I.J. Intelligent Systems and Applications, **2022, 3, 40-53**
Published Online June 2022 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2022.03.04

A Comprehensive Review of Machine Learning Techniques for Predicting the Outbreak of Covid-19 Cases

Arpita Santra[1], Ambar Dutta[2]

[1]Amity Institute of Information Technology, Amity University, Kolkata – 700135, India
E-mail: arpitasantra988@gmail.com

[2]Amity Institute of Information Technology, Amity University, Kolkata – 700135, India
E-mail: adutta@kol.amity.edu

Received: 29 September 2021; Accepted: 30 March 2022; Published: 08 June 2022

Abstract: At present, the whole world is experiencing a huge disturbance in social, economic, and political levels which may mostly attributed to sudden outbreak of Covid-19. The World Health Organization (WHO) declared it as Public Health crisis and global pandemic. Researchers across the globe have already proposed different outbreak models to impose various control measures fight against the novel corona virus. In order to overcome various challenges for the prediction of Covid-19 outbreaks, different mathematical and statistical approaches have been recommended by the researchers. The approaches used machine learning and deep learning based techniques which are capable of prediction of hidden patterns from large and complex datasets. The purpose of the present paper is to study different machine learning and deep learning based techniques used to identify and predict the pattern and performs some comparative analysis on the techniques. This paper contains a detailed summary of 40 paper based on this issue along with the use of method they applied to obtain the purpose. After the review it has been found that no model is fully capable of predicting it with accuracy. So, a hybrid model with better training should be employed for better result. This paper also studies different performance measures that researchers have used to show the efficiency of their proposed model.

Index Terms: Forecasting, Epidemic Covid-19, Machine Learning, Models, Performance Analysis.

1. Introduction

Machine learning is a science which promotes the study of computer algorithms such that the system could gain the capability in automatic learning and improve its functionality from past experiences. It aims at the development of computer programs so that it can access data and use it to learn from them. The algorithms are focused on building models from sample data called "training data". Machine learning algorithms are used in various fields – such as predictive analysis, natural language processing, sentiment analysis and many more. The term machine learning was first used by Arthur Samuel in 1959. Supervised, unsupervised and reinforcement learning are the three broad classification of machine learning algorithms. In supervised learning, the algorithms are trained with labels which means that for a given input the corresponding output is known. In unsupervised learning, no labels are provided to the learning algorithm. The machine itself must find patterns in its input. In reinforcement learning, the algorithms learn from a dynamic environment by using its own mode of solution. It uses trial and error methods and learns from its feedback which it tries to maximize. Some of the applications of machine learning include image and speech recognition, product recommendations, traffic prediction, self-driving cars, spam email and malware filtering, virtual personal assistant etc.

Today's world is full of various types of diseases which may be classified into two categories – infectious and non-infectious. Infectious diseases are caused by microorganisms belonging to different classes. These diseases can spread by various ways- air, water, saliva from infected people and even from other eatables. This may lead to mild to severe fever, cough, cold, and diarrhea and in some severe cases death. Majority of deaths in developing countries are caused due to these infectious diseases which are hard to handle, ones spread. Even today medicines are not there for many diseases. We are living in a country with a population more than 1 billion where adequate healthcare facilities are not available. This makes the situation worse. In order to avoid such deadly situations some mechanisms could be applied which will be capable of predicting such epidemics. Machine Learning has shown the ray of hope in this way. Machine Learning helped scientists and researchers in predicting various epidemics and pandemics along with understanding the pathogen

and also, to identify the specific drug to fight with it. It helps in analyzing tons of datasets along with the possible outcomes. Accurate data analysis would make easy detection of diseases and better care of patients. Prediction makes us visualize the upcoming situations and aware us of it. Many infectious diseases are there like – HIV, Ebola virus, Influenza virus and the very recent Coronavirus. Human Immuno Deficiency virus (HIV) was one of the deadly diseases causing great loss to mankind. It transmits through physical contact like – blood transfusion or sexual transmission. The main aim of this virus is to weaken the immune system of the patient along with killing those cells of the body which fights against infection. With time the patient becomes unable to fight any simple virus and meets its death. Similarly, Spanish flu is another example of one such disease. Spanish flu is an unforgettable pandemic with a death rate of nearly one-third of the population of the world. Since it first started in Spain thus called Spanish flu. It was caused due to H1N1 virus. Till today no specific vaccine or antibiotics are present against this flu.

The very recent COVID-19 disease is a communicable disease caused by the discovered Coronavirus. It originated in Wuhan city of China which further spread to all over the world causing mass destruction to mankind. Most people infected with COVID-19 will experience mild to moderate fever, cough and in severe cases it leads to death. The virus that causes COVID-19 is mainly transmitted through droplets generated from an infected person, coughs, sneezes etc. The United States, Brazil, India, Russia and the United Kingdom are among the most affected countries of the world. Thus, it's a high time to employ various machine learning algorithms in this path which can make us aware and help us in fighting such deadly diseases.

The seriousness of the pandemic can be observed with the help of the following statistics. Till now, approximately 23.36 crore Covid-19 cases have been observed throughout the world with 47.80 lakh death and 1.84 crore active cases. USA, India and Brazil are three most affected countries across the globe (Source: https://www.worldometers.info/coronavirus/). The following three figures Fig. 1 and 2 show the daily new cases and daily deaths respectively since January 2020.

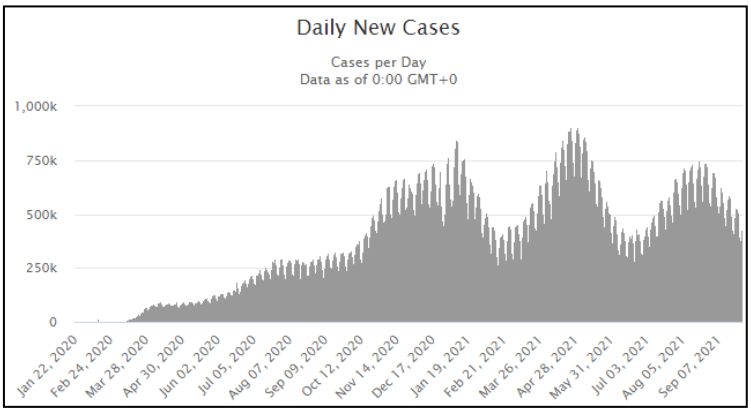

Fig.1. Daily new cases

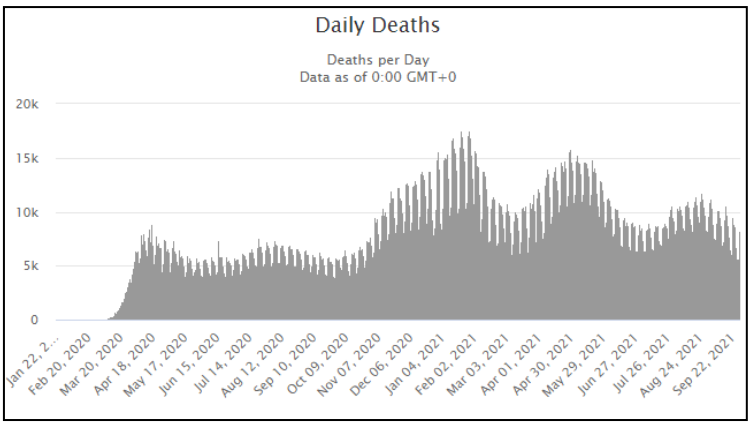

Fig.2. Daily Deaths

A lot of work is being done with the help of machine learning. PathAI's technology uses ML which will enable doctors for more accurate and faster diagnosis. The detection of breast cancer has also facilitated with this approach. The InnerEye project of Microsoft employed ML to detect tumor using 3D radiological images. In the same way we also should employ various prediction techniques which will aware us with upcoming scenarios. So that effective actions could be taken before time and prevent ourselves from such deadly pandemics. In this survey, we will explore various machine learning algorithms and methods used for the prediction of Covid-19 cases. We also worked on a detailed

summary of 40 papers based on this issue along with the use of method they applied to obtain the purpose. At last, we did a comparative analysis on the use of methods and their corresponding outcome.

