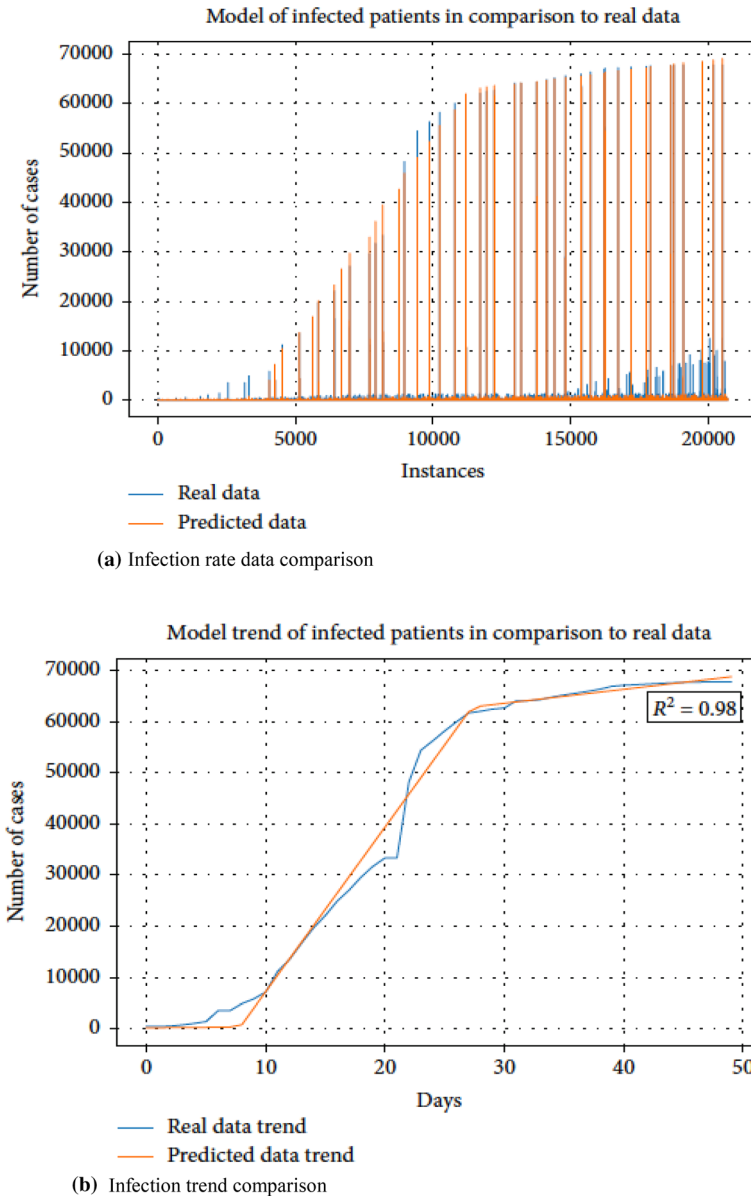


Fig. 6 **a** Infection rate data comparison, **b** Infection trend comparison

restrictions, IoT-enabled devices like Thermometers, Smart-wearable devices, and remote connection technologies have enabled telehealth. Patients are able to avail medical guidance from the confines of their homes and not have to visit a doctor's office or a hospital. While keeping patients safe from getting infected, these technologies are also helping to ease pressure on the already stretched medical staff.

