Prediction of Global Spread of Covid-19 Pandemic: A Review and Research Challenges

Saloni Shah [a], Aos Mulahuwaish [b]*, Kayhan Zrar Ghafoor [c], Halgurd S. Maghdid [d]

[ab] *Department of Computer Science and Information Systems Saginaw Valley State University, MI, USA*

[c] *Department of Software Engineering, Salahaddin University, Erbil, Iraq; School of Mathematics and Computer Science, University of Wolverhampton, Wolverhampton, UK*

[d]*Department of Software Engineering, Faculty of Engineering, Koya University, Kurdistan Region-F.R.Iraq.*

[a] *skshah@svsu.edu;* [b]*amulahuw@svsu.edu;* [c]*kayhan@ieee.org, halgurd.maghdid@koyauniversity.org*

Abstract

Since the initial reports of the Coronavirus surfacing in Wuhan, China; the novel virus currently without a cure has spread like a 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 analyse and predict the pattern and conditions that would influence the spread of this unforgiving virus. Using existing statistical modelling tools to latest AI 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 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 in compiling different research methods and comparing their accuracy in predicting the spread of COVID-19.

Keywords: COVID-19; Machine Learning; Deep Learning; Prediction Methods

1. Introduction

The global spread of the Coronavirus disease has not only become a healthcare concern, but the rapid mutation of the strain and high infectious rate of the disease has made it a socioeconomic issue for countries all over the globe. The World Health Organization (WHO) was alerted by Chinese officials about dozens of pneumonia-like diseases in the city of Wuhan as new celebration was taking place in the country. Then, 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 COVID-19 disease outbreak as a pandemic on 11 March 2020. Figure 1 shows the rapid increases in cases globally [13].

*Corresponding author. Email: amulahuw@svsu.edu

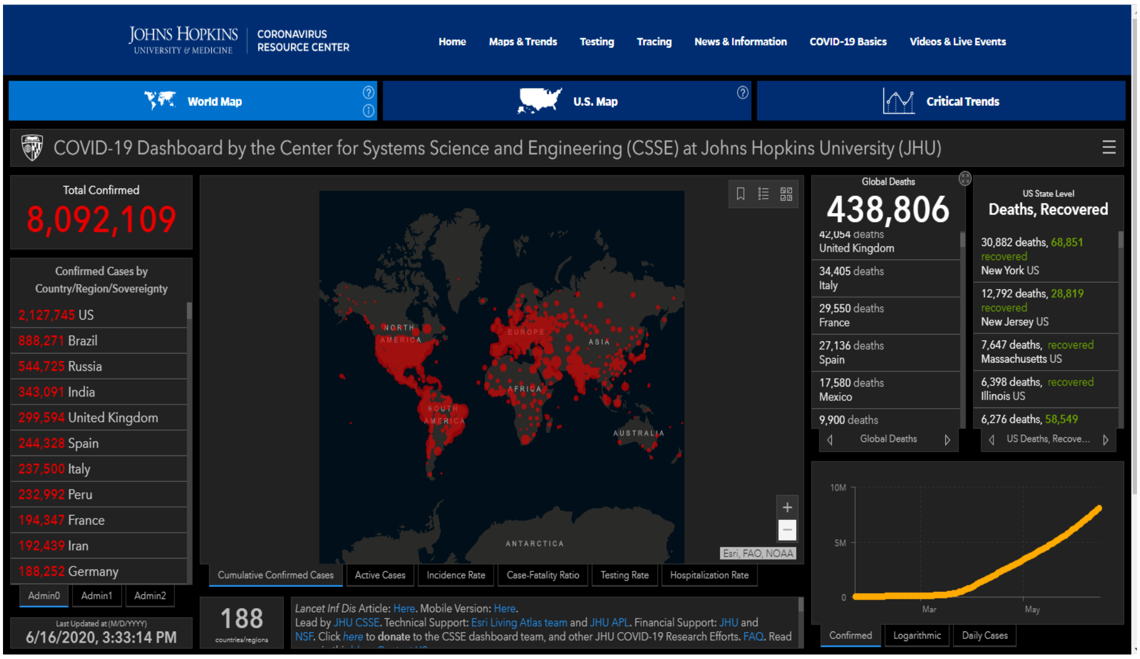

Figure 1: Snapshot of the Johns Hopkins University, Coronavirus Resource Centre (June 16, 2020)

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 so as to curb the spread, find a cure and in the meanwhile help local authorities all over the world to 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 are heavily relying 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 the fact 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 end up affecting 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 [3] and [14]. The social connectedness index discussed in Kuchler et al. [3] can have a major impact on identifying the links between mass population from one geographical area to the other. This connectedness identified by the social media can open possibilities to correlate the spread of the disease that depend on different demographics or geographies.

