Coherence: A Scientific Journal
A Peer- Reviewed Journal of the Science Section
The Bhawanipur Education Society College
Vol.1, Issue 1 (2022) 19-29

A Comprehensive Investigation on Suspected-Infected-Recovered-Deceased Cases of COVID-19 Using Machine Learning Techniques

Priyam Dhar, Sourabh Shreshta, Sayan Bose, Utsab Mukherjee* and Shawni Dutta
Department of Computer Science, The Bhawanipur Education Society College, Kolkata 700020, West Bengal

Abstract

The COVID-19 disease has left a negative mark on medical, social, economical aspects globally. The colossal damage has been done to the upcoming generations. It is highly recommended that the spread of this infectious disease has to be monitored by means of statistical models governed by Machine Learning (ML) based tools. Extensive research needs to be carried out to monitor the spreading of the epidemic. Nevertheless, the mortality rate and the recovery rate are equally significant to justify the current status of the transmission dynamics. This research is primarily aimed to investigate the Suspected-Infected-Recovered-Deceased (SIRD) Cases of COVID-19. To fulfill the objective of this study, recent literature that worked on the SIRD concept using ML is taken into consideration. These existing studies are summarized in terms of their working procedures, and shortcomings. A thorough assessment of the considered studies is carried out to direct future perspectives of the SIRD model.

Keywords: COVID-19, SIRD Model, AI, Machine Learning, Forecasting.

1. Introduction

Coronaviruses are a diverse group of viruses that commonly cause mild to severe upper respiratory tract illnesses like the common cold or flu. The first strain of the virus, known as SARS coronavirus because of the cell′s crown-like structure, was discovered in China in 2002. The epidemic was then followed by the discovery of a mutant version known as MERS in the Middle East. SARS-genetic CoV-2′s information is conveyed by its RNA. The virus′s protein structure allows it to undergo metamorphosis, resulting in a variety of strains. UV light, metals, and endogenous organism components all have an effect on mutation. Based on the immune response and vaccine efficiency, each strain has a unique rate of transmission capability, severity, and even fatality. It is critical to collect data and forecast the strains that are emerging.[1].

It was the 31st of December, 2019, when China notified the World Health Organization (WHO) regarding an unidentified pneumonia case in Hubei Province, China. COVID-19 was characterized as a

* Correspondent Author: email: utsab.mukherjee@thebges.edu.in (U. Mukherjee)

"Pandemic" by the World Health Organization (WHO) on March 11, 2020. These infections are transmitted from one person to another via respiratory droplets. After a period of 2 to 14 days, symptoms including fever, cough, and shortness of breath are observed as the virus′s effects. [2] Human to human contact is thought to be the cause of the disease′s explosive propagation in the population. Hence, it is highly recommended to track the exponential growth of the disease spreading. Mathematical models are a valuable tool for public health officials to use in the fight against epidemics, with the potential to result in large reductions in the number of infected individuals and deaths. Indeed, mathematical models may be used to produce short- and long-term forecasts, allowing decision-makers to maximize possible control tactics including containment, lockdowns, and vaccination programs. Models are also useful for a variety of additional purposes, including estimating transmission parameters, understanding contagion mechanisms, simulating various epidemic scenarios, and testing hypotheses [3].

Without artificial intelligence (AI) and machine learning technology, tackling a global epidemic is difficult to imagine (ML). As a result of the exponential increase in Covid-19 cases around the world, several health infrastructures have been devastated. Institutions, governments, and organizations were able to retaliate using modern technologies. AI and ML, which are frequently thought of as luxuries for high-end living and productivity, have emerged to be life-saving tools in combat against Covid-19 owing to their wide range of applications. Machine learning plays an important role in predicting any event by means of regression or classification. Based on the problem and the dataset different models can be deployed to get the desired analysis.

