

# Probable Forecasting of Epidemic COVID-19 in Using COCUBE Model for the State of Tamilnadu, India

Prasannavenkatesan Theerthagiri (✉ [prasannait91@gmail.com](mailto:prasannait91@gmail.com))

GITAM University

---

## Research Article

**Keywords:** COVID-19, future prediction, infection rate, COCUBE model, decease rate

**Posted Date:** July 16th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-41348/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on February 3rd, 2021. See the published version at <https://doi.org/10.4108/eai.3-2-2021.168601>.

# Abstract

The world has been struck due to the dangerous human threat called Corona Virus Disease 2019. This research work proposes a methodology to encounter the future infection rate, curing rate, and decrease rate. This uses the artificial intelligence algorithm to design and develop the proposed confirmed, cured, deceased (COCUDE) model. A machine learning model has been developed with several iterations to design the proposed COCUDE model. The Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Correlated Akaike Information criterion (AICc) metrics are analyzed to check the stationary and quality for the proposed COCUDE model. The prediction results are evaluated by the performance error metrics such as mean square error (MSE) and root mean square error (RMSE), in which the errors are lower for the proposed model. Thus, the prediction results indicate the proposed COCUDE model might accurately predict future COVID-19 infection rates. It might support the corresponding authorities to take the precautionous action on the required necessities for the medical and clinical infrastructures and equipment.

## 1. Introduction And Review Of Covid-19

The year 2020 makes a historical entry to the world not because of its fancy number but because of the major human infectious threat called novel Corona Virus Disease 2019 (COVID- 19). Even though COVID-19 virus was found during late November 2019 at the Chinese city Wuhan in the Hubei province, it became a deadly spreading disease throughout the world in 2020 [1, 2]. Still May 2020, no proven medicine or vaccine was found to break the transmission of the infections from COVID-19 virus. However, few precautionous potential therapeutic treatments are being given to infected patients to improve their immune system.

Apart from these, the source wherefrom coronavirus came is not yet scientifically proven. Although, few are reported that the source of COVID-19 infection is coming from wild animals especially from bats and pangolin, but it is not yet confirmed. Further, the COVID-19 virus spreads even with the physical touch from the infected human to human, from their used objects, and places. Based on various reports from across the globe, the incubation period is 3-14 days for COVID-19 virus [2, 3].

The COVID-19 virus is spreading rapidly and its infection severity to human as well as to the pet animals is very much higher. In countries throughout the world, the mortality rate due to the COVID-19 virus infection is climbing dangerously. The major spreading method of COVID- 19 is from the droplets of coughing, sneezing, and talking of the infected humans to the normal uninfected human. Likewise, the novel coronavirus is being widely spread by humans to human contact. As the infected humans are moving to the various cities and countries around the world, the spreading is rising with an increasing pace. In such a way that from the Hubei province of Wuhan city in China country is transmitted to the

global public and it makes the global pandemic situation. Such that a much needed instantaneous action should be taken to reduce the speed of the spread of COVID-19 to the worldwide [2, 3, 4].

Typically, being a few days to a week the COVID-19 infected human would not have any symptoms of infection. During or after a week, the COVID-19 symptoms are starting from mild cough, fever with respiratory problems and symptoms of pneumonia. Then based on the immune system of the individual, the COVID-19 infections affect the normal functions of human's vital organs and might lead to death. Most of the COVID-19 infected human is admitted to the hospital with mild respiratory problems. Commonly, the people with pre-existing chronic disease, weak immune system, and old age humans are targeted for the COVID-19 infection. As well as the decease rate from those patients is much higher than other patients [2, 5].

Owing to the outbreak of the COVID-19 virus infection to the global public, during 11th March 2020 the world health organization (WHO) had declared it as a pandemic. Even several countries had implemented the strict rules to secure public health namely, lockdown, quarantine, social distancing, etc. But the virus infection and decease rate are marginally higher. Various virology, medical, and clinical research developments are going on around the world. Besides, at this time much essential research should be concentrated on the prediction of the COVID-19 infections among the global public [1, 2].

As the number infected cases are increasing, the essential infrastructures, medical facilities, innovative technologies, diagnostic models, communications, informatics techniques, and advanced equipment's required for the future days are must be made available. It includes the required number of beds, medicines, ventilation, testing kits for the patients and PPE kit, gloves, N-95 masks and other safety equipment for the physicians and caretakers. An organized plan should be prepared for developing such medical materials and other essentials. It might support the clinical industry to serve the infected patients in a proper and calm way [2, 6, 7].

Therefore, to forecast the spreads and risk of infections an appropriate optimal methodology and mathematical modelling is much needed in order to challenge and control the spreads of the coronavirus and to safeguard the public health. Such that this research work proposes a confirmed, cured, deceased (COCUDE) model to predict the future COVID-19 infection rate for the public in the state of Tamilnadu, India [5].