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. More precisely, 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. The findings of this study presents that several parameters will contribute to the spreading of the virus such as social interactions measured by social networks. We also include a section on how AI and different technologies are being used to combat COVID-19. Our intent is to help the scientists and researchers to get a clear idea of research already done and guide them on how to proceed with their own studies.

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

2. Prediction Models

Prediction models can be categorized based on machine learning models and mathematical models, as depicted in Figure 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 Figure 3 [7]), to try 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 predication 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. Figure 3 shows the use of Big Data in different applications for fighting Covid-19.

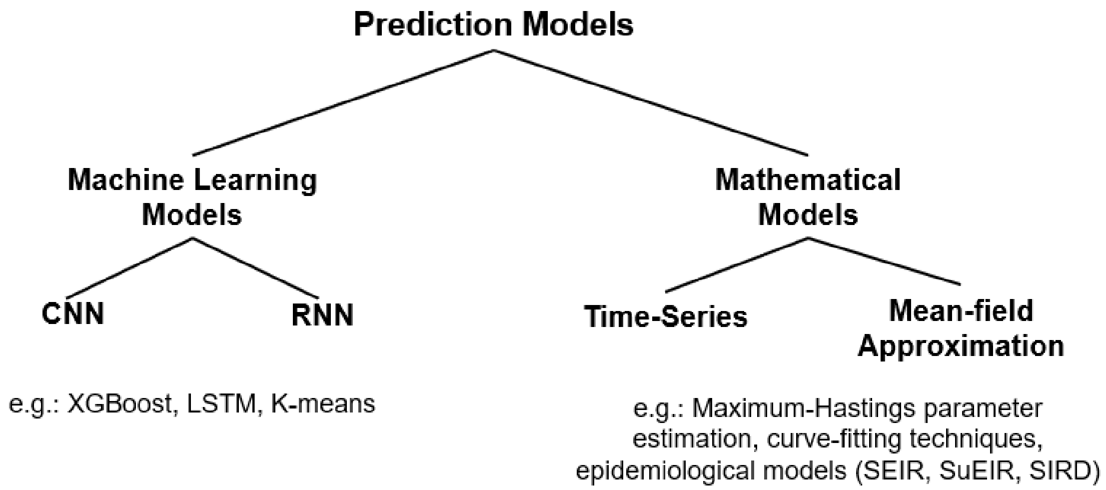

Figure 2: Categorization of prediction models based on machine learning and mathematical models

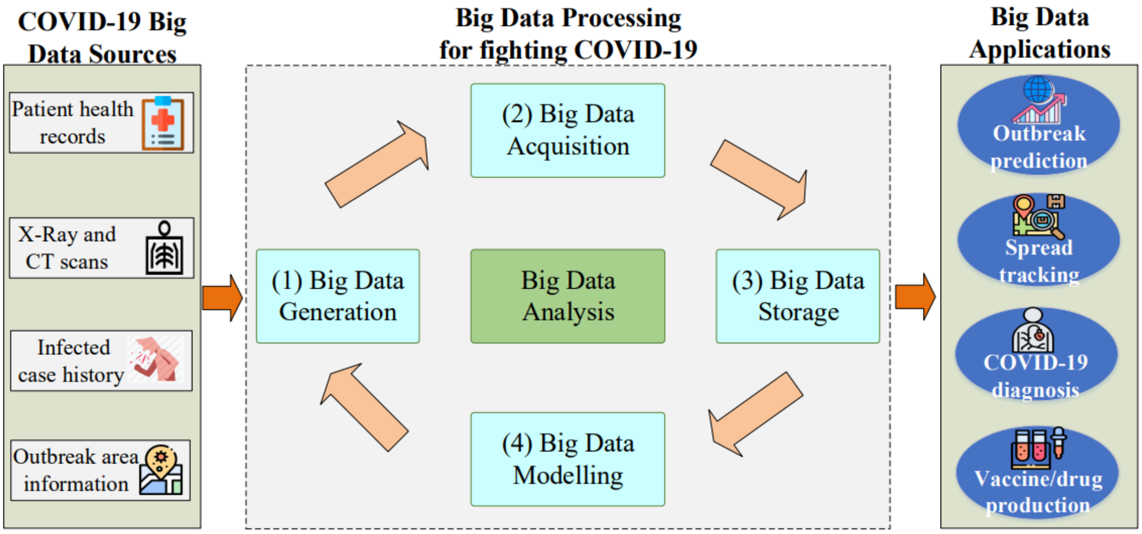

Figure 3: Big Data and its applications for fighting COVID-19 pandemic.