The rest of the paper is presented as follows: Various models and approaches that are frequently used for predicting Covid-19 using different machine learning methods are given in the Section II. Section III contains the literature review of papers based on this topic. Section IV contains analysis and discussions. Lastly, concluding remarks and future scope of work is provided in Section V.

2. Existing Models Used

A significant amount of work has been done in the field of describing and predicting epidemic/pandemic data over a period of time. Researchers have used various mathematical and statistical models for this purpose some of which are described below:

2.1. SIRD Model

It is an epidemiological model used to predict infectious diseases where the population (size N) is compartmented into four possible states: (S) Susceptible, (I) Infectious, (R) Recovered and (D) Death. The purpose of this model is to estimate basic infection rate, recovery rate and mortality rate. The model has some differential equations which help in meeting its purpose.

$$\frac{ds}{dt} = -\frac{\beta IS}{N}$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \mu I - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

$$\frac{dD}{dt} = \mu I$$

where N=S+I+R+D and γ, μ, ß are recovery rate, mortality rate and infection rate respectively with N as the total population

2.2. SEIR Model

It is also a compartmental epidemiological model mostly used on infectious diseases. The main difference between this model and SIRD model is the way in which they divide the population. For some important infections, there are individuals who, despite being infected, are not infectious during a significant incubation period. These individuals are compartmentalized to Exposed (E) category. This model divides the entire population into: (S) Susceptible, (E) Exposed, (I) Infectious and (R) Removed. The corresponding differential equations related to this model are

$$\frac{ds}{dt} = -\frac{\beta IS}{N}$$

$$\frac{dE}{dt} = \frac{\beta IS}{N} - \sigma E$$

$$\frac{dI}{dt} = \sigma E - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

where γ, ß, σ are the recovery rate, exposed rate and incubation rate respectively with N (=S+E+I+R) as the total population.

2.3. ARIMA Model

Auto-Regressive Integrated Moving Average (ARIMA) is a forecasting model which is a linear regression model which employs its own lags for prediction. This algorithm uses past values of time series to get future values. This model is used in cases where data has non-stationarity either in the form of mean, variance or covariance. It is most often

denoted by ARIMA (p, q, d) where p, q respectively represent order of Auto Regressive and Moving Average term and d controls the level of differencing to make the time series stationary. The mathematical formulation of the ARIMA (p, q, d) model using lag polynomial is given below:

$$\varphi(L)(1-L)^d y_t = \theta(L)\varepsilon_t$$

$$(1-\sum_{i=1}^{p}\phi_i L^i)(1-L)^d y_i = (1+\sum_{j=1}^{q}\theta_j L^j)\varepsilon_t$$

where p, q, d are non-negative integers. ARIMA model, suited for non-stationary, non-seasonal data, is widely used for prediction of economic and stock price data.

2.4. SARIMA Model

Similar to the ARIMA model, SARIMA stands for seasonal ARIMA and is used whenever a seasonal change is suspected. SARIMA model uses seasonal differencing of appropriate order to remove non-stationarity from the series. It is usually denoted by SARIMA (p,d,q)X(P,D,Q)s where "s" denotes the number of periods in each season. For monthly and quarterly time series, the values of s are 12 and 4 respectively. P, Q, D denote the auto-regressive, moving average and differencing terms for seasonal ARIMA.

2.5. ANN Model

ANN (Artificial Neural Network) is a computational model inspired from the structure of biological neural networks. It contains artificial neurons where for each input outputs are obtained which are then passed to other neurons. The neurons are connected with each other through a connection link. It is designed for extracting patterns from noisy data. It is mainly applied on non-linear and high-dimensional data. It accumulates information from patterns and relationships among data and learns from previous experiences rather than programming. Various implementations of ANN are used in different applications including pattern recognition, prediction, classification, modelling etc.

The mathematical expression for this model is

$$Y_{in} = \sum_{i=1}^{m} x_i w_i$$

where x_i (i=1 to m) are the inputs of the neuron and wi (i=1 to m) are the corresponding weight of each connection link. Y_{in} is the net input. The overall output is obtained by passing the net input through the activation function F.

$$Y = F(Y_{in})$$

2.6. RNN Model

RNN (Recurrent Neural Network) belongs to the class of ANN. It consists of nodes and connections between these nodes form directed graphs. It is mostly applied on sequential data. RNNs carry out the same task for each member of a series, with the output being based on the prior computations. In order to predict the next word in a sentence, knowledge of a few words before it is important for any NLP problem. Because of the presence of internal memory, it remembers its input which makes it more suitable for machine learning problems. It behaves similar to the human brain. It has short-term memory. It is then combined with Long-short term memory (LSTM) to overcome it. The equation showing evolution of RNN is given below:

$$O^t = f(h^t;\theta)$$

$$h_t = g(h^{t-1}, x^t, \theta)$$

where O^t is the output at time t, x^t is the input at time t, and h^t is the hidden layer(s) at time t. RNN model is specialized in handwriting recognition, speech recognition.

2.7. NNAR Model

The NNAR (Neural Network Autoregressive) model is inspired from the model of the brain. It is applicable to the time series which are non-stationary. It is a feedforward neural network which combines function with an activation

function. It is denoted by NNAR (p, k) where p and k denote the number of lagged inputs and the number of hidden layers respectively. A NNAR(p, 0) model resembles an ARIMA(p, 0, 0) without ensuring stationarity. Seasonal NNAR is denoted by NNAR (p, P, k). The NNAR (p, P, 0) model is similar to ARIMA (p, 0, 0)(P, 0, 0)m model without ensuring stationarity.

3. Literature Review

Z. Liu et.al [1] focused on the latency period of COVID-19 infection i.e the amount of time newly infected individuals was noninfectious and asymptomatic. This study inspects the asymptomatic, unreported and reported infectious cases in China. They tried to understand how the isolation, public closings and quarantine helped in the less expansion of the pandemic. As per this report, it generally takes 4-5 days to develop a symptom of the infection. They used two models i.e. ODE (ordinary differential equations) focused on infected individuals who are not infectious and DDE (Delay differential equation) which focused on time delay in recently infected individuals before.

S. Contreras et.al [2] aimed to present a multi group SEIRA model to find the spread of COVID-19 among heterogeneous populations. They had applied the model to a population having different characteristics like-different geographical nature of the territory, behavioral and social differences. Since this model had a general approach, it could be used for interaction between subpopulations and may help in better understanding of the evolution of this pandemic. Thus, also help in public-health policies. This model could also be easily modified as per the variations in the populations making it applicable to more populations. It also suggested that more variation can be included in this model and a deeper study on this may be helpful to make strong decisions.

N. Hasan [3] used EEMD technique to produce sub-signals and denoised the original data and used ANN model to train denoised data. Since the data were non-linear and non-stationary it was difficult to predict, so different models like ANN, SVM are used for multi research domain..The ANN model is employed due to its power processing capacity and its estimated capacity in any function up to a good level of precision. The performance of this model was assessed by both Mean Square Error (MSE) and R^2values.Even for accuracy- Regression analysis and moving average strategy are used. This model had performed very well for all types of data with no overfitting as all MSE is consistent.