As the origin of COVID-19 is China, the infection to human in India must be spread through the foreign infected traveller [8, 9]. Specifically, as this research focuses on the Tamilnadu state of India, the possibility the people are being infected by the COVID-19 transmission from travellers from outside the state. Apparently, they are spreading the infection to other regions of the state [5]. Therefore, this research work focus on the prediction of the number of COVID-19 infected, cured, and decease cases using the COCUDE model based on the dataset available from the Ministry of Health, Tamil Nadu, India.

## 2. Methodology

The proposed COCUBE model and its algorithm are extensively explained in this section for the prediction of future infection rate of COVID–19.

In epidemiology, the mathematical models are utilized for the extensive analysis of infectious diseases. The well-defined epidemic mathematical model gives the analysis of qualitative as well as quantitative information of the infectious diseases. Further, the vital importance of the epidemic mathematical model is the probable estimation of spread and decrease rate of infectious diseases. Also, the epidemic mathematical model has a strong influence on hospital emergency planning and risk management, infection control measures, health-economic related aspects, and decision making [10]. Typically, the epidemic mathematical model facilitates the estimations and predictions of infectious diseases by using complex computational modelling techniques [11, 12]

## 2.1 COCUBE Model

The proposed COCUBE uses the artificial intelligence-based machine learning methodology in order to predict future COVID–19 scenario. The artificial intelligence is a technique, intended to mimic the human brain by artificiality connected neuron nodes with several computations in order to arrive a solution or to make the decision for the problem. The machine learning is a kind of artificial neural network. Typically, machine learning is a self-adaptive learning algorithm, in which it increasingly analyses and processes the given problem or dataset and facilitates the solution based on its experience. Such that, the machine learning incorporates multi-level neurons of an artificial neural network algorithm for its computations.

In the proposed methodology, the machine learning algorithm utilizes the aspects of the SEIR (Susceptible-Exposed-Infected-Recovered) model and decrease rate, in order to assess and estimate the epidemic trends [13, 14, 15]. The proposed model is developed considering 1) the COVID–19 infection spreads from an infected human to other humans, 2) normal deaths are not considered for the decrease rate prediction, 3) the COVID–19 dataset for the proposed methodology is retrieved from the Department of Health and Family Welfare, Government of Tamil Nadu, 4) the training dataset for COCUBE model is taken from the date of the first infection in the state of Tamilnadu, India to 20th June 2020 [16, 17].

### 2.1.1 COCUBE Model Formulation:

In this proposed COCUBE model, a nonlinear artificial neural network NARX (Nonlinear Auto-Regressive model with exogenous input) model is employed for precise prediction of the each CO, CU, and DE models. The NARX neural network is a kind of dynamic recurrent neural network, in which the NARX is working with multi feedback connection of hidden layers for predictions. The NARX neural network model has been applied in several prediction applications of nonlinear processes. The following NARX equation is employed for the proposed COCUBE model.

$$CO(t) = f(CU_1(t - 1), DE_1(t - 1), CU_2(t - 2), DE_2(t - 2), \dots, CU_n(t - m), DE_n(t - m))$$

... .... (1)

$$CU(t) = f(CO_1(t - 1), DE_1(t - 1), CO_2(t - 2), DE_2(t - 2), \dots, CO_n(t - m), DE_n(t - m))$$

... .. (2)

$$DE(t) = f(CO_1(t - 1), CU_1(t - 1), CO_2(t - 2), CU_2(t - 2), \dots, CO_n(t - m), CU_n(t - m))$$

... .. (3)

Where CO denotes confirmed model, CU denotes Cured model, DE represents the decease model, 't' is the time with respect to inputs, 'n' is the number of training samples, and 'm' is the number of delays. The equation (1) is employed for future prediction confirmed cases (CO) based on the cured cases (CU) and deceased (DE) cases. In order to train the CO model, the CU and DE models are given as the input to the NARX model. As per the given equation (2), the future prediction of cured cases (CU) is supported by the inputs of confirmed (CO) cases and deceased cases (DE). Such that, the CU model gives future predictions on the basis of the CO and DE model.

Similarly, the equation (3) is adopted for future prediction deceased cases (CO) based on the confirmed cases (CO) and cured (CU) cases. In order to train the DE model, the CO and CU models are given as the input to the NARX model. So that, the COUDE model is supported by NARX neural network algorithm for the future prediction of the COVID-19 confirmed cases, cured cases, and deceased cases.

Figure.1 depicts the proposed COUDE model using the machine learning algorithm. In the proposed COUDE model, the metadata is collected from the repository [16, 17]. Then, the outliers and missing data compromised by pre-processing imputation using mean/median values. The pre-processed dataset has been taken for the training process using the machine learning algorithm. The COUDE model then predicts the future prediction using the training dataset and validates the predicted dataset the testing dataset. To estimate the stationary of the predicted dataset, it assessed with the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Correlated Akaike Information criterion (AICc) metrics [18].

