HiLL PUBLISHING

Journal of Applied Mathematics and Computation, 2022, 6(1), 121-126
https://www.hillpublisher.com/journals/JAMC/
ISSN Online: 2576-0653
ISSN Print: 2576-0645

Analysis of COVID-19 Prediction Models: Design & Analysis of New Machine Learning Approach

Raja Marappan*, S. Bhaskaran, N. Aakaash, S. Mathu Mitha

School of Computing, SASTRA Deemed University, Thanjavur, India.

How to cite this paper: Raja Marappan, S. Bhaskaran, N. Aakaash, S. Mathu Mitha. (2022) Analysis of COVID-19 Prediction Models: Design & Analysis of New Machine Learning Approach. *Journal of Applied Mathematics and Computation*, 6(1), 121-126.
DOI: 10.26855/jamc.2022.03.013

Received: January 10, 2022
Accepted: February 8, 2022
Published: March 18, 2022

***Corresponding author:** Raja Marappan, School of Computing, SASTRA Deemed University, Thanjavur, India.
Email: raja_csmath@cse.sastra.edu

Abstract

Recently the Coronavirus disease 2019 is the most worldwide health concern that is caused by the SARS-Cov2 virus. This results in a greater mortality rate with various clinical manifestations. The prediction of case trends and identification are becoming crucial in effective virus mitigation. There are different prediction methods such as statistical and epidemiological models are designed worldwide to determine and analyze the count of infected individuals with mortality rate. There are certain drawbacks in these models—in terms of lack of essential information and high uncertainty level results in lower prediction accuracy. To resolve the issues in these existing models, it is necessary to construct a new model using machine learning to predict the trends in the cases with better accuracy. This research focuses on the design of a new model using different soft computing methods. The performance of the model is evaluated to predict and analyze the infection and rate of mortality. The proposed model is expected to find the best results for predicting COVID case trends and mortality rates over the recent well-known methods.

Keywords

COVID-19, machine learning, SARS-CoV2, linear regression, perceptron, fuzzy systems

1. Introduction

The COVID-19 epidemic occurred in December 2019 in Wuhan China, which is caused by a novel Extreme Acute Respiratory Syndrome Coronavirus (SARS-CoV-2). The World Health Organization (WHO) announced on 30th January 2020 the occurrence is a pandemic and emergency worldwide. The clinical symptoms of COVID-19 are fatigue, tiredness, respiratory disorder, dry cough, etc. Some patients restore without any care but people with some diseases like diabetes and cardiovascular are vulnerable in this case. Hence, the social distance is required to slow down and prevent its transmission. This results in a great threat to all sectors of the country and healthcare. The government and all other organizations are struggling hard to mitigate the virus. Hence the use of soft computing techniques such as data mining, machine learning (ML), artificial intelligence (AI), deep learning (DL) plays a critical role in the contamination and diagnosis of the COVID-19 pandemic.

Over the last decade, ML has shown itself as a prominent domain of study for solving many complex real-world problems. ML is most used for forecasting and prediction. Different ML strategies have been applied for this pandemic prediction. Some of the neural network and regression models are applied to predict the patient's diseases. This research is focused on predicting COVID-19 case trends. These predictions can be very helpful in making decisions to manage the disease. This research also analyzes the drawbacks in the existing models and designs the new prediction model using ML strategies to achieve better accuracy.

2. Literature Survey of Recent COVID Prediction Methods

ML algorithms are being applied for various prediction and forecasting tasks. A supervised ML model is proposed for predicting COVID-19 infection [1]. The proposed method applied decision tree, logistic regression, naïve Bayes, Support Vector Machine (SVM) for prediction. The correlation coefficient analyzes the relationship between two sets of data characteristics. The performance of the models is analyzed based on their sensitivity and accuracy. It has been analyzed that the decision tree approach yielded good accuracy compared to other methods concerning specificity and sensitivity.

Some of the artificial neural networks (ANN) models also proposed for this pandemic prediction with the least absolute shrinkage & selection, Hybrid Multilayered Perceptron-Imperialist Competitive Algorithm (MLP-ICA) [2]. These models are used to predict semi-complex and simple datasets. The MLP-CIA gives good accuracy only when there is no occurrence of an interrupt.