Li Yan et al. [1] focused on using biomarkers, obtained via blood samples, to be able to predict severe COVID-19 cases that result in 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 patient with high accuracy. Furthermore, their trained models were able to predict worsening case well in advance, up to 10 days in some cases; giving medical professional a fighting chance to change treatment of a patient. This research showcases use of existing medical knowledge of biomarkers coupled with AI tools to help forecast severe cases in advance.

Elmousalami and Hassanien [10], use time series model to analyze and predict 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 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 [2], used Long Short-Term Memory (LSTM) – a type of Recurrent Neural Networks (RNN) method to predict 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 Figure 4 [2]) was also able to predict positive and recovered cases within a certain accuracy range. They deployed curve fitting technique to assess accuracy of their prediction model come up with close results. This tool has the potential to help local authorities with forecasting infection and recovery rates; giving them needed data to prepare for outbreak and plan deployment of limited medical resources.

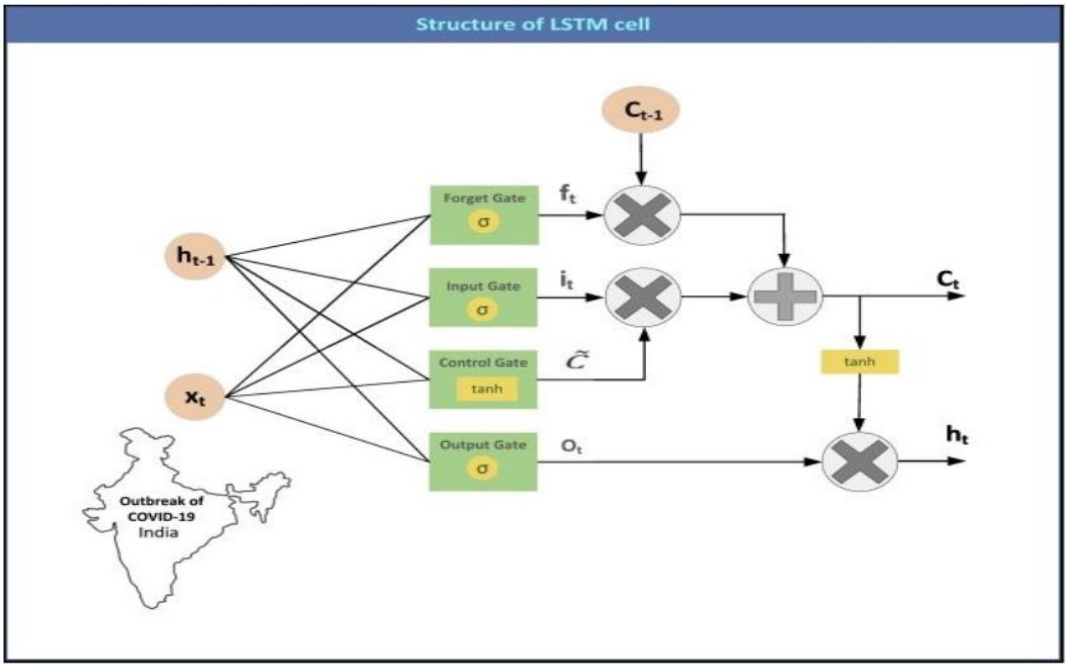

Figure 4: Basic structure of LSTM

Zhao et al. [4] use the Maximum-Hasting (MH) parameter estimation method and the modified Susceptible Exposed Infectious Recovered (SEIR) model 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. The authors study three possible scenarios to deploy: suppression, mitigation, or mildness in six African countries. In case of suppression, local authorities maintain a 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 controlled infection rate. In the second intervention method, 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 atleast 10 days. African nations of Algeria, Nigeria and Kenya show infection curve aligned with this intervention policy. The last intervention scenario lacks proper mitigation and will result in 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 [6], Yang et al. uses 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 impact of intervention measures. With a certain success rate, they attempt to predict the impact of delaying intervention leading into second outbreak/peak. They work on the Long Short-Term Memory (LSTM) model to predict 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 similarity in data trends with actual reported data – thus providing with a strong prediction model.