The lowest values of the AIC and BIC metrics corresponds to the better stationary model. Further, the predicted COVID-19 dataset has been evaluated with the performance error metrics for the accuracy evaluation. The performance error metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) are computed for the prediction results [19, 20, 21]. Moreover, the predicted future COVID-19 dataset has been analyzed with precision rate.

The proposed COUDE model had taken the three type of datasets of COVID-19 infection confirmed cases, cured/recovered cases, and decease cases for the prediction corresponding future rates. Such that the proposed

### 3. Performance Evaluation And Discussions

The proposed COCUDE model is trained, tested, validated, and its evaluation results are summarized in this section.

### 3.1 Results Analysis of CO model

Table.1 presents the results of the AIC, BIC, and AICc metrics. The COCUDE model has been iterated 10 items on each neuron; among that the lowest AIC iterations are finalized for each neuron. The number neuron 6 has the lowest AIC values. Such that the neuron 6 has the lowest AIC, BIC, and AICc values as 1380.5, 1385.5, and 1380.6 respectively. Its values are the lowest as compared to other neuron's stationary metrics. Therefore its corresponding neuron 6's dataset has been taken for the future prediction of CO model.

Table.1 Results of AIC, BIC, and AICc Metrics for CO model

Criterion/Neurons	AIC	BIC	AICc
2	1469.6	1477.6	1469.9
4	1483	1490.5	1483.3
<b>6</b>	<b>1380.5</b>	<b>1385.5</b>	<b>1380.6</b>
8	1423.2	1425.7	1423.3
10	1425.5	1430.5	1425.6
12	1580.2	1585.3	1580.4
14	1539.7	1544.7	1539.8
16	1520.2	1525.2	1520.3
18	1492.1	1497.2	1492.3
20	1450.5	1455.5	1450.6

Table.2 gives the performance error metrics such as MSE, RMSE, and MAE metrics resulted by the CO model. Similar to the stationary (AIC, BIC, and AICc) evaluation results, the neuron 6 has the lowest error values as compared to the results of other neurons. The neuron 6 produces the lowest errors as 1.2054, 1.0979, and 1.0693 for the MSE, RMSE, and MAE metrics respectively for the CO model.

Table.2 Performance Error Metrics resulted by CO model

Metrics/Neurons	MSE	RMSE	MAE
2	3.0678	1.7515	1.6886
4	1.5083	1.2281	1.0957
6	<b>1.2054</b>	<b>1.0979</b>	<b>1.0693</b>
8	1.6661	1.2908	1.2377
10	1.5575	2.4259	2.4088
12	3.518	1.8756	1.8016
14	2.9307	1.7119	1.6932
16	6.4948	2.5485	2.4928
18	4.6072	2.1464	1.9652
20	2.3903	1.546	1.4190

Figure 2 depicts the results of the proposed CO model for the future 90 days from the date of 10th June 2020. The results indicate the prediction of future infections in the state of Tamilnadu, India. The predicted results are validated with the official dataset and results of the CO model had produced very error values as shown in Table.2

### 3.2 Results Analysis of CU model

Table.3 presents the stationary metrics results such as AIC, BIC, and AICc for the CU model. The CU model has been iterated 10 items on each neuron; among that the lowest AIC iterations are finalized for each neuron. The number neuron 6 has the lowest AIC values. Such that the neuron 6 has the lowest AIC, BIC, and AICc values as 1479.258, 1486.791, and 1479.534 respectively. Its values are the lowest as compared to other neuron's stationary metrics. Therefore its corresponding neuron 6's dataset has been taken for the future prediction of the CU model.

Table.3 Results of AIC, BIC, and AICc Metrics for CU model

Criterion/Neurons	AIC	BIC	AICc
2	1742.149	1749.682	1742.425
4	1527.528	1535.061	1527.804
<b>6</b>	<b>1479.258</b>	<b>1486.791</b>	<b>1479.534</b>
8	1666.517	1674.049	1666.793
10	1801.991	1809.524	1802.267
12	1608.141	1615.673	1608.417
14	1534.959	1542.492	1535.235
16	1841.013	1848.546	1841.289
18	1791.097	1798.63	1791.373
20	1814.488	1822.021	1814.764

Table.4 Performance Error Metrics resulted by CU model

Metrics/Neurons	MSE	RMSE	MAE
2	6.7617	2.6003	2.5389
4	7.9903	2.8267	2.7624
<b>6</b>	<b>1.1093</b>	<b>1.0532</b>	<b>1.0322</b>
8	3.6793	1.9182	1.8653
10	6.7274	2.5937	2.639
12	1.1752	1.0841	1.0823
14	4.1006	2.0167	2.0038
16	3.3806	1.8386	1.7888
18	1.3845	1.1767	1.1232
20	3.6452	1.9092	1.8513

Table.4 gives the performance error metrics such as MSE, RMSE, and MAE metrics resulted by the CU model. The neuron 6 has the lowest error values as compared to the results of other neurons, as similar to the neuron 6's lowest AIC, BIC, and AICc evaluation results. The neuron 6 produces the lowest errors as 1.1093, 1.0532, and 1.0322 for the MSE, RMSE, and MAE metrics respectively for the CU model.