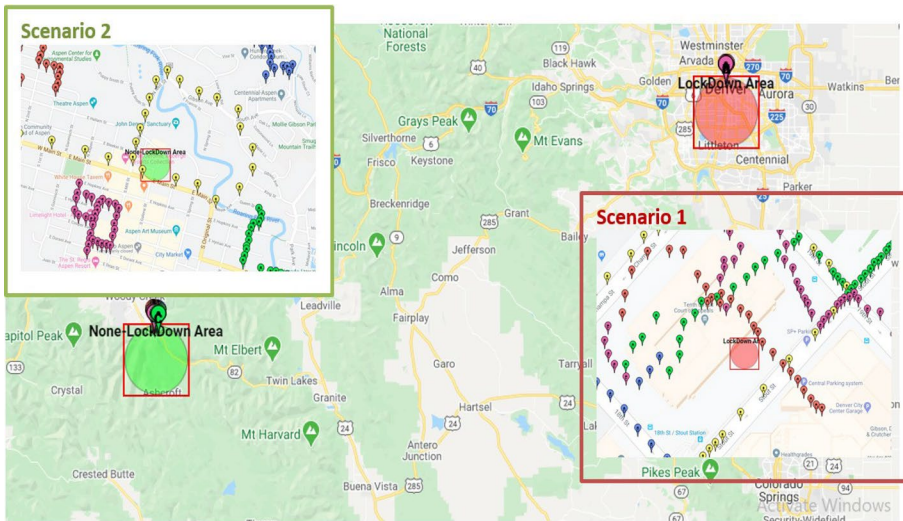


Fig. 7 The results of lockdown prediction model for two different scenarios

In the past decade, smartphone use has continued to rise globally. While enabling many underprivileged sections across the world, they have also led to commoditizing services via mobile apps. Mobile apps have brought many services to the household. They have further helped with the fight against COVID-19. Many private and public entities have come up with apps to share information related to this virus across the globe. Mobile apps have played a crucial role in contact tracing, a crucial tool needed to fight against the spread of COVID-19. In recent work, authors in (Maghdid et al. 2003) proposed a new framework to diagnose the coronavirus COVID-19 using onboard smartphone sensing data via multi-sensors technologies including camera, microphone, accelerometers, and fingerprint-temperature sensors. The proposed framework doesn't need any extra hardware, and it is working under the running application on the smartphone. Thus, such a solution enables doctors to use the application on their smartphones to diagnose the disease in a short time and at a lower cost in comparison to other exciting solutions.

Autonomous vehicular technology (vehicles, robots, unmanned aerial vehicles) also plays an important role in the fight against this virus. Their use has grown rapidly to address the need to deliver products (groceries, medical supplies, and other essential goods) without personal contact, enable surveillance to help curfew enforcement, spraying of disinfectants, and establish mobile temperature detection in open or large areas like malls, and hospitals. Figure 8 shows the AI and technologies to confront the COVID-19 pandemic.

Kumar et al. (2020) used the latest innovations in AI and Drone technology to propose a suite of applications in support of fighting COVID. Their proposals range from using the technology to enable indoor and outdoor monitoring, observations, sanitization, and data collection. Their solution expands the use of the Internet of Medical Things (IoMT) using AI-based algorithms, networking, cloud, and storage solutions. Overall, they pitch a comprehensive medical solution. They, however, note the limitations in real-life uses and adoption, which depend upon further development in programming and protocols to expand usage.