It is evident that to study the dynamics of any epidemic an extensive statistical analysis is needed. To solve this purpose a handful of research has been conducted using the aforementioned topic. Among them to ensure automation, Machine Learning(ML), Deep Learning(DL), and some other AI-based tools are taken into consideration. Machine Learning (ML) has the potential to identify the underlying associations within enormous datasets. Additionally, ML systems can be trained to categorize changing process conditions in order to represent variances in operating behavior. Machine Learning is nothing but a method by which the computer learns based on some existing data [4]. There are some advantages as well as shortcomings of the ML-based approach hence a newer algorithm was developed that can feed significantly enormous data by means of Deep Learning. Deep Learning is nothing but analogous to the situation of how a human brain identifies any stimulus based on the nerve cells. A similar architecture is used known as the Artificial Neural Network which serves the very same purpose to analyze or predict any situation without any human intervention [5]. It is needless to say that both DL and ML-based models have their respective pros and cons. Based on the dataset it is the role of the researcher to choose which model would suffice the need of the study. In some cases, it has been observed that the amalgamation of some existing models may give more reliable results hence in some cases hybrid models are also preferred.

In this research, several works have been reviewed that used machine learning-based models to predict the dependency of COVID-19 on the aforementioned topics. Based on some existing literature, a brief survey has been conducted that explains the role of machine learning, deep learning as well as hybrid methods to forecast the pandemic using the SIRD models. In a nutshell, the primary objective of the research is stated as follows:

a. To obtain an overview of the existing literature that covers the detection of COVID-19 transmission dynamics.
b. To address the above issue, the studies that incorporate the application of ML, DL, or hybrid models are considered.
c. The overall outline based on the objective, applied dataset and implemented models are summarized in Section 4 .
d. The analysis of the existing literature along with its limitations are provided in Section 5 .

2. Related Works

This section mostly contains recent COVID-19-related literature studies. An in-depth discourse analysis of the current literature was conducted. This analysis may be valuable in advancing pandemic research that is currently underway. We have studied available literature using the terms "COVID-19," "SIRD," “Machine Learning",” Confirmed Cases," and "Mortality." SIRD works are described in depth in this study. In order to have a full grasp of Covid-19 management schemes, scientific databases like Scopus, Web of Science and other Google Scholar papers were analyzed.

An analysis of the trend and pattern of COVID-19 transmission in India is shown in [6]. The trend forecast has been carried out by using Machine Learning algorithms like regression analysis, Support Vector Machine Learning algorithm, and Polynomial regression. They have used the data of the Ministry of health and family welfare of India as their data set. Using these algorithms they concluded that the methods that they had used to predict the rise in COVID-19 are 93% accurate and they can forecast the growth of Covid-19 Cases for the next 60 days. However, the study may have trained models with additional data sets to be able to forecast the trend pattern for more than 60 days [6].

Sarkar, Ovi et al [7] have made extensive research focused on understanding the effects of COVID-19. The authors have collected the dataset from the repository of WHO, and human information and applied different ML-based models like the Polynomial Regression algorithm, Clustering Algorithm, ARIMA model, and Facebook’s Prophet time series forecasting algorithm as well. An extensive comparison among these models was conducted and finally, it is concluded that Facebook′s Prophet model can outperform the other implemented models [7].

It′s been debated whether the number of tests matters in determining a global estimate of the number of persons who have been affected. As a result, Khan, Md H. R., et al [8] focused their research on building a framework using data from the Worldometer website. There seem to be ten variables in total: cumulative confirmed cases, newly confirmed cases, cumulative deaths, cumulative recovered patients, cumulative active cases, cumulative seriously critical patients, the infection rate in million, the death rate in million, cumulative tests conducted, and test rate in millions. To forecast the association between the number of tests and the COVID cases, researchers employed the CART algorithm and the Random Forest algorithm. Only variables including the number of active cases were shown to be significant in the study. They discovered that only variables like total active cases, total deaths, total recovered cases, new cases, and total serious cases are very important variables to forecast the cumulative number of cases of COVID-19 [8].

The motive of [9] is to demonstrate a comparison of machine learning and soft computing models for predicting COVID-19 outbreaks. They acquired COVID-19 patient information from the Worldometer website for five countries: Italy, Germany, Iran, the United States, and China. To forecast the outcome, they used two methods: Multi-layer Perceptron (MLP) and Adaptive Network-based Fuzzy Inference System (ANFIS). Standard epidemiological models have shown inadequate accuracy for long-term prediction due to the high degree of uncertainty and absence of critical data, as per their research. The findings of two ML models, namely, MLP and ANFIS for long-term prediction showed a strong generalization ability. In light of the findings presented in this research, and in light of the very complex nature of the COVID-19 outbreak as well as national disparities, this study implies that machine learning (ML) could be a useful method for modeling the outbreak as a time series. This work serves as an early benchmarking exercise to show how machine learning might be used in future research [9].