Patricia Melin [4] focused on temporal aspects like the prediction and forecast in different ways on coronavirus data. While this paper worked on the use of unsupervised self-organizing Kohonen-maps to group together different countries in the world having similar properties and effects of COVID-19. This would help the countries to take similar steps with those countries and fight against this virus. Countries are classified in 4 different classes Very high, High, Medium, Low. Then data of different countries on various aspects like-Confirmed Case, Recovered Case, Death cases were noted and evaluated and similarities were analyzed.

X. Duan [5] et.al focused on predicting the number of confirmed cases of COVID-19. To interpret the data sets and predict the required result, they decided to employ Auto-regressive Integrated Moving Average (ARIMA). Estimation, model identification, and forecasting for the ARIMA model were performed on two time series using the Box-Jenkins technique. The time series' stationarity was determined using the Augmented Dickey-Fuller(ADF) unit-root test. Even the R-package is used to generate ARIMA's numerical output. Finally, the generated model has a 95% confidence level of new confirmed instances per day for a 7-day period.

Z. Ceylan [6] focused on the estimation and prediction of the COVID 19 in Italy, Spain and France. Various statistical methods are used in the prediction like-time series model, multivariate linear regression, grey forecasting model, simulation model and much more. Since the outbreak depends on various factors therefore it was characterized by randomness and tendencies. But the statistical methods were unable to analyze the randomness and are hard to generalize. That is why the ARIMA model was applied for the purpose. It was applied because of its simple structure, fast applicability and ability to explain the data set. Since this model does not use mathematics and statistics which help in easy explanation and understanding. It's famous for its simplicity and systematic structure and acceptable forecasting performance.

S. R. Hanumanthu [7] mainly focused on the prediction of the epidemic using Intelligent Computing like machine learning, deep learning and other computing techniques. Machine Learning, Deep Learning and other techniques are used to predict epidemics like-Smallpox, Ebola virus, Cholera, Swine flu and many more. These techniques had been utilized by the physicians in prediction of imaging modalities in pneumonia. A deep learning model was utilized for detection and localization of pneumonia in chest X ray images since it was built on Mask-R CNN. Deep learning was preferred over machine learning due to greater performance, feature exaction without human interference that to without the use of engineering in the initial phase. Now these intelligent techniques have also been used in COVID 19 case. Several techniques of machine learning like SVM, RF, K-means were used in solving the related issues where DL, CNN are used in deep learning for pandemic prediction.

U. Şahina and T. Şahinb [8] used different types of prediction models like grey model, nonlinear grey model, Bernoulli model and fractional nonlinear grey Bernoulli model for prediction purpose and comparative analysis was done on it. Mean absolute percent error (MAPE), Root mean square error (RMSE) and R square values were used for measuring the performance of the model. With lowest MAPE and RMSE values and high R square value FANGBM gave the best performance. That is the reason this model was highly used by researchers as a forecasting tool. The FANGBM

(1,1) was the combination of nonlinear grey Bernoulli model and the r-th accumulated generation operation. The parameters of these models were fractional order value(r) and power index(γ). If r=1 and γ≠0 then FANGBM converts to NGBM and if γ=0 and r=1 it converts to GM.

C. Anastassopoulou [9] et.al used the SIRD model to find the estimations of the basic mortality rate, recovery rate, reproduction number. The parameters were mainly under consideration. To find the basic reproduction rate, fatality rate and recovery rate they solved the least squares problem using a rolling window. Initially they took 6 days in focus for the 1st estimation, then they added subsequent days in the rolling window with each day at a time. They report the coefficient of determination for each specific window which represents the proportion of variance in the dependent variable from independent variables and root mean square error. Further they forecast the evolution of the outbreak based on previous results. After all this still the accurate results and prediction has not been obtained. Even the unavailability of data or biased data may lead to ambiguous results.

N. C. D. Adhikari et al [10] used twitter for data for a specific country to create a locational network. These data were then integrated with algorithms to detect the outbreak of any diseases. The other approach could be the use of Twitter API to extract the tweets with the name of the epidemic. These then filtered on given criteria such as tweets by patients with the help of SVM. Here twitter data were used. For prediction purposes they used machine learning and various extraction techniques and algorithms for comparison or analysis purposes. The tweet data was transformed into structured from unstructured data converting it into a supervised learning problem. Those tweets are also classified as positive tweets, negative tweets, and neutral tweets. Naive Bayes using TF-IDF gave better results than other methodologies. After examining various tweets and cleaning the data and reducing the errors the required results were gained.

T. Chakraborty et.al [11] worked on forecasting analysis of the dengue epidemic to construct a system for the accurate prediction of dengue cases in Philippines, San Juan and Iquitos. They have used a hybrid approach combining ARIMA and neural network autoregressive models since the real-world data sets contain both linear and nonlinear data. The ARIMA model was used to filter out linear tendencies in data and then the residual data was passed to the NNAR model. The important assumption in this methodology was the additive relationship between the linear and nonlinear component of time series. This model suits for all linear-nonlinear to stationary - nonstationary time series. Hence best in this case. The approach applied to 3 dengue data sets and gave a satisfying result.

T. Ajayia et.al [12] used Random Forest for better interpretation of results and showed better accuracy (68%) compared to other techniques like neural nets (57%) and classification trees (55%). Though random forest has a better performance but its sensitivity value is mediocre. It indicates that the random forest may not be the best option for prediction. Thus future work is needed to bring a suitable model for the upcoming infectious diseases.

A. I. Sabaa et.al [13] mainly focused on the use of statistical and artificial intelligence to build models and forecast the upcoming situation regarding this pandemic in Egypt. They have used the data available by Egyptian ministry of health to train the model. Performance of the NARANN model was better than ARIMA. This model can be utilized for a multi-step forecast for multiple days. As per the study it has been found that cases are estimated to grow by 280% during the month of May. The NARANN has an absolute percentage error (APE) of less than 5% for all the forecasted cases.

M. Şahin [14] worked to find the relation between the weather conditions and its changes as with the pandemic COVID-19 by considering 9 cities of Turkey. Many researchers have worked on it and found that the mean temperature has a linear association with COVID-19 positive cases when the temperature is below 3oC and many more. In this paper temperature, wind speed, dew drop, humidity are the parameters under consideration. Since the incubation period is 14 days, these factors are examined within the interval 3 days, 7 days and 14 days and he used Spearman's correlation for this analysis. Evaluation of the correlation between the temperature and the number of cases in each city revealed that the temperature on the day of cases has high correlation (negative) i.e as temperature decreases, number of cases increases. Similarly, it has found that the higher the wind speed or humidity is, the more number of positive cases. The more the crowd, the more is the chances of contamination.

M. Ahmadi et.al [15] focused on the effects of climatic factors on COVID-19 in Iran. They worked on the 9 factors are-population density, average precipitation, average temperature, humidity, wind speed, average solar radiation, infected people, days of infection and intra-provincial movement. They have used the Partial correlation coefficient and Sobol-Jansen methods in the investigation process. According to the paper infected people, days of infection, intra-provincial movement and population density play a major role and are directly responsible for the spread of the virus whereas the other environmental factors are indirectly. As per the result the humid places have more density of people with viruses than others. However, in arid regions the relationship of humidity and infection is exactly opposite. Even states of Iran with higher populations have more infected people.