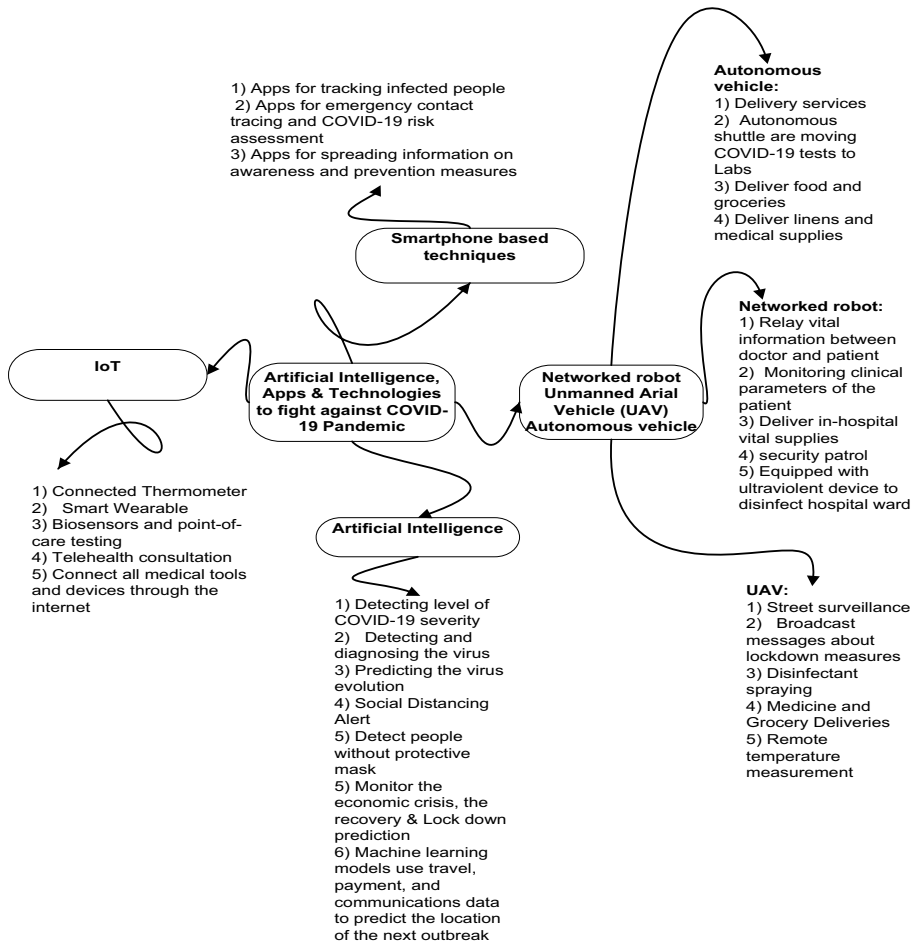


Fig. 8 AI and technologies to confront COVID-19 pandemic

Singhal et al. (2020) expounded on how using digital technologies, educational institutions adapted to the COVID lockdown. Their research suggested an iterative and evidence-based active learning process that was able to help students keep up with their studies and showed improved performance and satisfaction rates. While more research may be needed, this paper demonstrated how compulsions due to COVID lockdown might introduce long-term changes in how we conduct our day-to-day life, including something as traditional as education. This can potentially act as a positive disrupter in technical educational techniques, bringing in game-changing solutions that improve teachers' ability to deliver and for students to receive knowledge.

5 Lessons learned

This paper shows different models which are successfully used in predicting the spread of the virus. Considering that the study for the diseases is still in its infancy, using logical parameters and accurate values for calculations can help with reasonable predictions.

- Multiple studies have conducted their research with ideal conditions for data analysis. They have not factored in the impact of social distancing, lockdown controls, typical human social behavior that may lead to inaccuracy in predicting the spread of this virus.
- Many papers have also assumed typical scenarios of how COVID-19 may have started or spread in certain regions. Furthermore, they seem to have assumed typical symptoms as indicators of the spread of COVID-19. This may lead to false-positive in the dataset if patients that are not infected by the virus are tagged as having this virus.
- There have been some novel approaches used in papers we researched.
 - Zou et al. (2020), in their paper, considered unreported and untested cases to train their AI model. If their predictions are true, the inference that can be drawn from the research is that many patients who may recover or die from COVID-19 may go unaccounted for, skewing the overall infection/death rates.
 - In another unique approach, Kuchler et al. (2020) used data from a widely used social media platform to establish a relationship between people who are socially connected to the possibility of spreading the virus in regions where these socially connected people reside.
 - Lastly, Mollalo et al. (2020) utilized spatial models and socio-economic data to establish viral spread patterns amongst certain sections of society.
- Eliminating the unknowns and replacing them with strong, scientifically-based data can result in models that can help with accurate predictions and increasing the confidence level of regulatory authorities and the public in following recommendations to curb the spread of COVID-19.
- AI and other technologies can be used as an important tool in a fight for a diagnosis, a fight against the spread, and eventually designing a robust cure for COVID-19.
- The work is done by Tuli (2020b) and Shahid (2020) in establishing prediction models and comparing much-used approaches ultimately lays the groundwork for research that can help create a robust forecast model to track and predict the spread of COVID-19. It is pretty clear that Weibull-based Long-Short-Term-Memory (W-LSTM) and Bi-directional long short-term memory (Bi-LSTM) models so far performed better with regards to being closely fitted and being trained to produce forecast data with a higher level of accuracy.

6 Open research challenges

While the use of IoT, cloud computing, and the latest IT-based technologies enabled effective, fast, reliable, and scalable solutions in the fight against COVID-19, they also open up exploiting existing vulnerabilities with these technologies to be used by hackers and adversaries for malicious intents. Our hope is that the accuracy and reliability of these

technologies, along with increased adoption for COVID-19 related research, may act as a catalyst in further improving them.