The objective of the paper [10] is to find the methods of estimating the numbers of COVID-19 confirmed cases in the country and their implications in the future, using different learning models. They have used Sigmoid modeling, ARIMA, SEIR model, and LSTM machine learning models on the data sets collected from John Hopkins University to predict the numbers of COVID-19 confirmed cases in India. Upon research, they found that these models can be used to predict the numbers of COVID-19 up to 60 days. They came to a conclusion that the SEIR model has a trustworthy background in pandemic analysis and can detect spikes in COVID-19 events and the LSTM model also gave very promising results to other models. Hence, these models would support decision-makers to perceive the upcoming pandemic trajectory in the country and can take necessary actions for better effect of interventions [10].

Another study [11] explains how machine learning and deep learning can be utilized in epidemic research so that transmission can be forecasted and responded to quickly enough to stop or minimize transmission. The dataset of the paper uses the data from the Johns Hopkins dashboard, which consists of a daily case-reported daily time series summary table. This time series includes data for confirmed, death, and recovered cases of COVID-19 with the information of country, province, or state. The update frequency of the dataset was once per day. The models namely, Support Vector Regressor (SVR), Deep Neural Network (DNN), Long-short Term Memory (LSTM), and Polynomial Regression (PR) were trained using the daily case reported data. The result observed that the PR method best fits the growing trend [11].

Sujath, R et al [12] analyzed the spreading characteristics of the epidemic in India. For this purpose, a prediction framework was implemented based on a dataset taken from Kaggle. For the prediction of the spread, some machine learning-based algorithms like LR, MLP, and VAR were used on the dataset. This model showed good results in terms of accuracy. It would have been better if the considered dataset was larger and had more number of observations [12].

Mele, Marco, and Cosimo Magazzino established the correlation between pollution emissions, economic expansion, and Covid-19. Data from January 29 till May 18, in 2020 was considered focussing on confirmed death cases and air pollution accumulation levels for 25 prime Indian cities. Two different methodolies were used to explore the relationship between pollution emissions, economic growth, and most importantly the COVID-19 demise in India. From 1980 to 2018, the authors have performed Stationarity and Toda-Yamamoto causality tests using a time series approach. The findings point to a

one-way relationship between economic growth and pollution. The Oryx 2.0.8 protocol is then implemented in Apache using a D2C algorithm based on proportion-based causality. It is verified by an ML causal link between PM2.5, CO_2, NO_2, and COVID-19 deaths. In order to make the framework more reliable and accurate, a larger dataset should have been used [13].

The study makes use of hybrid ML-based techniques to analyze susceptible-infected-resistant (SIR) cases, with data from Hungary used to demonstrate their potential. The hybrid ML models employed in this investigation are the Adaptive Network-based Fuzzy Inference System (ANFIS) and the Multi-layered Perceptron-Imperialist Competitive Algorithm (MLP-ICA). These will be required to quantify the infected individuals′ time series as well as the fatality rate. The datasets used for this research are mainly online materials from Hungary that are related to the statistical reports of COVID-19 cases. The reports posted between 4th March and 19th April are mainly considered for this study. The traditional SIR-based systems were excellent in handling simple and semi-complex datasets but to handle a complex dataset we need a more robust system. That′s where the idea or concept of the hybrid machine learning models places itself. Specifically for this research the hybrid models mainly used are ANFIS and MLP-ICA. The MLP-ICA generates a platform for forecasting COVID-19 cases and fatality rate whereas the ANFIS facilitates both computing and learning techniques for nonlinear functions. The results of both these models are evaluated and a certain index score is reached. This index score is basically used for estimating the model accuracy. Therefore, based on the promising results of this study, an alternate modeling strategy can be reasoned to handle the complicated phenomena of the COVID-19 outbreak [14].