An alternative approach was proposed for COVID-19 prediction using the time series forecasting, logistic function, conjugate gradients (CG), Broyden-Fletcher-Goldfarb-Shanno (BFGS), Nelder-Mead, and limited memory bound-constrained BFGS (L-BFGS-B) [3]. This model has been trained on the COVID case data from Italy, Australia, UK. The prophet algorithm was used on time series forecasting which gave good accuracy and performance compared to logistic function.

Partial Least Square (SIMPLS) is developed to analyze and predict the mortality of the Covid-19 virus. It is a linear machine learning method [4]. The reduction of variables is being applied in removing the factors which are unnecessary from the dataset. The important values from the dataset are processed using the weighted sum of squares of the weights concerning SIMPLS analysis. The model was proposed to have a sensitivity of 80%, accuracy of 80% from the training data set, and accuracy of 87% from the validation set.

3. Analysis of Recent COVID Prediction Methods

There are different ML and DL models that have been developed recently to predict the COVID trends [1-7]. The accuracy of these models depends on several characteristics. The primary characteristics are the scale of the model, inconsistent dataset, missing values, testing, and training dataset. The most popular ML models for COVID case trends are linear regression, decision tree, ANN, Naive Bayes, MLP-ICA [1-3, 6-7]. It has been analyzed that decision tree, ANN, MLP-ICA provided good accuracy for case trend prediction. The partial least squares method is commonly used for predicting mortality with an accuracy of 80% [4]. The principal component analysis is commonly applied to extract the features from several datasets.

4. Need for Higher Accuracy ML Models for COVID Prediction

The high accuracy prediction model is required to predict the case trend and mortality rate for making informed decisions on controlling the coronavirus. Hence, it is required to implement different ML strategies, and then these models are to be integrated for the best accurate prediction. There is a need to primarily focus on feature extraction from the preprocessed dataset. This process is very important because this helps to discard less important features from the dataset. Each ML model must be trained with the dataset and this trained model will be evaluated based on performance metrics.

5. Design of New ML-based COVID-19 Prediction Model

Since there is a need for an accurate prediction model, initially the dataset is collected from some benchmarks like CSSE at Johns Hopkins University repository. The design of the ML model is sketched in Figure 1. The new model will be designed using the fuzzy system, linear regression, and multi-layer perceptron approaches. The individual prediction algorithms will be trained with the same preprocessed data for better evaluation. The model involves these operations—data gathering, data clearing and analysis, data transformation. Feature extraction, model building, training different models, and performance metrics are evaluated and visualized. For better accuracy and performance of the ML model, more data samples will be included to improve the context of the data. The model will be evaluated using hyperparameter fine-tuning and cross-validation approaches. The performance of the model will be evaluated to predict and analyze the infection and rate of mortality.

6. Results & Analysis

The proposed method is implemented using Python and the results are analyzed as follows:

6.1 Country Wise Heat Map

The countrywide heat map is shown in Figure 2 consists of all the countries and their number of confirmed cases, number of deaths, active cases, and mortality rate. The heat map is visualized in descending order of COVID cases. From

the heat map, we can understand the most affected country due to COVID is the US which is followed by India with 33,791,061 cases.

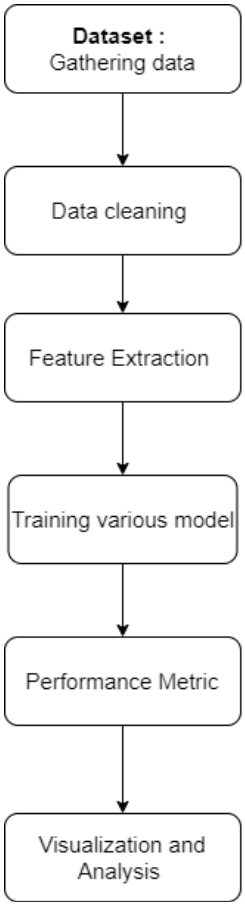

Figure 1. New COVID-19 Prediction model using ML.