R. Salgotra et.al [16] mainly focused on the COVID 19 pandemic, its effect on INDIA especially in three states - Maharashtra, Gujarat, Delhi and the possible prediction with the help of Genetic programming model. GP (Genetic programming) is an advanced form of GA where solutions are generated through computer programs instead of binary strings. GEP (Gene Expression Programming) is the more precise version of GP. This technique has been used by researchers in predicting models. This technique is also used in India for the prediction of the number of confirmed and death cases. The idea behind using this model is its reliability, efficiency than classical techniques and stability relative to ANN (Artificial Neural Network). It generates prediction equations where optimization is possible as the user wishes. In

order to predict the equations this model does not need prior information. Experimental data are used as data in this model instead of basic assumptions, which makes the model more reliable. As for future scope the predicted equations can be derived and the optimization can be done using Krill Herd Algorithm and naked mole- rate algorithm.

During the outbreak of the virus, China relied on AI and used facial recognition cameras for the tracking of infected people with travel history, drones to disinfect roads and public places, robots to deliver food and medicines and so on. AI also focuses on diagnosis of the patients, medical imaging process, diseases tracking and prediction. In some developing countries like India since the resources are limited, so to test the presence of virus X-rays and CT scan techniques can be made into action. Radiologists can perform the test with the help of deep learning. This also helps in monitoring the growth of the virus in lungs of the patients and the effects of the medicine. A. Kumar et.al [17] developed a forecast model based on the XGBoost calculation for the mortality risk prediction. Computational biologists are also working in this field through disease modeling and working on medications. Several applications are also available like "Arogya Setu App" which provides the distance of the user from an infected person with the help of Bluetooth and GPS system.

V. Zarikas et.al [18] worked on Johns Hopkins epidemiological data to cluster countries with respect to active cases, active cases per population and active cases per population per area. Their developed algorithm resulted in consistent and reasonable clustering. The code was in SPSS and Mathematica. The overall project follows the concept of hierarchical clustering. They used Euclidean distance between different time series. The failure of other known algorithms may encounter due to various reasons like difference in lengths, orders of magnitude and many more. Thus, at the end it has been found that small countries are in the most critical situation like Monaco, San Marino, Liechtenstein, Malta and so on.

Z. Zhao et.al [19] aimed to use modelling to estimate the spread of the epidemic in African countries under various epidemic scenarios and to propose a set of epidemic prevention and control measures that will be critical in restricting the spread of the epidemic in these countries. South Africa, Egypt, Algeria, Nigeria, Senegal, and Kenya are their key countries. To estimate the parameters of epidemic spread in each country, the MH parameter sampling optimization technique is utilized. According to the findings, the outbreak can be controlled by late April if the scenario is strictly controlled. The number of infected people will increase by 1.43–1.55 times under moderate control, and the date of the epidemic will be postponed by around 10 days. In case of weak control, the epidemic will be controlled by late May, the total number of infected cases will double.

M. Li et.al [20] built models to predict the numbers of cumulative confirmed cases (CCCs), new cases (NCs), and death cases daily in China. They also predicted the pandemic trend within and outside China. They used Eureqa to train the model so that it could predict the CCCs which in turn helps in predicting NCs and DCs. The epidemic was expected to peak around May 22, 2020, and be under control by February 21, 2020, according to the simulations. One weakness of this study is that it did not consider aspects other than demographics in developing predictive models for worldwide cases, such as politics, economy, and culture, which play a role in the transmission and outbreak of COVID-19.

Z. Liu et.al [21] focused on the conditions of people and the growth of infection in the Wuhan city of city. They used the SIRU model, in which S(t) represents the number of people susceptible to infection at time t, I(t) represents the number of asymptomatic infectious people at time t, R(t) represents the number of reported symptomatic infectious people at time t, and U(t) represents the number of unreported symptomatic infectious people at time t. The model incorporates pandemic characteristics such as (1) the importance of the timing and magnitude of major government public restrictions designed to reduce the severity of the epidemic; (2) the importance of both reported and unreported cases in interpreting the number of reported cases; and (3) the importance of asymptomatic infectious cases in disease transmission.

Shreshth Tuli et.al [22] used both machine learning and Cloud computing to prepare a model for prediction purposes. They developed a more accurate mathematical model to study and forecast the epidemic's spread. A machine learning-based enhanced model was used to forecast the potential hazard of COVID-19 in countries throughout the world. They used the data and put it on a cloud computing platform to make more precise and real-time predictions about the epidemic's growth characteristics. The model was also statistically efficient because to the adoption of a Robust Weibull model based on iterative weighting.

Kim Tien Ly [23] predicted the number of COVID-19 cases in the United Kingdom using Adaptive Neuro-Fuzzy Inference System (ANFIS) model to train the data collected. He worked on various factors of ANFIS to build a successful time-series prediction model. Neural Network has been used since it performs fine in case of long-term forecasting after training. Fuzzy modeling is the method of changing values with vagueness. This modeling was then implemented by Fuzzy Inference System (FIS) with a set of If-then rules on input-output system. This model worked well in the prediction of COVID-19 cases in Spain and Italy.

Farooq and Bazaz [24] predicted the spread of Covid-19 in different states of India using a combination of various models like Artificial Neural Network, Deep Learning and other models. They used Deep Learning to propose an ANN based online incremental learning technique to estimate various parameters of Covid-19. The main advantage of this model was its adaptability. Unlike other Deep Learning techniques, this model had a great power in adapting with the new data without any retraining or rebuilding from scratch. They chose the five most affected states of India and got a series of predictive values as per the states. As a future scope this model can also be used in various countries to get a satisfactory result.

Sahai et al [25] focused on the time series data of top 5 affected countries – USA, Brazil, India, Russia and Spain –

of the world by Covid-19 to predict the spread of this epidemic. They used ARIMA model for the prediction. The Hannan Rissanen algorithm was also used by them. They measured the forecast efficiency using mean absolute percentage error (MAPE) and mean absolute deviation (MAD). The proposed model predicted quite well and is preferable for diseases outbreak modeling. However, the ARIMA technique had a major drawback in forecasting and characterizing the time series model.

Appadu, Kelil and Tijani [26] mainly focused to know the total number of infected individuals and active cases of Covid-19 in South Africa, India, Germany, South Korea and Italy. Though this pandemic originated in the Wuhan city of China but later it spread worldwide killing lakhs of people all over the world. They employed different forecasting techniques like Euler's iterative method, spline interpolation and a hybrid Euler method for this purpose. They divided the methods into short, medium, long-term prediction. Cubic spline interpolation is useful for short- and medium-term prediction. In the same way the Euler method also worked better for medium- and long-term predictions. Thus, a fruitful result had been obtained from those methods. And as per the future scope these techniques could be further used for the other virus propagation.

Petropoulos, Makridakis and Stylianou [27] aimed at forecasting the number of confirmed and death cases of Covid-19. They proposed a short time forecasting model to predict the confirmed and death cases. They focused over a time period of 4 months and also worked on its accuracy and usefulness. They compared the performance of the model with other available forecasts and results were positive. Later they provided the forecast of 3 different countries – Denmark, Norway and Sweden. Thus, the study suggested that the method is consistent in forecasting. The limitations of this model where they did not consider the actions taken by the government. Their model will work only when things are stable as the model is univariate. Another drawback was they worked on short-term predictions which would help the policy-makers to take necessary steps.

Balli [28] analyzed the data related to Covid-19 and predicted the number of cases in different countries using various algorithms of machine learning. He used a time series prediction model to achieve various curves of the disease and to know the tendency of the pandemic. He took all the data of different countries from the World Health Organization. He also employed various machine learning algorithms like linear regression, support vector machine, multi-layer perceptron and random forest. The performance of the model was evaluated by mean absolute percentage error (MAPE) and root mean square error (RMSE). Experimental results revealed that Support Vector Machine (SVM) has the best fit. A further study on this is important to obtain more valuable results and figures of this pandemic.

