E-ISSN 2581 – 7957
P-ISSN 2277 – 3916

CVR Journal of Science and Technology, Volume 21, December 2021
DOI: 10.32377/cvrjst2116

A Survey on COVID-19 Future Forecasting using Machine Learning Models

Syed Asif Ali [1] and Bipin Bihari Jaya Singh [2]
[1] PG Scholar, CVR College of Engineering/ IT Department, Hyderabad, India
Email: syedasifali891@gmail.com
[2] Professor, CVR College of Engineering/ IT Department, Hyderabad, India
Email: bbjayasingh9@rediffmail.com

Abstract: **Predictive methods for Learning Machines (MLs) have proved their worth in expecting long-term performance results to improve decision-making over the course of future action. ML models have long been used in many application domains that need to be identified and prioritized against negative threats. Many prediction methods are widely used to manage prediction problems. This study demonstrates the ability of ML models to predict the future of affected patients by COVID-19 and they are currently considered to be a threat to humanity. Predictions are made by each Machine Learning model to determine the number of confirmed cases in upcoming days. In this paper, the review is undertaken for a few methodologies that are applied to the solution of COVID19 problem. The ML model architecture is proposed with a day wise and country wise analysis for confirmed cases of COVID19 problem.**

Index Terms: **COVID-19, Machine Learning methods, Time series methods future forecasting, adjusted R2 score, supervised machine learning**

I. Introduction

Over the past decade, machine learning (ML) has become an important element of research by solving very complex and sophisticated real-world problems. Applications include Healthcare, Autonomous Vehicles (AV), Business Applications, Natural Language Processing (NLP), Intelligent Robotics, Gaming, Climate Modeling, Voice and Image Processing. The goal of this research is to develop an early warning pattern for the spread of a new coronavirus, also known as SARS-CoV-2.

The World Health Organization (WHO) has designated COVID-19 as its official designation. COVID-19 is currently a major threat. Human life is under danger all across the planet. By the end of the year, the virus was first discovered in the Chinese city of Wuhan. When a huge number of people begin to experience symptoms such as Pneumonia It affects the human body in a variety of ways. severe acute respiratory syndrome with multi-organ failure that can result in mortality in a short period of time. Significant symptoms of COVID-19 flu include cough, shortness of breath and diarrhea in other patients. The biggest problem with the disease is that its symptoms usually appear after 14 days.

Estimates have been made for the number of new confirmed cases in the coming days. This work treats assessment as a regression problem; The research is therefore based on several state-of-the-art supervised machine learning regression models and time series models.

II. Literature Survey

E. Gambhir et al [1] This study attempts to give readers a better understanding of how a different Machine Learning model that is Support Vector Machine and Polynomial Regression Algorithm might be used in real-world scenarios. In SVM for classification or outlier detection, it creates a hyperplane or set of hyperplanes in an N-dimensional space. Linear Regression can be thought of as a specific case of Polynomial Regression. Linear regression works with continuous data that is known and two variables that are correlated (target variable and independent variable). What if we knew variables were associated but the connection did not appear to be linear? We could apply polynomial regression to fit a polynomial equation to our dataset.

In this study work, the current Covid-19 transmission trend in the world has been successfully studied. Estimating the prevalence of Covid-19, also known as the novel coronavirus, helps to implement appropriate control measures. In this article there is still a need for in-depth research of the viral outbreak scenario in India and further exploration for future improvements. Further research may include looking at datasets from India and estimating the number of cases in the future, as well as how the death rate will change as the number of cases increases.

Petropoulos F et al [2] The document details the timeframe of a live forecasting exercise with far-reaching ramifications for planning and decision-making, as well as objective projections for COVID-19 cases that have been confirmed. Simple time series forecasting algorithms are used to predict confirmed instances of COVID-19. The exponential smoothing family of models is used to provide forecasts. The exponential smoothing family has demonstrated strong forecast accuracy in several forecasting competitions, and it is particularly well suited to short series. Exponential smoothing methods can capture a wide range of seasonal and trend forecasting patterns.

The fact that the tendency slowed over this time indicated that COVID-19 would not cause major concerns, especially outside of Mainland China. That was not the case, unfortunately. The most recent predictions, covering the period 02/03/2020 to 21/03/2020, suggest a considerable increase in the global trend of cases, as well as a rise in the accompanying uncertainty. The study also featured a detailed analysis of the outbreak situation which must still be uncovered to make future improvements. We believe that our predictions will help governments and individuals make

Received on 30-09-2021, Revised on 03-11-2021, Accepted on 05-12-2021.

informed judgments and take the necessary steps to limit the virus's spread as much as possible.