Zou et al. [11] build upon exiting epidemic models like SIR and SEIR and propose a new model that takes into account lack of reporting of COVID-19 cases, resulting in inaccuracy of existing models. Their model SuEIR takes into account untested or unreported cases while predicting rate of cases (active or deaths) of COVID-19 infection. They used machine learning methods to train their model. The model (as shown in Figure 5 [11]) 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 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, the predictions of this model may not be able to be verified.

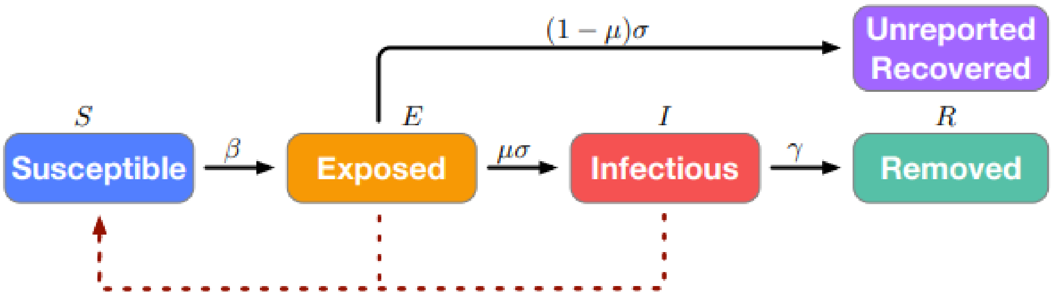

Figure 5: Illustration of the SuEIR model. Solid lines represent the transitions of individuals and dashed lines represent the routes of infection.

Fanelli and Piazza [5] use Mean-field approximation method on COVID-19 data from three hotspots across the world. They are able to substantiate and establish a prediction model for maximum number of infected individuals along with 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 following four classes: susceptible, infected, recovered and deaths (SIRD). Based on their assessment of dataset for one region, they are able to establish predictions for other regions as they approach 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 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. [8] worked on a 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 similarity of climate with the current COVID-19 hotspots. Key findings from their data analysis suggest hotspots to be concentrated in 30-50 deg 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 lack of human factors like intervention, other climatic factors like cloud cover, and viral factor like mutation of the virus which can lead to unpredictability of the model.

Mollalo et al. [9], utilized spatial models coupled with multiple socio-economic data factors to try explaining variation of COVID-19 in the Unites States (USA); based on geographic modelling. Again, their intent was to use well established GIS toolsets to help explain the distribution of COVID-19. They started with identifying 35 socioeconomic, behavioural, environmental, topographic, and demographic factors and used 5 different models (3 global and 2 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 find significant impact of environmental factors like temperature and air quality on distribution of COVID-19. This is an important point, as many studies have suggested 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 higher rate of COVID-19 spread amongst folks belonging to lower income groups. The authors acknowledged data availability, especially county level, as a limitation which 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; something which has been considered as 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. [3] used aggregated (anonymized) data from Facebook; of two early hotspots in an attempt to establish correlation between socially connected people on Facebook leading to 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 social platform like Facebook can predict potential spread of COVID-19 to new regions; as people from the hotspot potential move or are in physical contact of people from these new regions. In yet another attempt, local government have tried to use aggregated mobile phone data to zoom in on hotspots and in some cases using prediction techniques anticipate where the next hotspot may emerge or help with much needed 'contact tracing' [4]. 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.

In another vain, to predict how the coronavirus is spreading as well as how the lockdown area could be predicted in the crowded areas, Maghdid et al. [15] 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 area and consequently to mitigate the economic crisis. Since applying the lockdown area due to spreading coronavirus COVID-19 via most of the countries across the global has negative impact on the economic issues. To prove the validity of the prediction model, several experiments, scenarios, and hypothesis have been conducted and analysed. Figure 6 shows an example of the result model prediction for two different scenarios in Denver area Aspen and area in Colorado-USA.