Zeroual et al [29] conducted a comparative analysis of different Deep Learning methods to predict the new and recovered cases of Covid-19. They used various Deep Learning techniques like simple Recurrent Neural Network (RNN), Long short-term memory (LSTM), Bidirectional LSTM (BiLSTM), Variational AutoEncoder (VAE) and Gated recurrent units (GRUs) algorithms. They mainly considered 6 countries and their data – Italy, Spain, China, USA, Australia and France. They applied different techniques on the data available to them and checked the performance using mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE) and some more. From this comparative study they concluded that Variational AutoEncoder (VAR) had achieved the best forecasting results than other models. Hence this could be further used by the government and policyholder to take appropriate steps.

Elsheikha et al [30] aimed at predicting the figures of Covid-19 in Saudi Arabia using Deep Learning. They proposed a long short-long term memory network to predict the total number of confirmed, recovered and deaths in Saudi Arabia. They trained the model using official data and verified using root mean square error (RMSE), mean absolute error (MAE), coefficient of variation (COV), efficiency coefficient (EC), coefficient of residual mass (CRM), coefficient of determination (R2) overall index (OI). They predicted accuracy was then compared with two other models – ARIMA and NARANN (Nonlinear Auto-Regressive Artificial Neural Network). Then the forecasting model was also applied on 6 other countries – India, Brazil, South Africa, Saudi Arabia, Spain, and the USA. A major advantage of this model was the presence of feedback connection which allows the movement of data signals in backward directions thus increasing the accuracy level. This model was also capable of learning the non-linearity from training data.

Al-qaness et al [31] aimed to forecast the Covid-19 in hotspot regions. Covid-19, a viral disease which came up in December 2019 in the Wuhan province of China and later spread to all over the world creating huge destruction. Here they proposed a short-term forecasting model based on a more advanced adaptive neuro-fuzzy inference system (ANFIS). To improve the model's outputs and eliminate faults, a better version of the marine predator's algorithm (MPA) called chaotic MPA (CMPA) is used. They also compared the model to other AI models such as the original ANFIS, as well as improved variants of ANFIS that used the marine predator's algorithm and particle swarm optimization. Other statistical assessment criteria were used to examine the model's correctness, which yielded positive findings.

Gomes da Silvaa [32] focused on the number of new cases of Covid-19 in Brazil and the United States of America. They employed various machine learning algorithms like k-nearest neighbors, Bayesian regression neural network and support vector regression, cubist regression, quantile random forest for the prediction purpose. Along with these algorithms they also used the pre-processing variational mode decomposition (VMD) which converts the time series to different mode functions. The capability of this model was calculated based on performance criteria. The hybrid VMD had a greater result than all other forecasting models in terms of accuracy. It was also shown that different temperature; precipitation also had a major effect on the outcomes.

Kalantari [33] examined the advantages of Singular Spectrum Analysis (SSA) to predict the number of daily

confirmed cases, death cases and recovered cases of Covid-19. He proposed an algorithm to calculate the various parameters of SSA which includes window length and all other leading components. He then compared the obtained results of R-SSA and V-SSA with other forecasting techniques like Autoregressive Integrated Moving Average (ARIMA), Neural Network Autoregression (NNAR), Fractional ARIMA, Exponential Smoothing and TBATS. He selected the best model upon measuring the Root Mean Square Error (RMSE). Thus, the forecast from the model could be helpful for taking necessary actions.

Castillo and Melin [34] forecasted Covid-19 using the hybrid approach of both factual dimension and fuzzy logic. Here they employed a combined approach of fractal dimension and fuzzy logic. The concepts of factual dimension were used to determine the complexity of the dynamics in time series. Fuzzy logic was employed to present the uncertainty in the forecast process. They used the local datasets available of the 10 countries to build the model. The forecasted values were then rechecked with the actual values to determine its reliability. It gave an accuracy of 98%. This approach could be considered as a successful attempt and suggested to use by the government for taking decisions. As for future approaches this model could be used on other similar types of cases and other different concepts of fuzzy logic could be used. They also suggested using granular computing which also had the capacity of handling uncertainty.

Mojjada et al [35] focused on the prediction of the number of individuals infected from Covid-19 using Machine Learning modelling. Their analysis was based on supervised regression models, LASSO and exponential smoking (ES) models. When these techniques were applied on the datasets different models had different outputs. Those outputs were then evaluated on various parameters like-R-squared score, Adjusted R-squared score, mean absolute error (MAE), Mean square error (MSE), Root mean square error (RMSE). The result thus showed that ES gave better results. LR and LASSO had shown effectiveness in predicting the death rate to some degree. In all the Linear Regression had shown better results than other models. And these outcomes could be used by the policymakers to take necessary steps.

Maher, Majdalawieh and Nizamuddin [36] predicted all the details related to Covid-19 using hybrid models. Covid-19 occurred in the city Wuhan of China creating huge destruction which then later spread to different parts of the world killing lakhs of people. Here they used the hybrid model which combined the approach of both SEIRD (S-Susceptible, E-Exposed, I-Infected, R-Removed, D-Death) and ARIMA. Initially they estimated SEIRD model parameters i.e infected, recovered and deceased population using historical data for a perfect fit. Then they calculated the results of SEIRD with actual data. They trained three ARIMA models based on different parameters. It helped in the accuracy of prediction. The results were then checked using Mean Average Error (MAE), Mean squared error (MSE), Mean squared logarithmic error, Normalized mean average error and some others. In all, the outcomes of the model were quite satisfactory. A limitation of this model was that they did not consider the transition of population from one compartment to another.

Khana, Saeed and Ali [37] proposed a multivariate time series model called Vector Autoregressive model to forecast the number of new cases, recovered and death cases of Covid-19 by using time series model in Pakistan. Since different other time series models are available AR, MA, ARIMA but they used Vector Autoregressive because they considered 3 dependent variables which was not possible in other models. Estimation of the parameters was done using ordinary least square (OLS). They used histogram, ACF, PACF of the fitted model for checking the residuals. They also performed serial correlation test and normality test for statistical testing. They evaluated the model for ARCH (Autoregressive Heteroscedasticity) error and succeeded. This model had shown a 95% confidence level for various parameters and hence could be used for the prediction purpose.

Yousaf et al [38] used the time series ARIMA model to predict and forecast the number of confirmed cases, recoveries and death cases of Covid-19 in Pakistan. Since the ARIMA model had a higher accuracy rate and fitting capacity, it was better than other exponential smoothing. They used the statistical approach AIC for model selection and assessment. The outcomes gave a 95% confidence level. Thus, they suggested the use of this approach to take needed steps. One major disadvantage of this study was that they took some assumptions but if those things did not put in place the outcomes would be of no use. Another problem was the lack of availability of data also accepted that an amount of uncertainty was present in the prediction.

Chimmula and Zhang [39] used Deep Learning based on Long Short-term memory (LSTM) networks for forecasting Covid-19 in Canada using Deep Learning. They used the Canadian dataset to train the model. The validation of this model was confirmed by using root mean square error (RMSE). The outcomes of the model showed the capability to capture the transmission with minimum loss. The LSTM model was different from other models as the LSTM network exactly fits the real-time data without any assumptions. This model worked better and reduced the uncertainty. Thus, this study could be helpful for the citizens of Canada. This model could be used to predict the numbers in other countries.