Kim S et al [3] The document describes the slowdown in the spread of COVID-19 due to social isolation and behavioural adjustments in the Republic of Korea (ROK). In addition, if vaccines or antiviral drugs are not ready by the fall, the second wave of a pandemic is likely.

Furthermore, if no vaccine or antiviral medications are ready by the fall, a second wave of the epidemic is likely. The impact of non-pharmaceutical strategies on short- and long-term outbreak dynamics was studied in this study. A deterministic compartment model was used to create a dynamic model of COVID-19 transmission. Individuals were categorised into six categories: susceptible, behavior-changed susceptible, exposed, infected, isolated, and recovered. With the incubation of the virus, individuals in the susceptible group become infected and move to the exposed group. Individuals who have been exposed to the virus join the infectious group. As the number of confirmed and isolated cases grows, susceptible people change their behaviour and join the behavior-changed susceptible group, which has a decreased transmission rate.

When it comes to limiting the spread of new infectious diseases for which treatments are not available, social distance can be successful. It is important that the general public be actively involved and implement appropriate government policies. Susceptible individuals need to be aware of the risk of disease and follow proper preventive measures. On the other hand, it is not possible to remove COVID-19 by simply using non-pharmaceutical techniques. Long-term social isolation is linked to a substantial socioeconomic cost. The availability of vaccines and antiviral medications is critical to bringing the COVID-19 epidemic to an end. More research is needed to see how non-pharmaceutical and pharmacological therapies might be balanced.

O. Tutsoy et al [4] The Susceptible-Infected-Recovered (SIR) model, which is widely used to estimate COVID-19 mortality, is first reviewed, and analysed in this study. The paper next introduces the Suspicious-Infected-Death (SpID) model, which is a novel higher-order, multi-dimensional, tightly linked, and parametric model. The model's three compartments are represented by SIR. Susceptible persons are those who are at risk of becoming infected if they come into contact with infectious people. When the infection occurs, they might be patient. Infectious persons are represented by the infectious group. They can spread the disease to others who are susceptible, and they can recover in a set amount of time. The number of recovered pearson's R is not included in the SpID model as it is in the SIR model since the optimization algorithms primarily focus on minimization, such as the number of suspects, infected, and dead people, rather than maximising, such as the number of recovered people.

The results showed that the Susceptible-Infected-Recovered (SIR) model can accurately predict Turkey's casualties. Furthermore, the model forecasts that the number of infected and diseased people will be at their lowest in 300 days, while the number of suspicious casualties would be at its lowest in 1000 days. The suggested model considers suspicious, infectious, and death casualties, but excludes critical care and intubation, non-pharmacological policies, pharmaceutical policies, and unknown uncertainty. Due to this reason further research is required and needs improvement.

R. Kumari et al [5] The work examines recently established forecasting models in depth and anticipates the number of confirmed, recovered, and mortality cases caused by COVID-19 in India. For prediction, correlation coefficients and multiple linear regression were utilised, as well as autocorrelation and autoregressive to increase accuracy. The suggested model, the degree of variables in the dataset must be discovered and computed, and this information is useful in better preparing the dataset to fulfil the requirements of machine learning algorithms. Python software is used to do a recovery technique and a correlation analysis on data. It provides a statistical overview of confirmed, recovered, and death cases, as well as a high correlation between current data. With the use of confirmed and recovered cases, multiple regression analysis is utilised to forecast cases of death. More than one variable is used to predict the outcome in this regression procedure. When more than one independent variable is used to predict a target variable, it is useful. Multiple linear regression approaches are employed in predictive analysis to explain the link between two independent variables (confirmed and recovered cases) and one dependent variable (death cases).

The predicted and actual values are very close to each other. This prediction could be useful in resource management, such as health care, and prompt action could be performed with previous planning to decrease human life loss. The Multiple regression and Autoregressive model could be used to forecast when the epidemic would stop in a certain region. Further surveys are required to get exact results.

R. K. Singh et al [6] We show how a COVID-19 projection for India at various administrative levels may be produced using an easily transferable statistical model based on the standard Holt-Winters approach. We apply our statistical model to produce 48-day estimates of these values in India based on a daily time series of accumulated infections, active infections, and deaths, assuming little change in national coping strategies. Using these findings in conjunction with a supplemental SIR model, we predict that one-third of the Indian population could be infected with COVID-19 at some point, and that complete recovery from COVID-19 will take an estimated 450 days.

The Holt-Winters technique is a time series forecasting approach that can account for both trend and seasonality. The Holt-Winters approach is made up of three other, much simpler smoothing methods that are Simple Exponential smoothing (SES), Holt's Exponential Smoothing (HES) and Winters Exponential Smoothing (WES). In a closed community, a SIR model computes the theoretical number of people infected with an infectious illness over time. The name of this class of models comes from the fact that they use coupled equations to relate the number of vulnerable individuals, infected people, and recovered people.