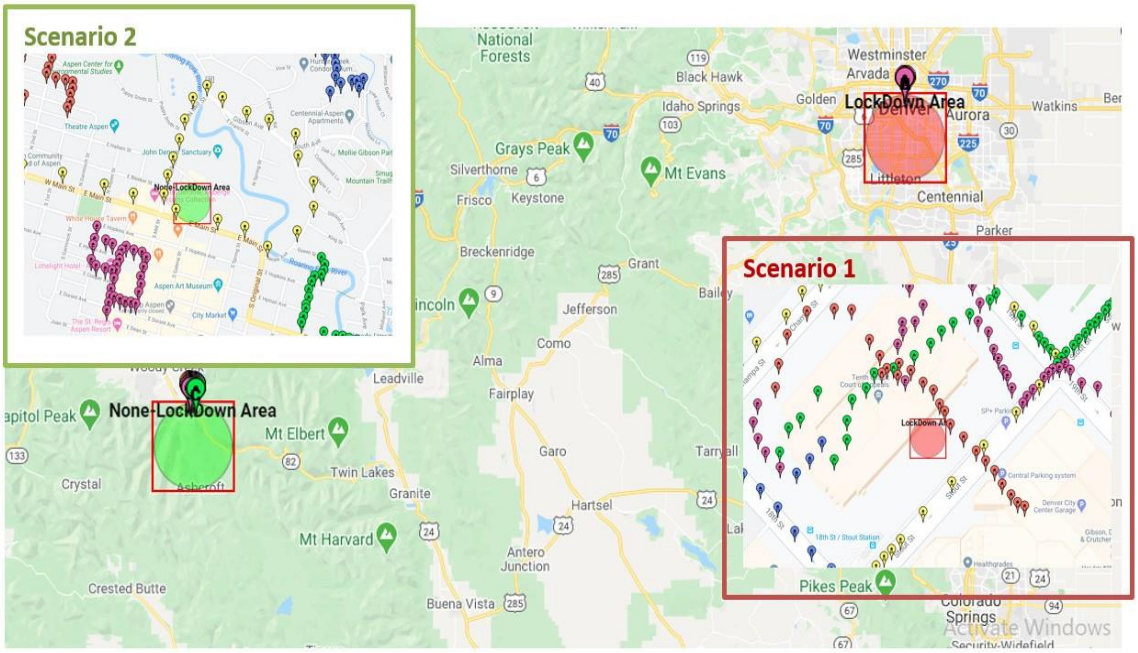

Figure 6: The results of lockdown prediction model for two different scenarios.

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

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. [1]	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 Hassanien [10]	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 [2]	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. [4]	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 Infectious 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)

Yang et al. [6]	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. [11]	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.
Fanelli and Piazza [5]	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. [8]	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.

Mollalo et al. [9]	Using spatial models coupled with multiple socio-economic data factors to try explaining variation of the spread of COVID-19 in the Unites 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 USAFacts (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. [3]	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. [12]	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
Maghdid et al.[15]	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.

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

In the past decade, smartphone use has continued to rise globally. While enabling many underprivileged sections across the world, they have also lead to commoditizing services via the use of mobile apps. Mobile apps have brought many services to the household. They have further helped with fight against COVID-19. Across the globe, many private and public entities have come up with apps to share information related to this virus. Mobile apps have played a crucial role in contact tracing, a crucial tool needed to fight against the spread of COVID-19. In a recent work, authors in [16] 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 solution enables doctors to use the application on their smartphones to diagnose the disease with a short time and at lower cost in comparison other exiting solutions.

Autonomous vehicular technology (vehicles, robots, unmanned aerial vehicles) is also playing a very 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 with curfew enforcement, spraying of disinfectants, and in establishing mobile temperature detection in open or large areas like malls, and hospitals. Figure 7 shows the AI and technologies to confront COVID-19 pandemic.