Figure 3 depicts the results of the proposed CU model for the future 90 days from the date of 10th June 2020. The results indicate the prediction of future infections in the state of Tamilnadu, India. The predicted results are validated with the official dataset and results of the CO model had produced very error values as shown in Table.4

### 3.3 Results Analysis of the DE model

Table.5 presents the stationary metrics results such as AIC, BIC, and AICc for the DE model. The DE model has been iterated 10 items on each neuron; among that the lowest AIC iterations are finalized for each neuron. The number neuron 12 has the lowest AIC values. Such that the neuron 12 has the lowest AIC, BIC, and AICc values as 585.7667, 593.2993, and 586.0426 respectively. Its values are the lowest as compared to other neuron’s stationary metrics. Therefore its corresponding neuron 12’s dataset has been taken for the future prediction of the DE model.

Table.5 Results of AIC, BIC, and AICc Metrics for DE model

Criterion/Neurons	AIC	BIC	AICc
2	595.592	599.1246	594.8678
4	707.7984	712.8201	707.9347
6	743.8817	751.4143	744.1576
8	635.7356	643.2682	636.0114
10	727.1973	734.7299	727.4732
12	<b>585.7667</b>	<b>593.2993</b>	<b>586.0426</b>
14	714.9297	722.4622	715.2055
16	661.058	668.5906	661.3338
18	696.2258	703.7584	696.5016
20	687.5292	692.5509	687.6655

Table.6 gives the performance error metrics such as MSE, RMSE, and MAE metrics resulted by the DE model. The neuron 12 has the lowest error values as compared to the results of other neurons, as similar to the neuron 12’s lowest AIC, BIC, and AICc evaluation results. The neuron 12 produces lowest errors as 3.4151, 1.8480, and 1.7925 for the MSE, RMSE, and MAE metrics respectively for the DE model.

Table.6 Performance Error Metrics resulted by DE model

Metrics/Neurons	MSE	RMSE	MAE
2	4.0340	2.0085	1.9684
4	4.1325	2.0328	2.9889
6	4.286	2.0702	2.147
8	4.7965	2.1901	2.0969
10	4.8489	2.2020	2.1658
12	<b>3.4151</b>	<b>1.8480</b>	<b>1.7925</b>
14	7.1387	2.6718	2.5672
16	5.6801	2.3833	2.3125
18	4.1346	2.0337	1.9865
20	9.4947	3.0813	3.0175

Figure 4 depicts the results of the proposed DE model for the future 90 days from the date of 10th June 2020. The results indicate the prediction of future infections in the state of Tamilnadu, India. The predicted results are validated with the official dataset and results of the CO model had produced very error values as shown in Table.2

### 3.4 Comparative Analysis of COCUDE model

The future prediction results of the COVID–19 using the COCUDE model is analyzed by using Table.7 and Table.8. Table.7 summarizes the AIC, BIC, and AICc metrics for COCUDE model. In Table.7, the neuron 6 had given the lowest AIC, BIC, and AICc metrics as 1380. 5, 1385.5, and 1380.6 respectively for the CO model; for the CU model its values are 1479.258, 1486.791, and 1479.534 respectively for the neuron 6. Similarly, the neuron 12 had given the lowest AIC, BIC, and AICc metrics as 585.7667, 593.2993, and 586.0426 respectively for the DE model using artificial neural network algorithm.

Table.7 AIC, BIC, and AICc Metrics for COCUDE model

Model	Criterion/Neurons	AIC	BIC	AICc
CO	6	<b>1380. 5</b>	<b>1385.5</b>	<b>1380.6</b>
CU	6	<b>1479.258</b>	<b>1486.791</b>	<b>1479.534</b>
DE	12	<b>585.7667</b>	<b>593.2993</b>	<b>586.0426</b>

Table.8 summarizes the MSE, RMSE, and MAE metrics for COCUDE model. In Table.7, the neuron 6 had

given the lowest MSE, RMSE, and MAE metrics as 1.2054, 1.0979, and 1.0693 respectively for the CO model; for the CU model its values are 1.1093, 1.0532, and 1.0322 respectively for the neuron 6. Similarly, the neuron 12 had given the lowest MSE, RMSE, and MAE metrics as 3.4151, 1.8480, and 1.7925 respectively for the DE model using artificial neural network algorithm.