It is also pretty clear that the accuracy of predictions will heavily depend on the adequacy of the available. However, due to limited coverage of the data collected, the data-driven model may not perform satisfactorily; if applied to a new epidemic with different characteristics. Furthermore, in most cases, the models did not consider the effects of different control strategies, and hence the rate of forecast varied from actual infection rates. This may be an open area of focus for future research so as to widen the coverage of possible progression profiles and incorporate the effects of different control measures on the epidemic progression profile.

7 Conclusion

In this paper, we have tried to compile a broad array of research that different teams have undertaken globally to explain the spread of COVID-19 from a multitude of angles, using data made available by public and private establishments. While some research has validated the trends seen during the early spread of this virus, others have built upon and forecasted this virus's future spread. Significant effort has been put into explaining the factors that may help control the spread or 'flatten the curve.' The paper explores the use of AI and other technologies in fighting COVID-19. Serious effort has also been invested in using 'data-driven analysis to dispel speculation related to COVID-19. Some articles (Hamzaha 2020) during the initial global outbreak used publicly available data to try not only to put predictions regarding the potential spread patterns but also examined social media posts to call out the prevalent public sentiment on the origin and containment efforts. The publications were helpful to the policymakers not only from assessing the political and economic influence of the virus spread; but also cautioned them on opinions or speculations forming amongst communities. Assessment of multiple models and studies (Wynants et al. 2020) showed the high risk of bias and creating over-optimistic estimates/forecasts and extreme cases, even misleading interpretation of available data. The reason for bias was found to mainly be due to the nonrepresentative selection of control patients, exclusion of certain patient types, and model overfitting.

While a lot of work has been done to come up with the best model that can improve the accuracy of the forecasts, it is pretty clear that active effort is still needed to continue to source the latest data, train the models to factor in non-epidemic constraints; so that predictions are able to assist medical and governmental agencies in coming up with a fighting chance to control, contain and hopefully eliminate this epidemic.

Acknowledgements This work was in part supported by Saginaw Valley State University.

References

- Bailey M, Cao R, Kuchler T, Stroebel J, Wong A (2018) Social connectedness: measurement, determinants, and effects. *J Econ Perspect* 32(3):259–280
- Car Z, Šegota SB, Anđelić N, Lorencin I, Mrzljak V (2020) Modeling the spread of COVID-19 infection using a multilayer perceptron. *Comput Math Methods Med* 5714714:10
- Fanelli D, Piazza F (2020) Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos Solitons Fractals* 134:109761