CVR Journal of Science and Technology, Volume 21, December 2021
DOI: 10.32377/cvrjst2116

The results of short-term statistical forecasts using the Holt-Winters approach imply that this method could be suitable for producing operational COVID-19 forecasts in India at various administrative levels. In general, Holt-Winters models record the level and most recent trend of a time series and are thus inadequate for long-term projections in circumstances like this one where there is no seasonal signal. More advanced mathematical-epidemiological models are necessary for such forecasts, necessitating additional research.

V. K. Gupta et al [7] In this paper we are solely looking for COVID-19 cases in India, which includes confirmed, deceased, and cured cases. We are doing this study based on cases that happened in different parts of India at different times. We cleansed the data and selected features, then forecasted all classes using random forest, linear model, support vector machine, decision tree, and neural network, finding that the random forest model outperformed the others. As a result, the random forest is used to predict and analyse all the results. The K-fold cross-validation is used to assess the model's consistency. The random forest is a classification algorithm that uses numerous decision trees to classify data. When creating each individual tree, it employs bagging and feature randomization to generate an uncorrelated forest of trees whose committee prediction is more accurate than that of any one tree. Multinomial logistic regression is a classification technique that extends logistic regression to issues with more than two discrete outcomes. We are utilising feed-forward neural networks with a single hidden layer and maybe skip-layer connections in this example. SVM can be used to classify or predict data. The input features are represented as vectors that are projected onto a higher-dimensional space. After that, an ideal hyperplane is built to separate the different instances of confirmed, death, and cured cases.

Because Random Forest is an ensemble model that uses bagging for sampling, we were blown away by its performance when compared to other models. Because the research is restricted, more research is required.

III. METHODOLOGIES

Machine learning is a branch of computer science that allows computers to learn without explicit programming. Machine learning is one of the most interesting methods. As the name suggests, it provides the computer with the ability to learn, which makes it more human. Machine learning is in vogue right now, probably in far more places than expected. Humans can learn from their past experiences, but computers cannot. Machine learning is a branch of artificial intelligence that helps systems learn from their past experiences and improve them without the need for human intervention. In other words, machine learning is similar to gardening. Seeds algorithms, nutrient data, you gardeners, and plants programs. Its main purpose is to create computer programs. Machine learning is inextricably linked to data analysis and statistics since the effectiveness of a learning algorithm is determined on the data used. Learning techniques, in general, are data-driven methods that combine core computer science concepts with notions from statistics, probability, and optimization. Machine learning is used in a variety of industries, including pharmaceutical, military, marketing, and security. The system uses machine learning algorithms to analyze datasets, including daily actual past data, and make predictions for future days.

Neural networks currently occupy a very important place in machine learning. Using neural networks makes it possible to capture abstract features from original data and therefore fulfill our goal of assessing the epidemic situation in a novel manner. The simplest and most intuitive neural network is a fully connected neural network. The neuron is its most basic component. Neurons can extract input from previous neurons in the previous layer and use the activation function to perform nonlinear modifications. A sophisticated nonlinear task occurs when all neurons combine with the appropriate structure, weighting, and bias. A neural network can theoretically equip any complex task with infinite neurons, which allows it to perform any task. On the other hand, better performance and shorter calculation time are not favorable. It is important to design a network with a specific feature to balance performance and time consumption.

A. Linear Regression

Regression analysis is a statistical study of dependencies and the relationship between two or more independent variables. There are still many regression techniques, but linear regression is the most commonly used.

In regression analysis, linear models and linear regression are mostly used. In that sequence, linear regression was also used. Linear regression is based on variables that use a large number of independent variables. From day to day, the independent variables vary. The built-in linear model involves the relationship between the underlying parameters and the data points. To determine accuracy and prediction, the rail and test approach is used. The system can be trained and tested with autonomy before making estimates.

B. The KNN algorithm

The KNN algorithm (k-nearest neighbor) is also widely used. Both taxonomy and regression approaches can be used with KNN. In the field of regression analysis, the KNN regression model is widely used. As a result, the KNN model was implemented, and the method performed superbly. The result of KNN regression is an attempt to give prominence to the object. The average value is KNN. It sets the average numerical target of the algorithm. The distance function is the same in KNN regression and taxonomy. etal [14] KNN's next 90 days (k-nearest neighbor) forecast This is the KNN-based forecast result. Prediction Linear regression is similar to the prediction result. Trained data using train and test approach. The machine isolated all this information. It splits the train automatically and tests the data after learning the data. Dhaka City data were compiled differently, so linear regression and k-nearest neighbors gave the same result. Its accuracy is determined, which helps us to understand the operation of the whole model.