	Country Name	Number of Confirmed Cases	Number of Deaths	Number of Recoveries	Number of Active Cases	Mortality Rate
0	US	43656368	701708	0.000000	42954660.000000	0.016073
1	India	33791061	448573	0.000000	33342488.000000	0.013275
2	Brazil	21445651	597255	0.000000	20848396.000000	0.027850
3	United Kingdom	7878571	137171	0.000000	7741400.000000	0.017411
4	Russia	7425057	204424	0.000000	7220633.000000	0.027532
5	Turkey	7182913	64264	0.000000	7118649.000000	0.008947
6	France	7111154	117535	0.000000	6993619.000000	0.016528
7	Iran	5601565	120663	0.000000	5480902.000000	0.021541
8	Argentina	5258466	115225	0.000000	5143241.000000	0.021912
9	Spain	4961128	86463	0.000000	4874665.000000	0.017428
10	Colombia	4959144	126336	0.000000	4832808.000000	0.025475

Figure 2. Country Heat Map.

6.2 State Wise Heat Map

The statewide heat map is shown in Figure 3 consists of the most affected states in the world. It also contains the total number of cases, deaths, active and mortality rate. The heat map is visualized in descending order of COVID cases. From the heat map, we can understand the most affected in England and follow up by Maharashtra with 6,553,961 cases.

6.3 Circle & Bar Plot

The circle plot structure is shown in Figure 4 is an interactive plot, which was developed using altair API. From the circle plot, we can understand that China had the most number of COVID cases initially, but they reduced the cases drastically over the period. On the other hand, we can see that the US has been continuously affected by COVID without any sign of a decrease in cases. We can also observe that the COVID cases in India are increasing continuously showing an upward trend in the plot.

6.4 Polynomial Regression

Figure 5 shows the prediction of the COVID case trend using the polynomial regression model. We chose polynomial because the COVID cases are unpredictable and the data is nonlinear in format. So using polynomial regression will give better results compared to linear regression. From the graph, the analysis gives better fitting of data in the curve initially. But it fails to fit data over time.

	Country Name	Number of Confirmed Cases	Number of Deaths	Number of Active Cases	Mortality Rate
13	US	43658244	701678	42956566.000000	0.016072
115	India	33791061	448573	33342488.000000	0.013275
171	Brazil	21445651	597255	20848396.000000	0.027850
9	United Kingdom	7878571	137171	7741400.000000	0.017411
50	Russia	7425057	204424	7220633.000000	0.027532
14	Turkey	7182913	64264	7118649.000000	0.008947
132	France	7111154	117535	6993619.000000	0.016528
113	Iran	5601565	120663	5480902.000000	0.021541
188	Argentina	5258466	115225	5143241.000000	0.021912
30	Spain	4961128	86463	4874665.000000	0.017428
157	Colombia	4959144	126336	4832808.000000	0.025475
109	Italy	4675758	130973	4544785.000000	0.028011
128	Germany	4249061	93781	4155280.000000	0.022071
114	Indonesia	4216728	142026	4074702.000000	0.033682

Figure 3. State Heat Map.