- Hamzaha FAB, Laub CH, Nazric H, Ligotd DV, Lee G, Tanf CL, Shaibg MKBM, Zaidonh UHB, Abdul-lahi AB, Chungj MH, Ongk CH, Chewl PY, Salungam RE (2020) CoronaTracker: worldwide COVID-19 outbreak data analysis and prediction. *Bull World Health Organ* 1:32
- Haytham H, Elmousalami, Hassanien AE (2020) Day level forecasting for coronavirus disease (COVID-19) spread: analysis, modeling and recommendations. arXiv preprint <https://arxiv.org/abs/2003.07778>
- Hui DS, Azhar EL, Madani TA et al (2020) The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health: the latest 2019 novel coronavirus outbreak in Wuhan, China. *Int J Infect Dis* 91:264–266
- Johns Hopkins CSSE (2020) Novel Coronavirus (COVID-19) Cases. <https://github.com/CSSEGISandData/COVID-19>
- Johns Hopkins University and Medicine, Coronavirus Resource Center: Johns Hopkins university and medicine, coronavirus resource center (2020). <https://coronavirus.jhu.edu/map.html>
- Kuchler T, Russel D, Stroebel J (2020) The geographic spread of COVID-19 correlates with structure of social networks as measured by Facebook. Tech. rep., National Bureau of Economic Research
- Kufel T (2020) ARIMA-based forecasting of the dynamics of confirmed COVID-19 cases for selected European countries. *Equilib Q J Econ Econ Policy* 15(2):181–204
- Kumar A, Sharma K, Singh H, Naugriya SG, Gill S, Buyya R (2020) A drone-based networked system and methods for combating coronavirus disease (COVID-19) pandemic. *Futur Gener Comput Syst* 115:1–19
- Maghdid HS, Ghafoor KZ (2020) A smartphone enabled approach to manage COVID-19 lockdown and economic crisis. *SN Comput Sci* 1(5):1–9
- Maghdid HS, Ghafoor KZ, Sadiq AS, Curran K, Rabie K (2020) A novel AI-enabled framework to diagnose coronavirus COVID-19 using smartphone embedded sensors: design study. arXiv preprint <https://arxiv.org/abs/2003.07434>
- Mollalo A, Vahedi B, Rivera K M (2020) Gis-based spatial modeling of COVID-19 incidence rate in the continental united states. *Sci Total Environ* 728:138884
- Pham Q-V, Nguyen DC, Huynh-The T, Hwang W-J, Pathirana PN (2020) Artificial intelligence (AI) and big data for coronavirus (COVID-19) pandemic: a survey on the state-of-the-arts
- Sajadi MM, Habibzadeh P, Vintzileos A, Shokouhi S, Miralles-Wilhelm F, Amoroso A (2020) Temperature, humidity and latitude analysis to predict potential spread and seasonality for COVID-19. Available at SSRN 3550308
- Shahid F, Zameer A, Muneeb M (2020) Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos Solitons Fractals* 140:110212
- Singhal R, Kumar A, Singh H, Fuller S, Gill SS (2020) Digital device-based active learning approach using virtual community classroom during the COVID-19 pandemic. *Comput Appl Eng Edu*. <https://doi.org/10.1002/cae.22355>
- The Guardian (2020) Facebook friendships can help predict COVID-19 spread, study finds. URL <https://www.theguardian.com/world/2020/apr/14/facebook-friendships-can-help-predict-COVID-19-spread-study-finds>
- Tomar A, Gupt N (2020) Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. *Sci Total Environ* 728:138762
- Tuli S, Tuli S, Tuli R, Gill SS (2020) Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. *Intern Things* 11:100222
- Tuli S, Tuli S, Verma R, Tuli R (2020) Modelling for prediction of the spread and severity of COVID-19 and its association with socioeconomic factors and virus types. medRxiv
- Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, Bonten MMJ, Dahly DL, Damen JAA, Debray TPA, de Jong VMT, De Vos M, Dhiman P, Haller MC, Harhay MO, Henckaerts L, Heus P, Kreuzberger N, Lohmann A, Luijken K, Ma J, Martin GP, Navarro CLA, Reitsma JB, Sergeant JC, Shi C, Skoetz N, Smits LJM, Snell KIE, Sperrin M, Spijker R, Steyerberg EW, Takada T, Tzoulaki I, van Kuijk SMJ, van Royen FS, Verbakel JY, Wallisch C, Wilkinson J, Wolff R, Hooft L, Moons KGM, van Smeden M (2020) Prediction models for diagnosis and prognosis of COVID-19 infection: systematic review and critical appraisal. *BMJ* 369:m1328
- Yan L, Zhang H-T, Goncalves J, Xiao Y, Wang M, Guo Y, Sun C, Tang X, Jing L, Zhang M, Huang X, Xiao Y, Cao H, Chen Y, Ren T, Wang F, Xiao Y, Huang S, Tan X, Huang N, Jiao B, Cheng C, Zhang Y, Luo A, Mombaerts L, Jin J, Cao Z, Li S, Hui X, Yuan Y (2020) An interpretable mortality prediction model for COVID-19 patients. *Nat Mach Intell* 2(5):283–288
- Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, Liu P, Cao X, Gao Z, Mai Z, Liang J, Liu X, Li S, Li Y, Ye F, Guan W, Yang Y, Li F, Luo S, Xie Y, Liu B, Wang Z, Zhang S, Wang Y, Zhong N, He J (2020) Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis* 12(3):165

- Zhao Z, Li X, Liu F, Zhu G, Ma C, Wang L (2020) Prediction of the COVID-19 spread in African countries and implications for prevention and controls: a case study in South Africa, Egypt, Algeria, Nigeria, Senegal and Kenya. *Sci Total Environ* 729:138959
- Zheng Z, Ke W, Yao Z, Zheng J, Chen J (2020) The prediction for development of COVID-19 in global major epidemic areas through empirical trends in China by utilizing state transition matrix model. *BMC Infect Dis* 20(1):1–12
- Zou D, Wang L, Xu P, Chen J, Zhang W, Guk Q (2020) Epidemic model guided machine learning for COVID-19 forecasts in the united states. medRxiv

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.