CVR Journal of Science and Technology, Volume 21, December 2021
DOI: 10.32377/cvrjst2116

C. Polynomial Regression

Linear regression can be considered as a specific context of polynomial regression. Linear regression works with both known continuous data and interrelated variables (target variable and independent variable). What if we know that the variables are connected, and the connection does not appear to be simple? We can apply polynomial regression to fit polynomial equations in our dataset. Polynomial Regression is a supervised Machine Learning Algorithm that is learned using previous data and then validated using another dataset.

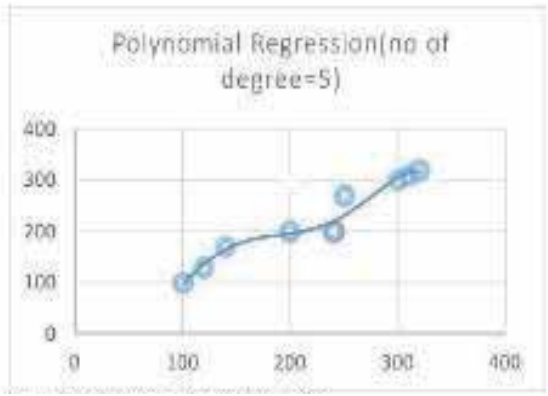

Figure 1. Polynomial Regression of degree 5

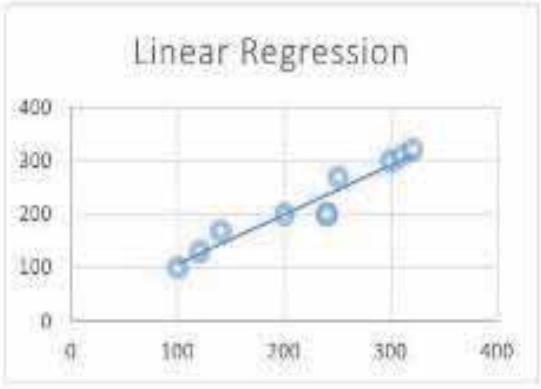

Figure 2. Linear Regression

M. Singh et al [13] Because the loss function and error rate in a simple linear model are high, the accuracy predicted by the simple linear model is lower than the accuracy predicted by the Polynomial model for non-linear data sets like those shown in Figures 1 and Figure 2. As a result, Polynomial Regression is a linear model with minor modifications that helps to improve accuracy for non-linear and complex datasets.

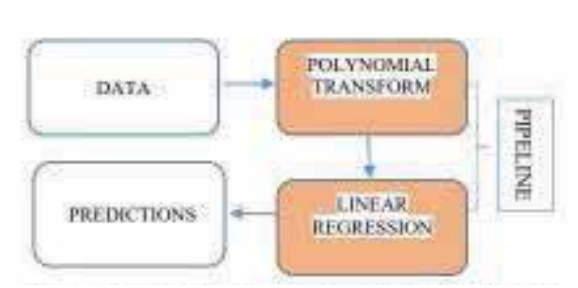

Figure 3. Pictorial representation of implementation of polynomial Regression

We utilize the Polynomial Features function to transform the data into a polynomial, and then use linear regression to fit the parameter in polynomial regression. The graphic illustration of this can be found in Figure 3. The polynomial characteristics convert the equation into an nth (degrees) equation. So, we must select it carefully because if the degree of polynomial is too low, the data will not fit, and if it is too high, the data will be overfit.

V. K. Gupta et al [7], The train and test data were converted for polynomial regression. the expected values from August 7, 2020, to August 28, 2020, as depicted visually. As a result, the polynomial Regression Algorithm has a 93 percent accuracy. Feature Selection To get the greatest results from our model, this stage involves feature extraction and selection. Having good and best features allows us to visualize the data's underlying structure. Feature Engineering has a substantial impact on the model's performance. It could entail separating or aggregating features to create new ones, or it could entail gathering data from external sources.

E. Gambhir et al [1], Dimensionality reduction makes it easier to evaluate and draw conclusions from a dataset. The dates were eliminated, and the useless parameters such as longitude and latitude were removed in order to draw better inferences from this dataset. Were transformed into a date-and-time object.

A. Bansal et al [10], Feature selection techniques are used in machine learning because they reduce training time, simplify the model so that users and researchers can understand it, reduce the dimensionality curse, and improve generalization. Another significant advantage of removing redundant or irrelevant features is that less redundant data translates to fewer decisions based on noise, resulting in less overfitting.