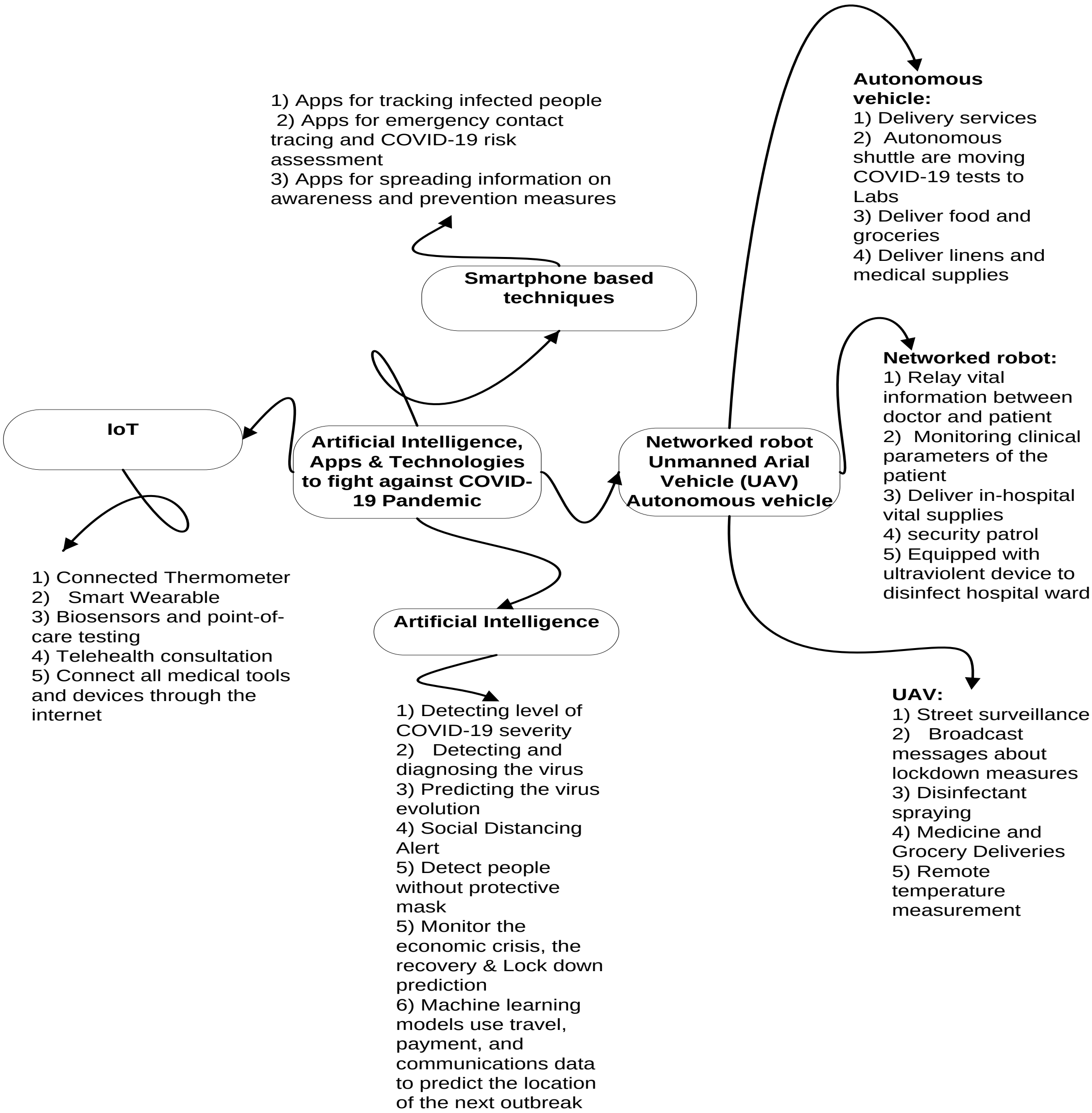

Figure 7: AI and technologies to confront COVID-19 pandemic

4. 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 impact of social distancing, lockdown controls, typical human social behavior that may lead to inaccuracy in predictions of 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 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. [11] in their paper took into consideration 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. [3] used data from a widely used social media platform to establish relationship between people who are socially connected to possibility of spread of the virus in regions where these socially connected people reside.
    - Lastly, Mollalo et al. [9], utilized spatial models and socio-economic data to establish patterns of spread of virus amongst certain sections of the 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 the diagnosis, a fight against the spread and eventually designing a robust cure for COVID-19.

5. 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 the future spread of this virus. Significant effort has been put on explaining the factors that may help control the spread, or ‘flatten the curve’. The paper explores use of AI and other technologies in fighting COVID-19. Serious effort has also been invested in using a ‘data-driven’ analysis to dispel speculation related to COVID-19.