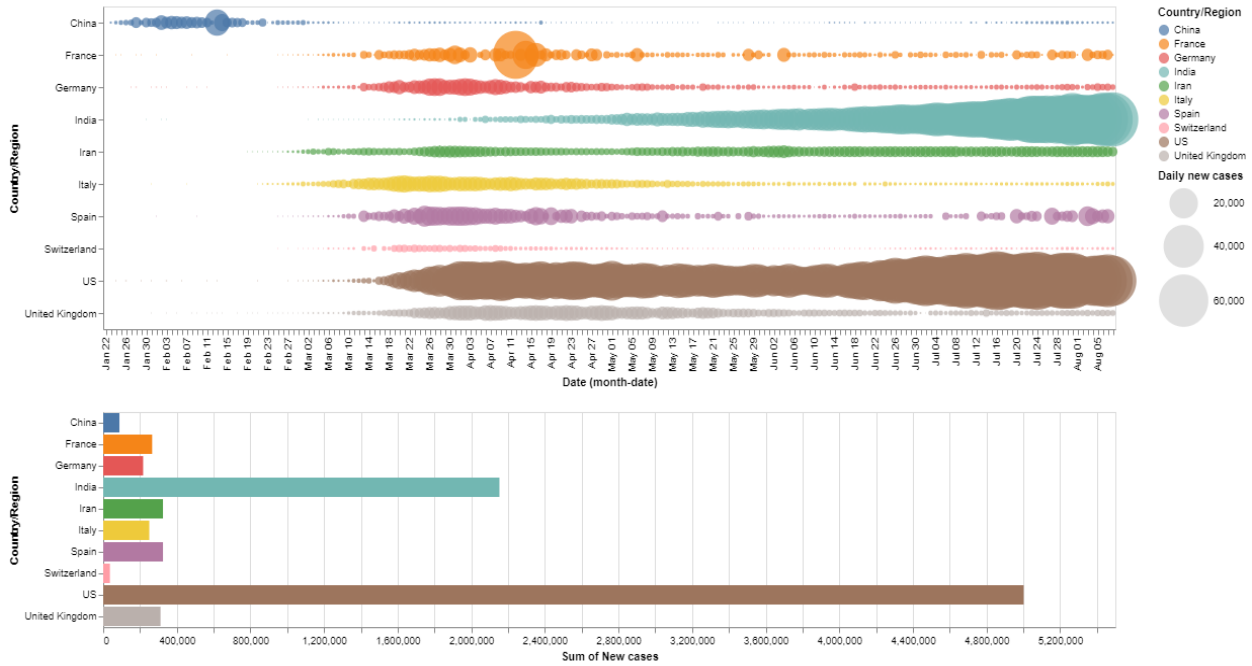

Figure 4. Circle Plot.

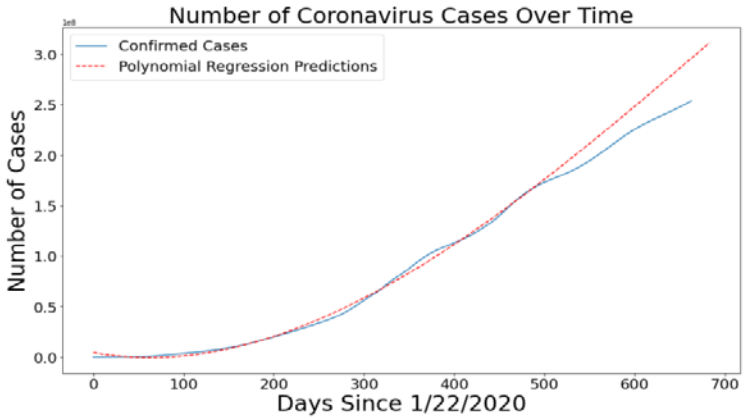

Figure 5. Polynomial Regression.

6.5 SVM

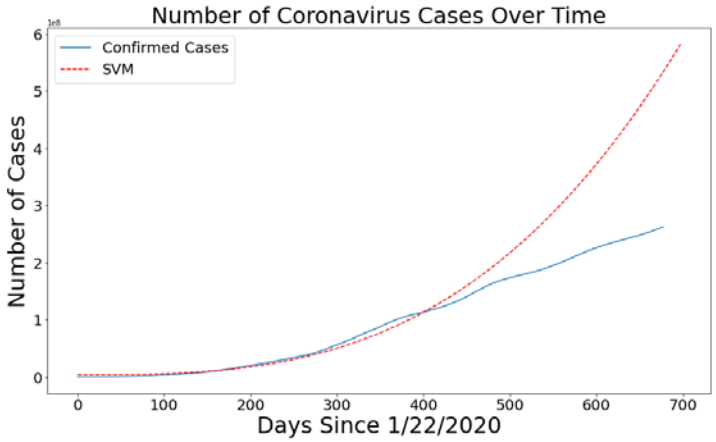

Figure 6. SVM Plot.

Figure 6 shows the prediction done by the SVM. An SVM model is fitting fewer data points in the curve compared to polynomial regression. Similar to the polynomial model, SVM also fits the data in the curve at the initial stages and fails to fit the data later in the curve.