Information gain is used in mutual information feature selection to calculate the surprise or reduction in entropy caused by dataset transformation in some way. Typically, information gain is used by evaluating the information gain of each variable and then selecting the variable that minimizes entropy by maximizing information gain and splitting the dataset best into groups for effective classification. Each variable's gain in the context of the target variable is evaluated for the purpose of using information gain in feature selection. The calculation is reciprocal between the two randomly chosen variables.

CVR Journal of Science and Technology, Volume 21, December 2021
DOI: 10.32377/cvrjst2116

Recursive feature elimination removes recursive features and builds the model on the attributes that remain. The draper feature elimination technique is illustrated by recursive feature elimination. Model accuracy is used in this case to determine which attributes from the list of attributes would contribute the most to helping predict the target variable.

Correlation is defined as a measure of how two variables change in relation to one another in correlation feature selection. It is not uncommon for some features, despite being designed to measure different qualities, to be influenced by a common mechanism and to vary in tandem.

D. Fbprophet

Fbprophet An open-source algorithm developed by Facebook to estimate time-series data using an additive model to match non-linear trends with annual, monthly, and daily seasonal as well as holiday effects. It works best with time series with significant seasonal effects and multi-season historical data. The Prophet apologized for the missing data and trend shifts and it usually handles outliers well. We validate our data against the rolling mean, just like any other time series model.

There are three target classes in our dataset, each with many discrete instances. The following are the target classes:

(1) Confirmed cases: The number of instances that have been confirmed as of a specific date. It can be increased or decreased based on the following day, time, and location, which is only applicable to Indian states.

(2) Death cases: The total number of death cases at any given time. It can be increased or decreased based on the following day, time, and location, which is only applicable to Indian states.

(3) Cured cases: The total number of cured cases at any given time. It can be increased or decreased based on the following day, time, and location, which is only applicable to Indian states.

E. Performance Tuning

V. K. Gupta et al [7], The popular prediction models utilized in the investigation are listed in Figure 4, and the packages employed by these models are open-source libraries written in the R programming language and licensed under the GNU GPL. All of the packages are used here, each with its own way for model building. which are tuned for better results.

Model	Method	Required package	Tuning parameter
Random forest	randomForest	randomForest	mtry=2, ntree=500
SVM	svm	e1071	kernal=radial, degree=3
Decision tree	rpart	rpart	usesurrogate=0
Neural network	nnet	nnet	size=10
Multinomial logistic regression	multinome	nnet	maxit=1000

Figure 4. Machine learning models and their tuning parameters

IV. Study on Covid-19

Many researchers have participated in the study of the novel coronavirus after an outbreak in Wuhan, China in late December 2019 and developed a variety of models to predict its spread, transmission, and mortality. Some studies and research are related to drug development and a tool to diagnose the epidemic. Here are a few of the latest lessons discussed here.

Zhong et al. [15] developed a statistical model for forecasting timely coronavirus outbreaks in China. Hamzah et al. [16] developed an online platform for providing real-time information related to COVID-19 and statistical analysis of data. The Susceptible-Exposed-Infectious-Recovered (SEIR) prediction model has been used for daily forecasting. They have improved their small services to download data from different sources.

Morawska and Coo [17] discussed how COVID-19 spreads, especially in the air. Li et al. [18] investigated genetic mutations COVID-19 virus. This study has shown that the novel coronavirus has genetic similarities with the coronavirus found in rhinolophus sinicus, paradoxurus hermaphroditus, paguma larvata, aselliscus stoliczkanus, and civet, while homologous analysis shows that it is almost identical to the bat coronavirus.

Ma et al. [19] analyzed the effect of humidity and changes in the temperature of COVID-19 patients, but the study was limited to the city of Wuhan only. This study established a combination of temperature and humidity variations in daily deaths from the virus. Singh et al. [20] studied and compared SARS, MERS, and COVID-19 viruses based on transmission cycle, etiology, genetics, management, diagnosis, birth rates, laboratory diagnoses, clinical features, and radiation characteristics.

Pal et al. [21] showed the separation of ribonucleic acid bacterial group and origin of acute coronavirus respiratory and virion structure and genetic COVID-19 genes. Dutheil et al. [22] investigated the role of COVID-19 in reducing air pollution as many industries closed again and traffic is also very low.

Singh et al. [23] analyzed time series data and predicted the enrollment, death, and death rates for each reported case (death rate) based on COVID-19 global health data. This study concluded that the normal mortality of COVID-19 is positively correlated with the number of confirmed cases. It can also depend on the human diet process and the strength of the immune system. The study suggested that an emergency could arise prior to a proper injection. Other critical issues were assessed by several researchers, taking into account individual countries, provinces, and specific conclusions. Bhatnagar et al. [24] presented a detailed analysis of the COVID-19 epidemic with the help of boxplot and Q.