Table.8 Performance Error Metrics resulted by COCUBE model

Model	Metrics/Neurons	MSE	RMSE	MAE
CO	6	1.2054	1.0979	1.0693
CU	6	1.1093	1.0532	1.0322
DE	12	3.4151	1.8480	1.7925

## 4. Conclusion

This research work had focused on the prediction future COVID–19 population in the state of Tamilnadu, India using the artificial neural network-based machine learning algorithms. The proposed COCUBE model has been trained and testing using the given official dataset. The predicted dataset of the presented COCUBE model has been tested and it gives the lowest error metric values. The presented CO and CU model had produced the lowest MSE error metrics values as 1380.5, and 1479.258 respectively for the neuron 6. Also, the neuron 6 had outperformed these for these two models as compared to neurons. Further, the neuron 6 of CO and CU model gives the lowest AIC values as 1.2054 and 1.1093 respectively with the highest stationary and quality. Similarly, for the DE model, the neuron 12 had offers the lowest MSE metrics with 3.4151 as well as lowest AIC value with 585.7667 as compared to other neurons by employing the artificial neural network with NARX algorithm. Therefore, the proposed COCUBE model predicts the precise future infection cases, cured cases, and decease cases. Such that the results of the proposed COCUBE model might more helpful for respective authorities to plan future any necessities because of the COVID–19 pandemic.

## Declarations

Competing interests: The authors declare no competing interests.

## References

1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KS, Lau EH, Wong JY, Xing Early transmission dynamics in Wuhan, China, of novel coronavirus–infected New England Journal of Medicine. 2020 Jan 29.
2. World Health Organisation (2020). Coronavirus disease 2019 (COVID-19): [Press release].

3. European Centre for Disease Prevention and Control (2020). Novel coronavirus disease 2019 (COVID19) pandemic: increased transmission in the EU/EEA and the UK –: [Press release]. ECDC, Stockholm.
4. Tang Z, Li X, Li Prediction of new coronavirus infection based on a modified SEIR model. medRxiv. 2020 Jan 1.
5. Ayubali AA, Satheesh SR. On predicting the novel COVID-19 human infections by using Infectious Disease modelling method in the Indian State of Tamil Nadu during 2020. medRxiv. 2020 Jan
6. Li L, Yang Z, Dang Z, Meng C, Huang J, Meng H, Wang D, Chen G, Zhang J, Peng H, Shao X. Propagation analysis and prediction of the COVID-19. Infectious Disease Modelling. 2020 Jan 1;5:282-92.
7. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, Bleicker T, Brünink S, Schneider J, Schmidt ML, Mulders DG. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Eurosurveillance. 2020 Jan 23;25(3):2000045.
8. Ranjan R. Predictions for COVID-19 outbreak in India using Epidemiological models. medRxiv. 2020 Jan
9. Gupta R, Pal Trend Analysis and Forecasting of COVID-19 outbreak in India. medRxiv. 2020 Jan 1.
10. Read JM, Bridgen JR, Cummings DA, Ho A, Jewell CP. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. MedRxiv. 2020 Jan 1.
11. Shao Y, Wu J. IDM editorial statement on the 2019-nCoV. Infectious Disease Modelling. 2020;5:233.
12. Hethcote HW. The mathematics of infectious diseases. SIAM review. 2000;42(4):599-653.
13. Li MY, Wang L. Global stability in some SEIR epidemic models. In Mathematical approaches for emerging and reemerging infectious diseases: models, methods, and theory 2002 (pp. 295-311). Springer, New York,
14. Zhou X, Cui J. Analysis of stability and bifurcation for an SEIR epidemic model with saturated recovery rate. Communications in nonlinear science and numerical simulation. 2011 Nov 1;16(11):4438-50.
15. Levin SA. New directions in the mathematics of infectious disease. In Mathematical approaches for emerging and reemerging infectious diseases: models, methods, and theory 2002 (pp. 1-5). Springer, New York, NY
16. <https://api.covid19india.org/csv/>
17. <https://nhmt.n.maps.arcgis.com/apps/opsdashboard/index.html#/095ad0a1c0254b058fa36b32d1ab1977>
18. Theerthagiri, (2019). CoFEE: Context-aware futuristic energy estimation model for sensor nodes using the Markov model and autoregression. International Journal of Communication Systems, e4248
19. Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi Application of the ARIMA model on the COVID-2019 epidemic dataset. Data in brief. 2020 Feb 26:105340.

20. Anne R. ARIMA modelling of predicting COVID-19 infections. medRxiv. 2020 Jan

21. Theerthagiri, P. FUCEM: futuristic cooperation evaluation model using Markov process for evaluating node reliability and link stability in mobile ad hoc Wireless Netw (2020).