Using two statistical methods and a deep learning (DL) model, Kim, Meejoung. [15] examines and forecasts the COVID-19 daily case counts using the autoregressive integrated moving average (ARIMA), the generalised autoregressive conditional heteroskedasticity (GARCH), and the stacked long short-term memory deep neural network (LSTM DNN). From December 31, 2019, to February 22, 2021, the experiment uses 10 datasets: nine country datasets and world datasets provided by the WHO. The data size and vaccination were explored based on the performance. The numerical outcomes show that performance is influenced by the dates and immunizations used in the data. It also shows that the LSTM DNN outperforms the two statistical models in terms of prediction. According to the experimental data, LSTM DNN improves mean absolute error by up to 88.54 percent (86.63 percent) and 90.15 percent (87.74 percent) when compared to ARIMA and GARCH, respectively (root mean squared error). While the results of ARIMA and GARCH differ depending on the datasets. The findings of the study might be used as a benchmark for the efficiency ranges and accuracy rate of COVID-19 daily confirmed cases.

The research [16] is concerned with predicting and analyzing COVID-19 by implementing various ML models. The datasets included in this study were collected mainly from the Kaggle website. The disease is characterized by a serious infection spread disease, similar to any other widespread disease. Due to the enormous growth of the number of infections throughout the pandemic, the healthcare infrastructure of many places is struggling to recognize relevant and efficient diagnoses. Supervised ML-based classification tools for this epidemic were implemented in this paper using K-Nearest Neighbor (KNN), Decision Tree, SVM, and also voting-based models like Random Forest models. All of the models

mentioned above are trained on 80% of the data and tested on 20% of the data. The KNN model outperforms all other models used for the analysis with respect to accuracy (98.34%), recall (97%), and F1 score (0.97). SVM, on the other hand, emerged as the most promising model among all models with 97% precision while performing worse in other metrics than KNN [16].

To identify the patients who had a low risk of dying from COVID-19, an effort has been made by Banoei et.al. They have used the data PCR tests of 400 patients who were admitted at the University of Miami Hospital. On this data, they had applied the Statistically Inspired Modification of Partial Least Square (SIMPLS) algorithm to find the results. On research they found that the mortality rate depends on many factors and out of those who have a history of smoking, they have more toxins in their body compared to others so they have more risk of dying than normal patients [17].

M. Mahdavi et al. [18] are interested in developing and comparing prognosis prediction machine learning models based on invasive and noninvasive laboratory, demographic, and clinical data obtained on the day of admission for patients. The information was gathered mostly from the computerised medical records of 628 patients admitted to Masih Daneshvari Hospital between February 20 and May 4, 2020. For maybe the first time in Iranian patients, this research develops and analysed three feasible prognostic models using invasive and non-invasive data since the first day of treatment to estimate COVID19 mortality. Furthermore, the predictive efficacy of non-invasive and invasive feature groups was evaluated across the temporal and feature number spectrum, producing some unexpected results. The non-invasive model surpassed the invasive approach at lower, sparse feature dimensions, indicating that certain non-invasive characteristics have a heavy influx of prediction information. On the other hand, invasive features showed a more dispersed distribution of predictive information. Furthermore, whereas invasive features were good predictors of impending expiry, non-invasive features outperformed them for a longer period.

GROOMS or Group of Optimized and Multisource Selection approach was proposed by Fong S. J and et al. It is nothing but an ensemble of a collection of five groups of predictive approaches. The research was carried out using a polynomial neural network coupled with corrective feedback (PNN-cf). The performance of the framework was checked using certain parameters like Root Mean Squared Error or RMSE. It was evident in the research that PNN-cf showed better performance as compared to the other tools. [19] the research was carried out using a small dataset, but because of the reliability of the PNN-cf architecture, the research forecasted reliable outcomes.

Calafiore GC et al [20] implemented a time-varying SIRD model to study the infection spreading nature in Italy. To explain and forecast the nature of the COVID-19 outbreak in Italy, a parameter varying variant of the model was developed. The research also stated that the models would be able to forecast in the long term as well as short term with significant accuracy. Abdul et al [21] have made a framework that deals with the use of federated learning to predict the recovery and infection rate using a federated learning model. The model was proved to have better efficacy as compared to traditional ML-based models, however, the federated learning-based model is slow as compared to the traditional models.

3. Mathematical models

In the domain of epidemiology, mathematical analysis to predict the count started back in the early 18th Century by Swiss mathematician Bernouli [22]. It was observed during the World War era, that to

analyze the behavior, growth dynamics of any pandemic, a comprehensive relationship between susceptibility, infection, recovered, and deceased must be studied. [23]. It is also obtained in the works of Kermack and McKendrick that the epidemic takes place in Gaussian pattern.