Shastri et al [40] presented a comparative analysis of the data of India and USA of Covid-19 using Deep Learning models. They proposed a model using recurrent neural networks (RNN) based on various long short-term memory (LSTM) like Stacked LSTM, Convolutional LSTM and Bi-directional LSTM to design and predict the numbers. As per the graphs and histological diagrams it could be concluded that the performance of Convolutional LSTM had performed better than the other two LSTM while the performance of Stacked LSTM was the worst. The outcome of the model was evaluated with other parameters like Mean Absolute Percentage Error (MAPE). Thus, this approach could be used by the policymakers to take necessary steps.

Renald et al [41] developed a deterministic model to assess the impact of various rabies management approaches in metropolitan areas with animals. The authors considered the population of household dogs, stray dogs, and Maasai dogs

in their model. For rabies transmission control, the model included the contributions of vaccination, culling, and their combination.

For the transmission dynamics of COVID-19, Abriham et al [42] developed a mathematical model (dynamic SEQISINHR model). The influence of prevention and control techniques was studied by the writers. They created a self-protection parameter to examine the effects of physical separation, staying at home, wearing masks, and cleaning hands, among other things. Finally, they concluded that increasing the pace of isolation and quarantine is the best way to control the disease.

4. Analysis and Discussion

4.1. Performance Measures

Researchers have used a number of performance measures to evaluate the performance of their proposed models, some of which are mentioned below:

If we denote Y_t and Z_t for the observation at time t and its forecasts for n observations time-series data, then the forecasting error e_t is $(Y_t - Z_t)$.

Mean Absolute Error (MAE) = $\frac{1}{n}\sum_{i=1}^{n}|e_t|$

Mean Square Error (MSE) = $\frac{1}{n}\sum_{i=1}^{n}e_t^2$

Root Mean Square Error (RMSE) = $\sqrt{MSE}$

Mean Absolute Percentage Error (MAPE) = $\frac{1}{n}\sum_{i=1}^{n}\frac{e_t}{Y_t}$

Symmetric MAPE (sMAPE) = $\frac{1}{n}\sum_{i=1}^{n}\frac{2|e_t|}{|Y_t|+|F_t|}$

Unscaled Mean Bounded Relative Absolute Error (UMBRAE) is a new measure which combines the good features of different performance measures to address some issues in them. UMBRAE is defined as follows:

UMBRAE = $\frac{MBRAE}{1-MBRAE}$

with MBRAE = $\frac{1}{n}\sum_{i=1}^{n}\frac{|e_t|}{|e_t|+|E_t|}$

where E_t denote the forecasting error using some benchmarking method.

4.2. Discussion

Now-a-days, the outbreaks of the pandemic Covid-19 is a worldwide matter of concern. Thus, in order to restrict the spread of the disease there is a serious requirement of the proposal of efficient and accurate prediction model. In this paper a large number of contributions in the field have been studied, and only a representative set of techniques is discussed. From the study, it is revealed that most of the researchers utilized machine learning and deep learning based techniques to predict from large and complex datasets.

Researchers in the field have used compartmental model like SIRD, SIER models, ARIMA and SARIMA models, ANN, CNN, RNN and NNAR based models. However, most of the researches applied hybrid models considering the merits of various models and obtained a good accuracy in predicting the Covid-19 cases. However, it has been observed from the studies that for the prediction of the outbreak of the pandemic, one of the most popular models used were various auto-regression models including ARIMA and SARIMA models which are followed by different machine learning and deep learning techniques including support vector machine, k-nearest neighbors, and other neuro-fuzzy techniques. A brief summary of 10 representative set of models is given in Table-I. The table includes a brief description of each paper including the techniques used, input data taken, performance metrics used and results obtained. It is further observed from the study that most of the models gave an accuracy of more than 90% in the prediction of Covid-19 outbreaks with the help of different performance measures. However, since different studies focused on different parts of the globe, there will be a requirement of tuning of the parameters for some of the models.

5. Conclusion

Machine learning algorithms for the prediction of epidemic/pandemic diseases has been among the most popular

research interests today which will help the healthcare department to make correct medical decisions. As the entire world is badly suffering from the outbreak of Covid-19, researchers have proposed huge number of machines learning based approaches for the prediction and forecasting of the Covid-19 cases. This survey shows how machine learning and artificial intelligence methods are utilized to predict Covid-19 outbreaks. It presents an extensive literature review of a number of approaches for the prediction of Covid-19 outbreaks around the world using machine learning and includes some comparative performance analysis of some of the representative works.

Table 1. Comparative Analysis of Some Selected Models

Reference	Purpose	Technique Used	Input	Performance Metrics used	Result
[23]	Predicting the number of COVID-19 cases in the United Kingdom	Fuzzy logic and Artificial neural network (ANN)	Training details came from Github Center for Systems Science and Engineering (CSSE) of Johns Hopkins University.	Total Error Rated Rate (UMBRAE)	The study designs a final model with a better approach among ANFIS comparable items. If UMBRAE <1, the proposed model works (1-UMBRAE) * 100% better than the benchmark method. If UMBRAE> 1, the system is almost (UMBRAE-1) * 100% worse than the benchmark method.
[36]	To predict all the details related to Covid-19	SIRD and ARIMA	Data from COVID tracking project (COVID Tracking Project, 2020)	Medium Average Error (MAE), Square Mean Error (MSE)	The research designs a model combining the properties of SIRD and ARIMA. The worst and best cases differ from the last estimate of 5.7% of people infected with the virus, 9.2% of people who have recovered, and 12.7% of people who die.
[25]	Forecasting of Covid-19 in US, Brazil, India, Russia and Spain	ARIMA and Hannan Rissanen algorithm	Data obtained from online database	Mean absolute Deviation (MAD) and Mean absolute percentage error (MAPE)	Models designed to predict Covid-19 cases with ARIMA.MAD was the lowest in Spain followed by Russia. The MAPE of India, Brazil and the US were 3.701%, 1.844% and 2.885% respectively. It was significantly lower in Russia and Spain at 1.090% and 0.832% indicating the accuracy of the solid weather forecast. The results showed that the situation was better for Russia and Spain whereas adverse for India, US and Brazil.
[27]	Forecasting the number of confirmed and death cases of Covid-19	ARIMA	Data per country obtained from the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (2020)	Absolute percentage error (APE), Mean absolute percentage error (MAPE),	A short-term forensic prediction model for Covid is proposed. In 2020-05-21 they predicted that there was a 5% chance that the confirmed cases would exceed 15 million while the confirmed cases by the end of 2020-05-20 were 5 million.
[28]	Predict the numbers of COVID-19 in different countries	Multi-layer perceptron and Random forest, Support Vector, Logistic Regression,	COVID-19 data between 20/01/2020 and 18/09/2020 for USA, Germany and the world is available on the World Health Organization website.	Mean absolute percentage error (MAPE), Root mean square error (RMSE)	Data from USA, Germany was analyzed to predict Covid-19 and used various machine learning algorithms.The SVM method provides better data performance worldwide, Germany and USA compared to other algorithms.
[29]	A comparative study of different deep learning ways to predict new and returning cases of Covid-19	Deep learning strategies such as Recurrent Neural Network (RNN), short-term memory (LSTM), Bidirectional LSTM (BiLSTM), agated recurring units (GRUs) and Variational AutoEncoder algorithms (VAE)	The data sets were made public by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (https://github.com/CSSEGISandData / COVID-19 accessed 17/06/2020).	Root mean square error (RMSE), Mean absolute error (MAE), Mean absolute percentage error (MAPE)	It focuses on the data of a fixed time series of 6 major countries in Italy, Spain, China, USA, France and Australia. The VAE method provides a better prediction of COVID-19 certified cases compared to other models considered in almost every conceivable country except Italy. The VAE model achieved MAPE values of 5.90%, 2.19%, 1.88%, 0.128%, 0.236%, and 2.04% in COVID-19 in 6 countries.