## Figures

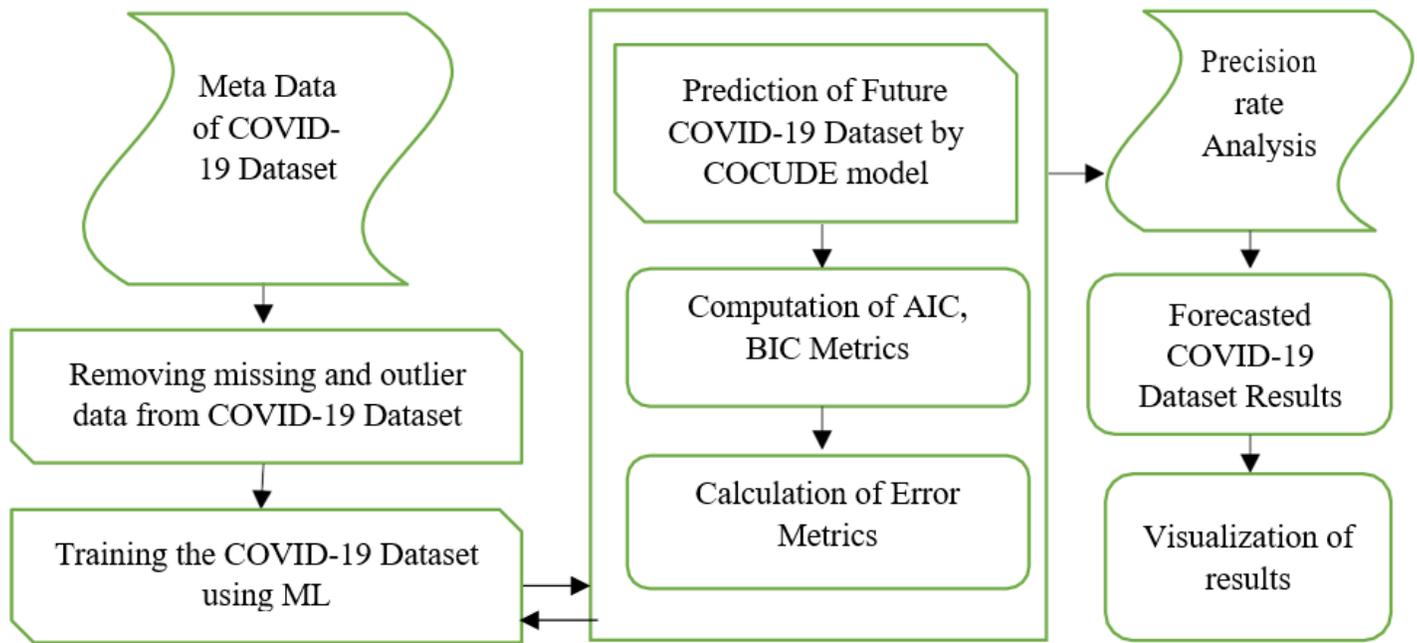


Figure 1

Block diagram of the proposed COCUDE model