The equations that underlie the relation between S, I, R, D are stated as follows. It is needless to say that the existing ML and DL based models [24] rely on these equations to solve the above mentioned problem.

Let s(t) represents the number of suspected people; i(t) be the function that signifies the number of infected people. The number of recovered people is equally important to analyze any pandemic dynamics so r(t) be the number of recovered people. Besides these the fourth pillar of our model is the death count given by d(t). here every variable – be it suspected count, infected count or recovered cases or even number of deaths are dependent on time. Now, we explain the new count of suspected people as sn(t). The newly infected people count is given by in(t). The most recently recovered people are counted as rn(t) and the new deceased cases are represented as dn(t). The mortality rate is given by η. The rate of the recovered people is given by λ. One of the most significant parameter that has to be kept in mind in number of people coming in proximity or contact to one another due to travel history or social gathering is given by μ. If the total population taken under consideration is N, then the SIRD equations are stated as follows:

$$s_n(t) = -\frac{\mu s(t).i(t)}{N} \quad \dots [3.1]$$

The number of new suspected patient entirely depends on infection rate and previous suspected rate. The probability of getting accurate data depends on the considered population size as well. Equation [3.1] depicts the number of newly suspected people.

$$i_n(t) = \frac{\mu.i(t).s(t)}{N} - (\lambda + \eta).i(t) \quad \dots. [3.2]$$

The recent infection count is given by the equation [3.2]. it primarily depends on people coming into contact and at the same time recovery and mortality rates also contribute a vital part in the aforementioned topic.

$$r_n(t) = \lambda.i(t) \quad \dots. [3.3]$$

The epidemic dynamics is understood best with the recovery count. It is directly proportional to the infection count. However recovery rate due to lockdown and medications also play a crucial role. Equation [3.3] depicts the count of the newly recovered persons.

$$d_n(t) = \eta.i(t) \quad \dots [3.4]$$

Last but not the least, the death count is explained by the equation [3.4] it is directly related to the mortality rate.

4. Discussions and result analysis

Artificial intelligence-based analysis is always an alternate tool for traditional statistical analysis to study any kind of widespread. To forecast the outbreak of an epidemic, the study of suspected-infected-recovered-deceased study is necessary. According to epidemiology [25], it must be taken into account what percent of people are getting affected due to travel history, besides this lockdown and partial lockdowns play an important role as well. The recovered rate must be monitored to get the efficacy of the system. This section briefly summarizes numerous existing literature studies that emphasize the SIRD model as shown in Table 1.

Table 1 : Analysis of existing literature on SIRD model

4.1 Machine Learning-based study

Reference	Findings of the research	Dataset description	Model Description
[6]	The authors of this study evaluated the pattern of COVID-19 infection in India and estimated the number of COVID-positive patients for 60 days with a 93 percent accuracy.	Ministry of health and family welfare of India	Regression analysis, SVM, and polynomial regression
[7]	ML techniques were utilized to forecast the severity of covid-19; the 'prophet time series' model produced the best results.	The data was retrieved from the website of the World Health Organization (WHO) and besides that, human information association has been collected	The Polynomial Regression algorithm, Clustering Algorithm, ARIMA model analysis, and Facebook prophet time series forecasting algorithms were used.
[8]	The primary finding was to check whether the number of tests and the confirmed cases of COVID-19 are dependent or not.	Data was extracted from wordometer website	CART model was used to perform regression
[12]	Predicted the spread of the epidemic	kaggle	LR, MLP, VAR models
[20]	Implemented a time varying SIRD model to study the contagion in Italy.	Github repository	Least squared optimization method has been adopted.
[17]	Prediction of death rate	PCR tests carried out for 400 patients admitted at the University of Miami Hospital	SIMPLS algorithm

4.2 Deep learning based study

Reference	Findings of the research	Dataset description	Model Description
[9]	The nature of the outbreak was predicted by multi-layer perceptron and Adaptive Network-based Fuzzy Inference System	The data was collected from COVID-19 Patients from the Worldometer website for five countries like Italy, Germany, Iran, the USA, and China.	MLP and ANFIS algorithms
[10]	The authors have forecasted the status of the pandemic in India for upcoming 60 days	John Hopkins University - JHU CSSE	Sigmoid modeling, ARIMA, SEIR model, and LSTM machine learning models