6.6 Neural Prophet

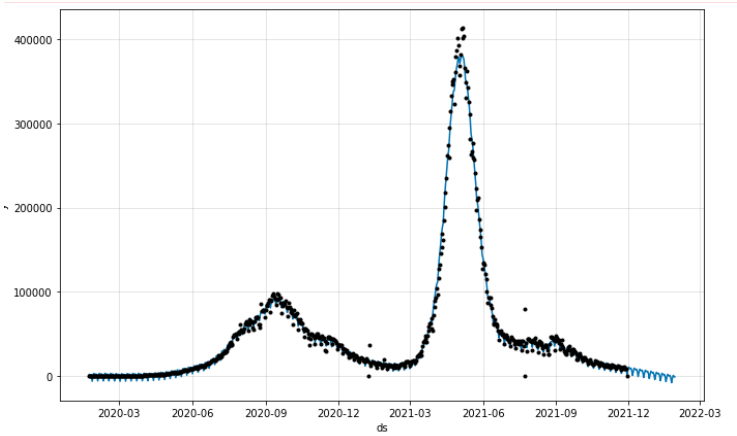

Figure 7. Neural Prophet Prediction.

Figure 7 shows the prediction done by the neural prophet model. The model fits the data points along the predicted line almost exactly. The above graph predicts the next 20 days of cases in India. As the neural prophet is a tailor-made time series analysis, it provides better accuracy compared to other models.

6.7 Performance Evaluation

Table 1. Performance comparison

Methods	RMSE
Neural prophet	161134.9
Polynomial regression	26204310
SVM	146864900

The performance comparison is shown in Table 1. For the performance metrics, we followed RMSE to compare the best ML model. The lower the RMSE value, the greater the performance. Here in the picture, we observed that the Neural Prophet has the lower value, so we can conclude that the Neural Prophet is the best model for COVID-19 prediction.

7. Conclusions & Future Work

The recent trends in COVID-19 prediction models are explored in this research. The importance of an accurate prediction model for COVID-19 is discussed. Evaluation of best prediction models for Coronavirus is proposed. In the proposed model, the individual prediction algorithms are trained with the same preprocessed data for better evaluation. From this analysis, we understood that neural prophet provides better accuracy when compared with other ML models. As there is a lack of a standardized model in predicting COVID case trends. As neural prophet uses a modular architecture and is highly scalable, the model can easily scale when given a huge dataset. This analysis can help the government and other organizations in finding the best model for COVID prediction, which could save valuable lives. Neural Prophet uses a deep learning model because the execution time is increased when provided with a huge dataset and many hidden neuron layers. As a plan, tuning the hyperparameter of the model in such a way that it will further reduce the execution time and will yield more accurate results.

The following are the future work:

- Data cleaning and data preprocessing will be done on the dataset using the new approach.
- Feature extraction will be done on the dataset using specific features.
- The prediction model will be implemented using Python.
- The predictions will be evaluated using performance metrics.

References

[1] Muhammad, L. J., Algehyne, E. A., Usman, S. S., Ahmad, A., Chakraborty, C., and Mohammed, I. A. (2020). Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset. SN Computer Science, 2(1). https://doi.org/10.1007/s42979-020-00394-7.

[2] Pinter, G. (2020, June 2). COVID-19 Pandemic Prediction for Hungary; A Hybrid Machine Learning Approach. MDPI. https://www.mdpi.com/2227-7390/8/6/890.

[3] Rahimi, I. (2021, March 3). Analysis and Prediction of COVID-19 Using SIR, SEIQR, and Machine Learning Models: Australia, Italy, and UK Cases. MDPI. https://www.mdpi.com/2078-2489/12/3/109.

[4] Banoei, M. M. (2021, September 8). Machine-learning-based COVID-19 mortality prediction model and identification of patients at low and high risk of dying. Critical Care. https://ccforum.biomedcentral.com/articles/10.1186/s13054-021-03749-5.

[5] Furqan Rustam, Aijaz Ahmad Reshi, Arif Mehmood, and Saleem Ullah. (2020, May 25). COVID-19 Future Forecasting Using Supervised Machine Learning Models. IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/document/9099302.

[6] Ardabili, S. F. (2020, September 1). COVID-19 Outbreak Prediction with Machine Learning. MDPI. https://www.mdpi.com/1999-4893/13/10/249.

[7] Kolla Bhanu Prakash, Sagar Imambi, S., Mohammed Ismail, Pavan Kumar, T., and Naga Pawan, Y. V. R. (2020, May 1). Analysis, Prediction and Evaluation of COVID-19 Datasets using Machine Learning Algorithms. Http://Www.Warse.Org/. http://www.warse.org/IJETER/static/pdf/file/ijeter117852020.pdf.