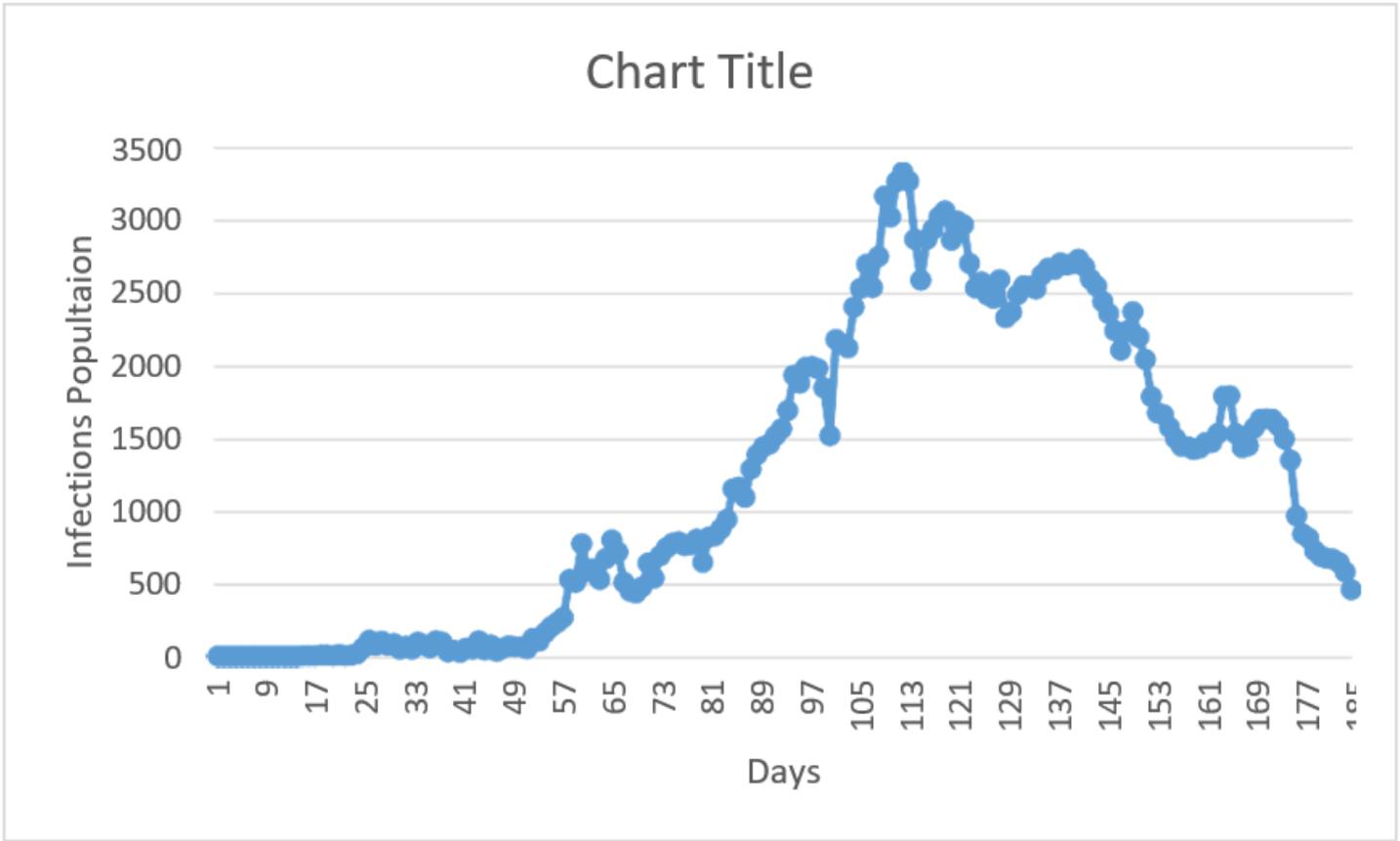
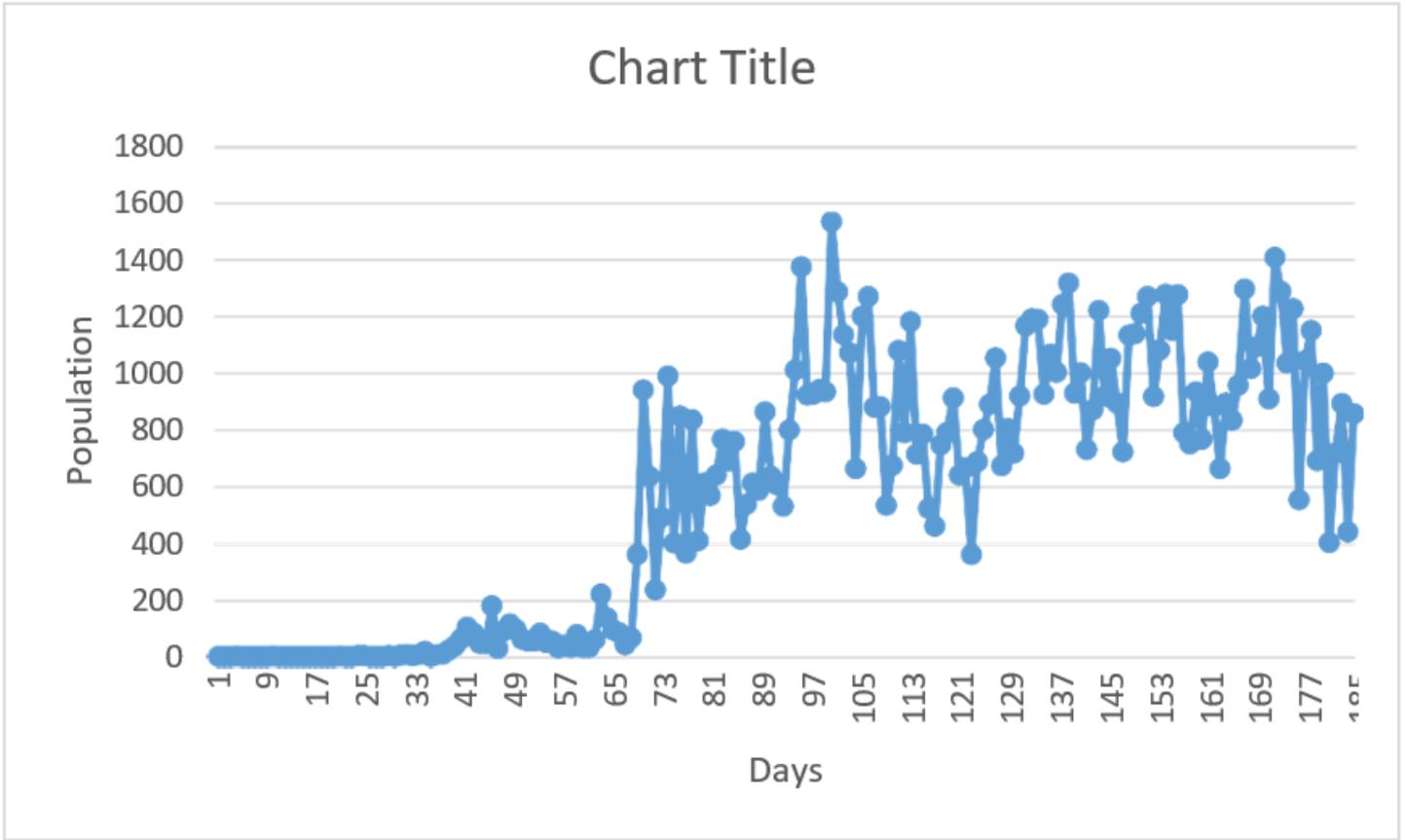