Acknowledgment

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

References

1. Li Yan, Hai-Tao Zhang, Jorge Goncalves, Yang Xiao , Maolin Wang , Yuqi Guo , Chuan Sun, Xiuchuan Tang , Liang Jing , Mingyang Zhang , Xiang Huang , Ying Xiao , Haosen Cao, Yanyan Chen, Tongxin Ren, Fang Wang, Yaru Xiao , Sufang Huang, Xi Tan, Niannian Huang, Bo Jiao, Cheng Cheng , Yong Zhang , Ailin Luo, Laurent Mombaerts, Junyang Jin, Zhiguo Cao, Shusheng Li, Hui Xu and Ye Yuan: An interpretable mortality prediction model for COVID-19 patients. Nature Machine Intelligence pp. 1– 6 (2020)
2. Anuradha Tomar and Neeraj Gupt: Prediction for the spread of covid-19 in India and effectiveness of preventive measures. Science of The Total Environment p. 138762 (2020)
3. Theresa Kuchler, Dominic Russel, Johannes Stroebel: The geographic spread of covid-19 correlates with structure of social networks as measured by Facebook. Tech. rep., National Bureau of Economic Research (2020)
4. Zebin Zhao, Xin Li, Feng Liu, Gaofeng Zhu, Chunfeng Ma and Liangxu Wang: 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. Science of the Total Environment p.138959 (2020)
5. Duccio Fanelli and Francesco Piazza: Analysis and forecast of covid-19 spreading in china, Italy and France. Chaos, Solitons & Fractals 134, 109761 (2020)
6. Zifeng Yang, Zhiqi Zeng, Ke Wang, Sook-San Wong, Wenhua Liang, Mark Zanin, Peng Liu5, Xudong Cao, Zhongqiang Gao, Zhitong Mai, Jingyi Liang, Xiaoqing Liu, Shiyue Li, Yimin Li, Feng Ye, Weijie Guan, Yifan Yang, Fei Li, Shengmei Luo, Yuqi Xie, Bin Liu, Zhoulang Wang, Shaobo Zhang, Yaonan Wang, Nanshan Zhong, Jianxing He: Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. Journal of Thoracic Disease 12(3), 165 (2020)
7. Quoc-Viet Pham, Dinh C. Nguyen, Thien Huynh-The, Won-Joo Hwang, and Pubudu N. Pathirana: Artificial intelligence (AI) and big data for coronavirus (COVID-19) pandemic: A survey on the state-of-the-arts (2020)

8. Mohammad M. Sajadi, Parham Habibzadeh, Augustin Vintzileos, Shervin Shokouhi, Fernando Miralles-Wilhelm and Anthony Amoroso: Temperature, humidity and latitude analysis to predict potential spread and seasonality for COVID-19. Available at SSRN 3550308 (2020)
9. Abolfazl Mollalo, Behzad Vahedi and Kiara M. Rivera: Gis-based spatial modeling of COVID-19 incidence rate in the continental united states. Science of The Total Environment p. 138884 (2020)
10. Haytham H. Elmousalami and Aboul Ella Hassanien: Day level forecasting for coronavirus disease (COVID-19) spread: analysis, modeling and recommendations. arXiv preprint arXiv:2003.07778 (2020)
11. Difan Zou, Lingxiao Wang, Pan Xu, Jinghui Chen, Weitong Zhang and Quanquan Guk: Epidemic model guided machine learning for COVID-19 forecasts in the united states. medRxiv (2020)
12. M. Bailey, R. Cao, T. Kuchler, J. Stroebel, and A. Wong: Social connectedness: Measurement, determinants, and effects. Journal of Economic Perspectives 32(3), 259 – 80 (2018)

13. Johns Hopkins University and Medicine, Coronavirus Resource Center: Johns Hopkins university and medicine, coronavirus resource center (2020). URL https://coronavirus.jhu.edu/map.html
14. T. Guardian: Facebook friendships can help predict COVID-19 spread, study finds (2020). URL https://www.theguardian.com/world/2020/apr/14/facebook-friendships-can-help-predict-covid-19-spread-study-finds
15. H.S. Maghdid, K.Z. Ghafoor: A smartphone enabled approach to manage covid-19 lockdown and economic crisis. arXiv preprint arXiv:2004.12240 (2020)
16. H.S. Maghdid, K.Z. Ghafoor, A.S. Sadiq, K. Curran, K. Rabie.: A novel ai-enabled framework to diagnose coronavirus COVID-19 using smartphone embedded sensors: Design study. arXiv preprint arXiv:2003.07434 (2020)