4.3 Hybrid and others

Reference	Findings of the research	Dataset description	Model Description
[19]	How small dataset can even provide reliable results for widespread of the disease based on hybrid way of training the dataset.	Chinese archives	polynomial neural network coupled with corrective feedback (PNN-cf)
[21]	The recovery rate and infection rate of the disease was predicted keeping data privacy in mind using a decentralized model.	The datasets from COVID-19 Patients' Chest X-ray radiography were extracted	Federated learning model
[11]	Focused on forecasting the growth of the pandemic, thereby reducing the risks of any potential damage. Polynomial Regression proved to be the best model to predict this problem	Johns Hopkins dashboard	SVR, DNN, LSTM and PR.

5. Conclusions and future directions

The most deadly virus of the past few decades is the novel coronal virus. It spreads at an alarming rate. To monitor the rate of spreading and the mortality rate, traditionally, during the previous epidemics like Spanish flu or the black death statistical analysis was done, however, AI-based models prove to be more promising to ensure satisfactory outcomes. It is quite evident that the infection spread is highly time-dependent so continuous monitoring is recommended for a much better strategic approach to prevent more harm to mankind.

It has been observed that to predict the peak of the pandemic, analysis of the SIRD model is highly required. Some of the works are taken into account where we have reviewed how their accuracy of prediction and some future scopes are stated along with their pros and cons. The common limitation of existing spreading based research articles can be outlined as follows

i. Adequate amounts of data should be incorporated to construct a reliable model in terms of virus spreading pattern prediction. The adequate data basically implies that the data should cover a higher number of COVID testings within a larger time frame. In fact, the models should be efficient enough to modify themselves dynamically so that any abrupt and/or unknown changes in the spreading pattern could be analyzed. Nevertheless, the use of a larger dataset for a significant amount of time period may be beneficial to investigate the impact of different pandemic 'waves'.

ii. COVID has many virus variants and each of them possesses different spreading characteristics. Hence, all the variants along with their infection patterns should be emphasized individually. This is highly recommended as different variants require different medicines and vaccines as potential strategies to combat the infection. Hence, analyzing each variant's spreading pattern will provide an insight to the Government as well as the frontline healthcare providers to take necessary precautions to tackle such pandemic.

iii. Instead of focusing only on either ML or DL-based methods, the SIRD model can even be constructed by using a model that has ML as well DL methods as its constituent. The hybrid ML-DL based method may be employed to ensure more reliability and efficiency.

iv. Use of federated learning may be incorporated into SIRD model prediction to ensure more promising predictive analytics. Very few studies such as [21] have exhibited federated learning for conducting forecasts. As a result, this topic could be investigated for further research. However, this federated learning requires expensive computational power and it is slow as compared to traditional models.

Acknowledgement

We conducted an extensive examination of the SIRD Model to enhance our comprehension of the COVID-19 pandemic. Our gratitude extends to the entire Science Section; without their support, this research endeavour would have been unattainable. Additionally, we express our appreciation to the library (e-repository) of The Bhawanipur Education Society College, enabling us to access relevant journals.

References

[1] Roy, B., et al. "Global variants of COVID-19: Current understanding." *Journal of Biomedical Sciences* 8.1 (2021): 8-11.

[2] Deng, Sheng-Qun, and Hong-Juan Peng. "Characteristics of and public health responses to the coronavirus disease 2019 outbreak in China." *Journal of clinical medicine* 9.2 (2020): 575.

[3] Calafiore, Giuseppe C., Carlo Novara, and Corrado Possieri. "A time-varying SIRD model for the COVID-19 contagion in Italy." *Annual reviews in control* 50 (2020): 361-372.

[4] Lison, Pierre. "An introduction to machine learning." *Language Technology Group* (LTG) 1.35 (2015): 1-35.

[5] Lauzon, Francis Quintal. "An introduction to deep learning." *2012 11th International Conference on Information Science, Signal Processing and their Applications* (ISSPA). IEEE, 2012.

[6] Gambhir, Ekta, et al. "Regression analysis of COVID-19 using machine learning algorithms." *2020 International conference on smart electronics and communication* (ICOSEC). IEEE, 2020.