Ivanov et al. [25] analyzed and predicted the effects of the ongoing epidemic on land supply chains. They also performed a cost-based analysis of the condition of supply chains and the impact of COVID-19 on supply chains and associated risks. Hou et al. [26] conducted a SEIR model

CVR Journal of Science and Technology, Volume 21, December 2021
DOI: 10.32377/cvrjst2116

analysis to evaluate the functioning of the segregation of people especially the city of Wuhan and to develop new variations of the SEIR model. They conclude that isolation and separation are two powerful and unique tools to reduce the risk of infection. Roosa et al. [27] developed the COVID-19 predictive system in real time in China for some time. Tuli et al. [28] used the latest technology, such as machine learning and cloud computing, to predict the growth rate of the COVID-19 epidemic with the help of the Weibull model.

Xu et al. [29] described the pathogenesis of COVID-19 and compared it to SARS and MERS. These pathological features are very similar to SARS and MERS. The study provided some recommendations for doctors so that they could arrive on time, develop a patient treatment plan. Kucharski et al. [30] performed a mathematical model and analyzed four data sets. The study found that the transmission rate is between 1.6 to 2.6. Here they classify patients into four distinct categories: they can be infected, exposed (but not yet acquired), contagious, and removed (i.e., separated, recovered, or no longer infected). Yuvaraj et al. [31] A comprehensive neural network was used to analyze SARS-CoV-2 protein-ligand interactions against selected drugs. Other studies focused on the mental health of farmers involved in the poultry business [32].

Researchers are also working on experimental procedures and are trying to reduce the testing time. In this sequence, Assad et al [33] suggested that sample integration is the best way to do it, reducing the probationary period leading to a reduction in mortality but by a minimum of 10% positive cases. If the positive charges are very low, binary elimination algorithms are the best option.

V. Steps for ML Algorithms Implementation

Italy, USA, UK, and France are the two countries currently in Section 4 While India is in Section 3. This study attempts to establish a system for pre-distribution of the number of cases affected by COVID-19 using machine learning methods. The data used for the study include daily reports of cases of recent infections worldwide because of COVID-19. This is a shocking situation in the world as the number of confirmed cases is increasing day by day. The number of people affected by the Covid-19 pandemic in different parts of the world is not well known.

A. Data Preprocessing

The purpose of this step is to convert raw data into a form suitable for machine learning. Systematic and clean data allows the data scientist to obtain accurate results from the machine learning model used. This process includes data formatting, cleaning, and modeling. At the modelling stage, the data scientist trains many models to explain which one of them gives the most accurate predictions.

B. Model Training

The data scientists first collect the data and divide it into three subsets, he could proceed with model training. This process involves “feeding” the algorithm with training details. The algorithm will process the data and extract a model that is able to find the target value (attribute) in the new data - the answer you want to get by predicting analysis. The purpose of model training is to improve the model. Two styles of model training are very common - supervised and supervised learning. The choice of each style depends on whether you have to predict certain attributes or elements of group data in the same way.

Supervised Learning allows processing of data by targeted signals or labeled data. These qualities are mentioned in the historical data prior to training. With supervised learning, a data scientist can solve planning and deferred problems.

Unsupervised learning, with this training style, the algorithm analyzes unlabeled data. The purpose of model training is to find hidden connections between data objects and structural similarities or differences. Unsupervised learning aims to solve problems such as assembly, learning about the rules of assembly, and size reduction. For example, it can be used in data, furthering the phase to reduce the complexity of the data.

C. Model Testing

The purpose of this step is to develop a simpler model that is able to create a target value quickly and efficiently. A data scientist can achieve this goal by modifying the model. That is the use of model parameters to achieve the best performance of the algorithm One of the most effective methods of model testing and correction of the opposite

D. Cross Validation

Validation is very common and the planning method is used. Includes separate training database into ten equal parts (folders). The model provided is training only nine folders and then tested in the tenth (the one that had been left out). Training continues until the entire herd is set aside and used for testing. As a result of the model performance measurement, the specialist calculates the accumulated points for each set of parameters. A data scientist trains models with different sets of hyperparameters to define which model has the highest predictive accuracy. Guaranteed points fall on the performance of the middle model in ten catch folders. There the data scientist examines models with hyperparameter values that find the best guaranteed points. There are various error metrics for machine learning activities.