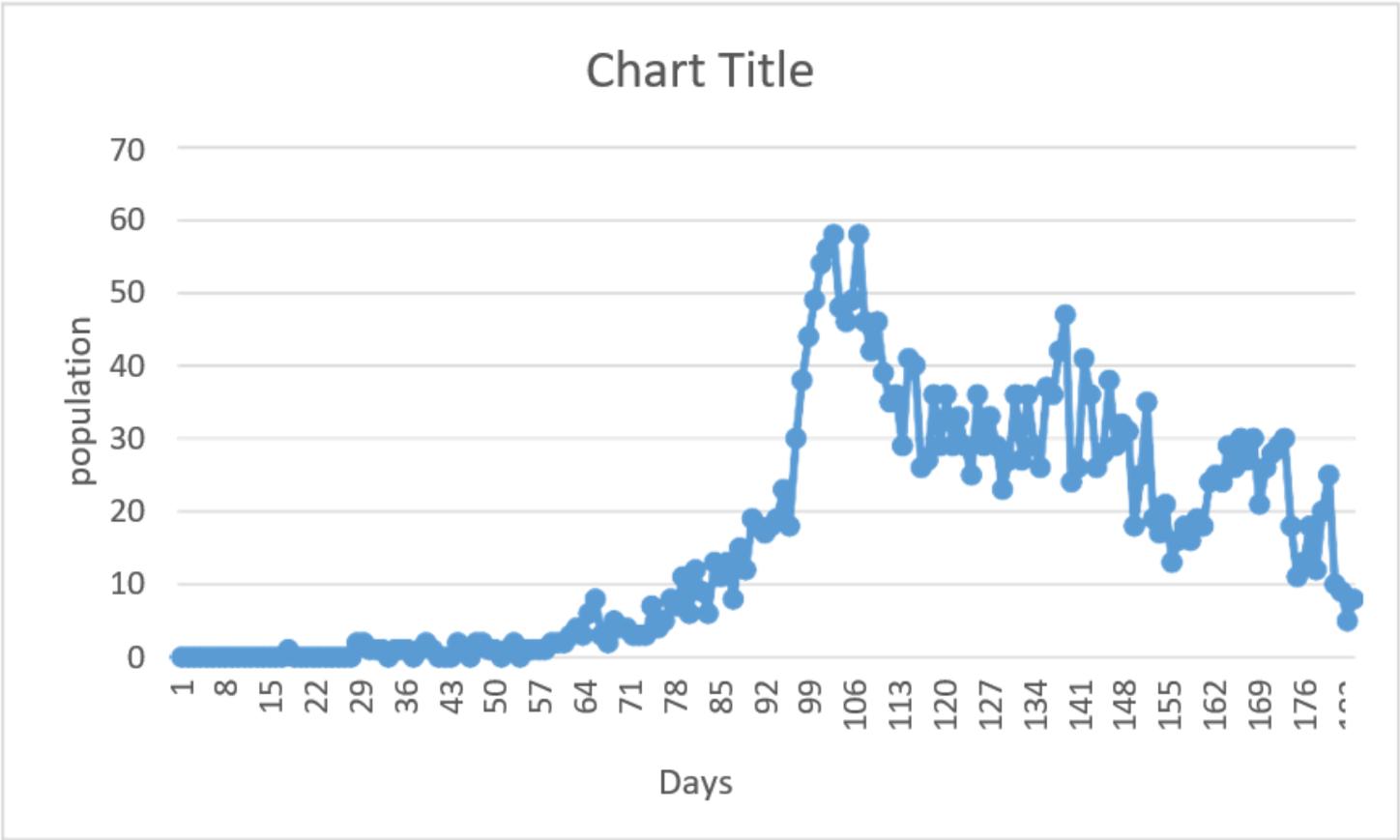
Figure 2

Future Predicted COVID-19 infection rate by CO model



**Figure 3**

Future Predicted COVID-19 Curing rate by CU model



**Figure 4**

Future Predicted COVID-19 Decease rate by DE model