[7] Sarkar, Ovi, Md Faysal Ahamed, and Pallab Chowdhury. "Forecasting & Severity Analysis of COVID-19 Using Machine Learning Approach with Advanced Data Visualization." *2020 23rd International Conference on Computer and Information Technology* (ICCIT). IEEE, 2020.

[8] Khan, Md Hasinur Rahaman, and Ahmed Hossain. "Machine learning approaches reveal that the number of tests do not matter to the prediction of global confirmed COVID-19 cases." *Frontiers in artificial intelligence* 3 (2020): 561801.

[9] Ardabili, Sina F., et al. "Covid-19 outbreak prediction with machine learning." *Algorithms* 13.10 (2020): 249.

[10] Jakka, Aishwarya. "Forecasting COVID-19 cases in India using machine learning models." *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics* (ICSTCEE). IEEE, 2020.

[11] Punn, Narinder Singh, Sanjay Kumar Sonbhadra, and Sonali Agarwal. "COVID-19 epidemic analysis using machine learning and deep learning algorithms." *MedRxiv* (2020): 2020-04.

[12] Sujath, R. at al, Jyotir Moy Chatterjee, and Aboul Ella Hassanien. "A machine learning forecasting model for COVID-19 pandemic in India." *Stochastic Environmental Research and Risk Assessment* 34 (2020): 959-972.

[13] Mele, Marco, and Cosimo Magazzino. "Pollution, economic growth, and COVID-19 deaths in India: a machine learning evidence." *Environmental Science and Pollution Research* 28 (2021): 2669-2677.

[14] Pinter, Gergo, et al. "COVID-19 pandemic prediction for Hungary; a hybrid machine learning approach." *Mathematics* 8.6 (2020): 890.

[15] Kim, Meejoung. "Prediction of COVID-19 confirmed cases after vaccination: based on statistical and deep learning models." *SciMedicine Journal* 3.2 (2021): 153-165.

[16] Rohini, M., et al. "A comparative approach to predict corona virus using machine learning." *2021 international conference on artificial intelligence and smart systems* (ICAIS). IEEE, 2021.

[17] Banoei, Mohammad M., et al. "Machine-learning-based COVID-19 mortality prediction model and identification of patients at low and high risk of dying." *Critical Care* 25.1 (2021): 1-14.

[18] Mahdavi, M., Choubdar, H., Zabeh, E., Rieder, M., Safavi-Naeini, S., Jobbagy, Z., ... & Kamrani, E. (2021). A machine learning based exploration of COVID-19 mortality risk. *Plos one*, 16(7), e0252384.

[19] Fong, Simon James, et al. "Finding an accurate early forecasting model from small dataset: A case of 2019-ncov novel coronavirus outbreak." *arXiv preprint arXiv:2003.10776* (2020).

[20] Calafiore, Giuseppe C., Carlo Novara, and Corrado Possieri. "A time-varying SIRD model for the COVID-19 contagion in Italy." *Annual reviews in control 50* (2020): 361-372.

[21] Abdul Salam, Mustafa, Sanaa Taha, and Mohamed Ramadan. "COVID-19 detection using federated machine learning." *PLoS One* 16.6 (2021): e0252573.

[22] Bernoulli, D. "Essai d'une nouvelle analyse de la mortalité causée par la petite vérole. Mem Math Phy Acad Roy Sci Paris (1766). English translation entitled 'An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it'In: Bradley L (ed) Smallpox Inoculation: An Eighteenth Century Mathematical Controversy." *Adult Education Department, Nottingham* (1971).

[23] Razzell, Peter. The conquest of smallpox: the impact of inoculation on smallpox mortality in eighteenth century Britain. Caliban Books, 13 The Dock, Firle, Sussex BN8 6NY, 1977.

[24] Fernández-Villaverde, Jesús, and Charles I. Jones. "Estimating and simulating a SIRD model of COVID-19 for many countries, states, and cities." *Journal of Economic Dynamics and Control* 140 (2022): 104318.

[25] Ma, Kevin Sheng-Kai. "Integrating travel history via big data analytics under universal healthcare framework for disease control and prevention in the COVID-19 pandemic." *Journal of clinical epidemiology* 130 (2021): 147.