VI. Conclusions

Machine learning (ML) is the study of computer algorithms that improve themselves over time using data and experience. Artificial intelligence is considered a factor in this. Machine learning algorithms design a model based on sample data, referred to as "training data", without being specifically programmed to make predictions or decisions. Machine learning algorithms are used in a wide variety of applications, including medicine, email filtering, speech recognition and computer vision, where it is difficult or impossible to develop traditional algorithms to perform the required tasks. This paper highlights the recent research on Covid-19 Future forecasting using Machine Learning

Algorithms, focusing mainly on approaches based on number of newly infected cases, number of deaths and number of recovered. Different methods were compared in terms of processing time, complexity, discrimination, and rigidity. In terms of discrimination, complexity, and accuracy, we may conclude that Machine Learning approaches are the best option. By publishing this survey study, we hope that researchers in this field will be encouraged to pay more attention and collaborate on the use of localized approaches in machine learning systems.

REFERENCES

[1] E. Gambhir, R. Jain, A. Gupta and U. Tomer, "Regression Analysis of COVID-19 using Machine Learning Algorithms," *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, 2020, pp. 65-71, doi: 10.1109/ICOSEC49089.2020.9215356.

[2] Petropoulos F, Makridakis S. Forecasting the novel coronavirus COVID-19. PLoS One. 2020 Mar 31;15(3):e0231236. doi: 10.1371/journal.pone.0231236. PMID: 32231392; PMCID: PMC7108716.

[3] Kim S, Ko Y, Kim YJ, Jung E. The impact of social distancing and public behavior changes on COVID-19 transmission dynamics in the Republic of Korea. PLoS One. 2020 Sep 24;15(9):e0238684. doi: 10.1371/journal.pone.0238684. PMID: 32970716; PMCID: PMC7514094.

[4] O. Tutsoy, Ş. Çolak, A. Polat and K. Balikci, "A Novel Parametric Model for the Prediction and Analysis of the COVID-19 Casualties," in IEEE Access, vol. 8, pp. 193898-193906, 2020, doi: 10.1109/ACCESS.2020.3033146.

[5] R. Kumari et al., "Analysis and predictions of spread, recovery, and death caused by COVID-19 in India," in Big Data Mining and Analytics, vol. 4, no. 2, pp. 65-75, June 2021, doi: 10.26599/BDMA.2020.9020013.

[6] R. K. Singh et al., "Short-Term Statistical Forecasts of COVID-19 Infections in India," in IEEE Access, vol. 8, pp. 186932-186938, 2020, doi: 10.1109/ACCESS.2020.3029614.

[7] V. K. Gupta, A. Gupta, D. Kumar and A. Sardana, "Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model," in Big Data Mining and Analytics, vol. 4, no. 2, pp. 116-123, June 2021, doi: 10.26599/BDMA.2020.9020016.

[8] Y. Liu and Y. Xiao, "Analysis and Prediction of COVID-19 in Xinjiang Based on Machine Learning," 2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT), 2020, pp. 382-385, doi: 10.1109/ISCTT51595.2020.00072.

[9] A. U. Mandayam, R. A.C, S. Siddesha and S. K. Niranjan, "Prediction of Covid-19 pandemic based on Regression," 2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), 2020, pp. 1-5, doi: 10.1109/ICRCICN50933.2020.9296175.

[10] A. Bansal and U. Jayant, "Covid-19 Outbreak Modelling Using Regression Techniques," 2021 International Conference on Innovative Practices in Technology and Management (ICIPTM), 2021, pp. 113-118, doi: 10.1109/ICIPTM52218.2021.9388347.

[11] Z. Yang and K. Chen, "Machine Learning Methods on COVID-19 Situation Prediction," 2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE), 2020, pp. 78-83, doi: 10.1109/ICAICE51518.2020.00021.

[12] C. V. S. S. Nikil, H. Dalmia and G. J. R. Pavan Kumar, "Covid-19 Outbreak Analysis," 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), 2020, pp. 347-350, doi: 10.1109/ICSTCEE49637.2020.9276790.

[13] M. Singh and S. Dalmia, "Prediction of number of fatalities due to Covid-19 using Machine Learning," 2020 IEEE 17th India Council International Conference (INDICON), 2020, pp. 1-6, doi: 10.1109/INDICON49873.2020.9342390.

[14] A. Abdullha and S. Abujar, "COVID-19: Data Analysis and the situation Prediction Using Machine Learning Based on Bangladesh perspective," 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), 2020, pp. 1-8, doi: 10.1109/iSAI-NLP51646.2020.9376812.

[15] L. Zhong, L. Mu, 1. Li, 1. Wang, Z. Yin, and D. Liu, Early prediction of the 2019 novel coronavirus outbreak in the mainland China based on simple mathematical model, IEEE Access, vol. 8, pp. 51 761-51 769,2020.