[30]	Predicting the number of Covid-19 cases in Saudi Arabia	Deep Learning techniques- long short-long term memory network	The official reported COVID-19 confirmed cases, recovered cases, and deaths by the Saudi ministry of health (https://covid19.moh.gov.sa/)	Efficiency coefficient (EC), coefficient of variation (COV), and coefficient of residual mass, coefficient of determination (R2) overall index (OI), Root mean square error (RMSE), mean absolute error (MAE),	Deep learning methods were used to predict Covid-19 cases in Saudi Arabia. The proposed LSTM model successfully predicted the number of cases in the previous week with better accuracy compared to NARANN and ARIMA. The RMSE data predicted using LSTM was less than 11 and 28% of those of ARIMA and NARANN, respectively. The LSTM model also worked in other countries such as Brazil, India, Saudi Arabia, South Africa, Spain, and the USA.
[31]	Predicting Covid-19 in tropical regions of Russia and Brazil	Adaptive neuro-fuzzy inference system (ANFIS)	The official cases of COVID-19 reported in Russia and Brazil declared by the WHO from March 26 to June 1, 2020 were adopted.	Mean absolute percentage error (MAPE), mean absolute error (MAE), and Root mean square relative error (RMSRE), Root mean square error (RMSE),	Russia and Brazil data were used to train models ANFIS, PSO, MPA, and CMPA. The predicted CMPA results have a slightly lower percentage error related to that of ANFIS, PSO, and CMPA, indicating higher performance of CMPA performance than other investigated models. Among all the investigated models, CMPA has the lowest RMSE, MAE, MAPE, and RMSRE values of 833, 667, 0.22, and 0.0024 in the case of Russia and 1407, 1073, 0.30, and -0.004 in the Brazilian state.
[32]	Predicting the number of new Covid-19 cases in Brazil and the United States of America	Machine learning algorithm like- k-nearest neighbors, and support vector regression, cubist regression, quantile random forest, Bayesian regression neural network.	In the Brazilian context, the data set was from the API (Application Program Interface) and the USA context, the data was from the "COVID-19 Data Repository" on Github provided by the Center for Systems Science and Engineering (CSSE) in -Johns Hopkins University.	Improvement percentage (IP) index, symmetric mean absolute percentage error (sMAPE), and relative root mean squared error (RRMSE)	The machine learning methods developed by BRNN, CU-BIST, KNN, QRF, and SVR, as well as the VMD method, were used in the forecasting process. The CU-BIST compliant VMD model are the ideal tools for predicting COVID-19 events. Level of models in all Brazilian regions VMD – CUBIST, VMD-BRNN, SVR, CUBIST, VMD – SVR, BRNN, VMD – QRF, QRF, VMD – KNN, and KNN, and in the USA are -VMD – CUBIST. , BRNN, CUBIST, SVR, VMD – BRNN, VMD – SVR, VMD– QRF, QRF, KNN, and VMD – KNN.
[33]	To evaluate the benefits of Singular Spectrum Analysis (SSA) to predict daily value of verified cases, death cases and Covid-19 conviction	Forecasting techniques like Neural Network Autoregression (NNAR), ARIMA, Fractional ARIMA, Exponential Smoothing and TBATS	Database of Center for Systems Science and Engineering (CSSE) at Johns Hopkins University	Root Mean Square Error (RMSE)	SSA model hired to predict Covid-19 cases hired. R-SSA and V-SSA predictive strategies surpass other models ARIMA, ARFIMA, ETS, TBATS, and NNAR. There will be a rapid increase in the number of certified cases in France, Spain and the UK but the rate will slow down in Russia and Argentina.

It is evident from the survey that a number of machine learning approaches have been proposed by the researchers using ARIMA, SARIMA, SVM, k-NN etc for the prediction and forecasting of Covid-19 cases since December 2019. However, there is a need for more advanced machine learning techniques which may successfully deal with large and complex datasets for the prediction of hidden patterns. Another challenge which the researchers are facing is the missing and inaccurate data. The future scope of the present research work is to propose an advanced and hybrid machine learning model that will be able to predict Covid-19 cases successfully, with a special focus on India.

References

[1] Z. Liu, P. Magal Ousmane Seydi and G. Webb, A COVID-19 epidemic model with latency period, Infectious Disease Modelling (2020), doi: https://doi.org/10.1016/j.idm.2020.03.003

[2] S. Contreras, H. Andrés Villavicencio, D. Medina-Ortiz, J. Pablo Biron-Lattes, Á. Olivera-Nappa, A multi-group SEIRA model for the spread of COVID-19 among heterogeneous populations, Chaos, Solitons and Fractals (2020)