[16] F. A. B. Harnzah, C. H. Lau, H. Nazri, D. V. Ligot, G. Lee, C. L. Tan, M. K. B. M. Shaib, U. H. B. Zaidon, A. B. Abdullah, M. H. Chung, et aI., Coronatracker: Worldwide COYID-19 outbreak data analysis and prediction, Bull World Health Organ, http://dx.doi.org/1O.2471IBLT.20.255695.

[17] L. Morawska and J. Cao, Airborne transmission of SARS CoY-2: The world should face the reality, Environment International, vol. 139, p. 105730, 2020.

[18] C. Li, Y. Yang, and L. Ren, Genetic evolution analysis of 2019 novel coronavirus and coronavirus from other species, Infection, Genetics and Evolution, vol. 82, p. 104285, 2020.

[19] Y. Ma, Y. Zhao, 1. Liu, X. He, B. Wang, S. Fu, J. Yan, J. Niu, 1. Zhou, and B. Luo, Effects of temperature variation and humidity on the death of COYID-19 in Wuhan, China, Science of the Total Environment, vol. 724, p. 138226,2020.

[20] A. Singh, A. Shaikh, R. Singh, and A. K. Singh, COYID-19: From bench to bed side, Diabetes & Metabolic Syndrome: Clinical Research & Reviews, vol. 14, no. 4, pp. 277-281, 2020.

[21] M. Pal, G. Berhanu, C. Desalegn, and V. Kandi, Severe acute respiratory syndrome coronavirus-2 (SARS-CoY-2): An update, Cureus, vol. 12, no. 3, pp. 1-13,2020.

[22] F. Dutheil, J. S. Baker, and V. Navel, COYID-19 as a factor influencing air pollution? Environmental Pollution, vol. 263,p.114466,2020.

[23] D. Singh, V. Kumar, V. Yadav, and M. Kaur, Deep convolutional neural networks-based classification model for COYID-19 infected patients using chest X-ray images, International Journal of Pattern Recognition and Artificial Intelligence, doi: 1O.1142/S0218001421510046.

[24] V. Bhatnagar, R. C. Poonia, P. Nagar, S. Kumar, V. Singh, L. Raja, and P. Dass, Descriptive analysis of COYID-19 patients in the context of India, Journal of Interdisciplinary Mathematics, doi: 10.1080/09720502.2020.1761635.

[25] D. Ivanov, Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoY-2) case, Transportation Research Part E: Logistics and Transportation Review, vol. 136, p. 101922, 2020.

[26] C. Hou, J. Chen, Y. Zhou, L. Hua, J. Yuan, S. He, Y. Guo, S. Zhang, Q. Jia, C. Zhao, et aI., The effectiveness of quarantine of Wuhan city against the corona virus disease 2019 (COVID-19): A well-mixed SEIR model analysis, Journal of Medical Virology, vol. 92, pp. 841-848, 2020.

[27] K. Roosa, Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, 1. Hyman, P. Yan, and G. Chowell, Real-time forecasts of the COYID-19 epidemic in China from February 5th to February 24th, 2020, Infectious Disease Modelling, doi: 10.1016/j.idm.2020.02.002.

[28] S. Tuli, S. Tuli, R. Tuli, and S. S. Gill, Predicting the growth and trend of COYID-19 pandemic using machine learning and cloud computing, Internet of Things, vol. 11, p. 100222, 2020.

[29] Z. Xu, L. Shi, Y. Wang, J. Zhang, L. Huang, C. Zhang, S. Liu, P. Zhao, H. Liu, L. Zhu, et aI., Pathological findings of COVID-19 associated with acute respiratory distress syndrome, The Lancet Respiratory Medicine, vol. 8, no. 4, pp.420-422,2020.

[30] A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R. M. Eggo, F. Sun, M. lit, J. D. Munday, et aI., Early dynamics of transmission and control of COYID-19: A mathematical modelling study, The Lancet Infectious Diseases, vol. 20, no. 5, pp. 553-558, 2020.

[31] N. Yuvaraj, K. Srihari, S. Chandragandhi, R. A. Raja, G. Dhiman, and A. Kaur, Analysis of protein-ligand interactions of SARS-CoY-2 against selective drug using deep neural networks, Big Data Mining and Analytics, doi: 10.26599/BDMA.2020.9020007.

[32] . Dhiman, The effects of coronavirus (COVID 19) on the psychological health of indian poultry farmers, Coronaviruses, doi: 10.2174/26667967019992 00617160755.

[33] A. Assad, M. A. Wani, and K. Deep, A comprehensive strategy to lower number of COYID-19 tests, SSRN Electronic Journal, doi: 1O.2139/ssrn.3578240