[3] N. Hasan, A Methodological Approach for Predicting COVID-19 Epidemic Using EEMD-ANN Hybrid Model, Internet of Things (2020), doi: https://doi.org/10.1016/j.iot.2020.100228
[4] P. Melin, J. Cesar Monica, D. Sanchez, O. Castillo, Analysis of spatial spread relationships of Coronavirus Pandemic in the world using self organising maps, Chaos, Solitons and Fractals (2020)
[5] X. Duan and X. Zhang, ARIMA modeling and forecasting of irregularly patterned COVID-19 outbreaks using Japanese and South Korean data, Data in Brief (2020), doi: https://doi.org/10.1016/j.dib.2020.105779
[6] Z. Ceylan, Estimation of COVID-19 prevalence in Italy, Spain, and France, Science of the Total Environment 729 (2020)
[7] S. R. Hanumanthu, Role of Intelligent Computing in COVID-19 Prognosis: A State-of-the-Art Review, Chaos, Solitons and Fractals (2020), doi: https://doi.org/10.1016/j.chaos.2020.109947
[8] U. Şahina and T. Şahinb, Forecasting the cumulative number of confirmed cases of COVID-19 in Italy, UK and USA using fractional nonlinear grey Bernoulli model, Chaos, Solitons and Fractals (2020)
[9] C. Anastassopoulou, L. Russo, Athanasios Tsakris, Constantinos Siettos,Data-based analysis, modelling and forecasting of the COVID-19 outbreak, PLoS ONE 15(3): e0230405. https://doi.org/10.1371/journal.pone.0230405
[10] N. C. D. Adhikari, A. Alka, V. Kumar Kurva, S. S Hitesh Nayak, K. Rishav, A. Kumar Nayak, S. Kumar Nayak, V. S. Karthikeyan, S. Nayak, Epidemic Outbreak Prediction Using Artificial Intelligence, International Journal of Computer Science & Information Technology, Vol 10, No 4, August 2018
[11] T. Chakraborty, S. Chattopadhyay, I. Ghosh, Forecasting dengue epidemics using a hybrid methodology, Physica A 527 (2019)
[12] T. Ajayia, R. Darab, Z. Poljaka, Forecasting herd-level porcine epidemic diarrhea (PED) frequency trends in Ontario (Canada), Preventive Veterinary Medicine 164 (2019)
[13] A. I. Sabaa, A. H. Elsheikhb, Forecasting the prevalence of COVID-19 outbreak in Egypt using nonlinear autoregressive artificial neural networks, Process Safety and Environmental Protection 141 (2020)
[14] M. Şahin, Impact of weather on COVID-19 pandemic in Turkey, Science of the Total Environment 728 (2020)
[15] M. Ahmadi, A. Sharifi, S. Dorosti, S. Jafarzadeh Ghoushchi, N. Ghanbari, Investigation of effective climatology parameters on COVID-19 outbreak in Iran, Science of the Total Environment 729 (2020)
[16] R. Salgotra, M. Gandomi, A. H Gandomi, Time Series Analysis and Forecast of the COVID-19 Pandemic in India using Genetic Programming, Chaos, Solitons and Fractals (2020), doi: https://doi.org/10.1016/j.chaos.2020.109945
[17] A. Kumar, P. Kumar Gupta, A. Srivastava, A review of modern technologies for tracking COVID-19 pandemic, Diabetes & Metabolic Syndrome: Clinical Research & Reviews 14 (2020)
[18] V. Zarikas, S. G. Poulopoulos, Z..Gareiou, E. Zervas, Clustering analysis of countries using the COVID-19 cases dataset, Data in Brief (2020), https://doi.org/10.1016/j.dib.2020.105787
[19] Z. Zhao, X. Li, Feng Liu, G. Zhu, C. Ma, L. Wang, Prediction of the COVID-19 spread in African countries and implications for prevention and control: A case study in South Africa, Egypt, Algeria, Nigeria, Senegal and Kenya, Science of the Total Environment 729 (2020)
[20] M. Li, Z. Zhang, S. Jiang, Q. Liu, C. Chen, Y. Zhang, X. Wang, Predicting the epidemic trend of COVID-19 in China and across the world using the machine learning approach, medRxiv preprint doi: https://doi.org/10.1101/2020.03.18.20038117
[21] Z. Liu, P. Magal, O. Seydi, G. Webb, Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data, arXiv:2002.12298v1 [q-bio.PE] 27 Feb 2020
[22] S. Tuli , S. Tuli, R. Tuli, S. S. Gill, Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing, https://doi.org/10.1016/j.iot.2020.100222
[23] K. Tien Ly, A COVID-19 forecasting system using adaptive neuro-fuzzy inference, Finance Research Letters, https://doi.org/10.1016/j.frl.2020.101844
[24] J. Farooq, M. A. Bazaz, A deep learning algorithm for modeling and forecasting of COVID-19 in five worst affected states of India, https://doi.org/10.1016/j.aej.2020.09.037
[25] A. K. Sahai, N. Rath, V. Sood, M. P. Singh, ARIMA modelling & forecasting of COVID-19 in top five affected countries, https://doi.org/10.1016/j.dsx.2020.07.042
[26] A. R. Appadu, A. S. Kelil, Y. O. Tijani, Comparison of some forecasting methods for COVID-19, https://doi.org/10.1016/j.aej.2020.11.011
[27] F. Petropoulos, S. Makridakis, and N. Stylianou, COVID-19: Forecasting confirmed cases and deaths with a simple time-series model, International Journal of Forecasting, https://doi.org/10.1016/j.ijforecast.2020.11.010
[28] S. Ballı, Data analysis of Covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods, https://doi.org/10.1016/j.chaos.2020.110512
[29] A. Zeroual, F. Harrou, A. Dairi, Y. Sun, Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study, https://doi.org/10.1016/j.chaos.2020.110121
[30] A. H. Elsheikha, A. I. Saba, M. A. Elaziz, S. Lu, S. Shanmugan, T. Muthuramalingam, R. Kumar, A. O. Mosleh, F. A. Essa , T. A. Shehabeldeen, Deep learning-based forecasting model for COVID-19 outbreak in Saudi Arabia, https://doi.org/10.1016/j.psep.2020.10.048
[31] M. A. A. Al-qaness, A. I. Saba, A. H. Elsheikh, M. A. Elaziz, R. A. Ibrahim, S. Lu, A. A. Hemedan, S. Shanmugan, A. A. Ewees, Efficient artificial intelligence forecasting models for COVID-19 outbreak in Russia and Brazil, https://doi.org/10.1016/j.psep.2020.11.007
[32] R. G. da Silvaa, M. H. D. M. Ribeiro, V. C. Mariani, L. S. Coelho, Forecasting Brazilian and American COVID-19 cases based on artificial intelligence coupled with climatic exogenous variables, https://doi.org/10.1016/j.chaos.2020.110027
[33] M. Kalantari, Forecasting COVID-19 pandemic using optimal singular spectrum analysis, Chaos, Solitons and Fractals, https://doi.org/10.1016/j.chaos.2020.110547
[34] O. Castillo, P. Melin, Forecasting of COVID-19 time series for countries in the world based on a hybrid approach combining the fractal dimension and fuzzy logic, Chaos, Solitons and Fractals, https://doi.org/10.1016/j.chaos.2020.110242
[35] R. K. Mojjada, A. Yadav, A.V. Prabhu, Y. Natarajan, Machine Learning Models for covid-19 future forecasting, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2020.10.962

[36] M. Ala'raj, M. Majdalawieh, N. Nizamuddin, Modeling and forecasting of COVID-19 using a hybrid dynamic model based on SEIRD with ARIMA corrections, Infectious Disease Modelling, https://doi.org/10.1016/j.idm.2020.11.007
[37] F. Khana, A. Saeed, S. Ali, Modelling and forecasting of new cases, deaths and recover cases of COVID-19 by using Vector Autoregressive model in Pakistan, Chaos, Solitons and Fractals, https://doi.org/10.1016/j.chaos.2020.110189
[38] M. Yousaf, S. Zahir, M. Riaz, S. M. Hussain, K. Shah, Statistical analysis of forecasting COVID-19 for upcoming month in Pakistan, Chaos, Solitons and Fractals, https://doi.org/10.1016/j.chaos.2020.109926
[39] V. K. R. Chimmula, L. Zhang, Time series forecasting of COVID-19 transmission in Canada using LSTM networks, Chaos, Solitons and Fractals, https://doi.org/10.1016/j.chaos.2020.109864
[40] S. Shastri, K. Singh, S. Kumar, P. Kour, V. Mansotra, Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study, Chaos, Solitons and Fractals, https://doi.org/10.1016/j.chaos.2020.110227
[41] Edwiga Kishinda Renald, Katharina Kreppel, Dmitry Kuznetsov, " Desirable Dog-Rabies Control Methods in an Urban setting in Africa - a Mathematical Model ", International Journal of Mathematical Sciences and Computing(IJMSC), Vol.6, No.1, pp.49-67, 2020. DOI: 10.5815/ijmsc.2020.01.05
[42] Akalu Abriham, Demsis Dejene, Tadele Abera, Abayneh Elias," Mathematical Modeling for COVID-19 Transmission Dynamics and the Impact of Prevention Strategies: A Case of Ethiopia ", International Journal of Mathematical Sciences and Computing(IJMSC), Vol.7, No.4, pp. 43-59, 2021. DOI: 10.5815/ijmsc.2021.04.05

Authors' Profiles

Arpita Santra

After her Bachelors in Mathematics from University of Calcutta, she is presently pursuing her MCA degree from Amity University, Kolkata. Her research interests include data analytics and machine learning.

Dr. Ambar Dutta

Dr. Dutta is at present working as Associate Professor in Amity Institute of Information Technology, Amity University, Kolkata. With more than 18 years of experience in teaching and research, Dr. Dutta authored a book and has published more than 50 papers in reputed journals/conferences. His research interest includes Data Analytics, Machine Learning, Image Processing, Information Retrieval. He is senior member of various professional bodies. He is active reviewer of many reputed journals of Elsevier, Springer, IET.

How to cite this paper: Arpita Santra, Ambar Dutta, "A Comprehensive Review of Machine Learning Techniques for Predicting the Outbreak of Covid-19 Cases", International Journal of Intelligent Systems and Applications(IJISA), Vol.14, No.3, pp.40-53, 2022. DOI: 10.5815/ijisa.2022.03.04

