Chapter 10
Impact of COVID-19 in India and Its Metro Cities: A Statistical Approach

Radha Gupta, Kokila Ramesh, N. Nethravathi, and B. Yamuna

Abstract The infectious coronavirus disease is spreading at an alarming rate, not only in India but also globally too. The impact of coronavirus disease (COVID- 19) outbreak needs to be analyzed statistically and modelled to know its behaviour so as to predict the same for future. An exhaustive statistical analysis of the data available for the spread of this infection, specifically on the number of positive cases, active cases, death cases and recovered cases, and connection between them could probably suggest some key factors. This has been achieved in this paper by analyzing these four dominant cases. This helped to know the relationship between the current and the past cases. Hence, in this paper, an approach of statistical analysis of COVID-19 data specific to metropolitan cities of India is done. A regression model has been developed for prediction of active cases with 10 lag days in four metropolitan cities of India. The data used for developing the model is considered from 26th April to 31st July (97 days), tested for the month of August. Further, an Artificial Neural Network (ANN) model using back propagation algorithm for active cases for all India and Bangalore has been developed to see the comparison between the two models. This is different from the other existing ANN models as it uses the lag relationships to predict the future scenario. In this case, data is divided into training, validation and testing sets. Model is developed on the training sets and is checked on the validation set, tested on the remaining, and then, it is implemented for prediction.

Keywords COVID-19 · Data analysis · ANN modelling · Prediction

---

R. Gupta (✉) · N. Nethravathi · B. Yamuna
Department of Mathematics, Dayananda Sagar College of Engineering, Bangalore, Karnataka, India
e-mail: radha.gaurav.gupta@gmail.com

N. Nethravathi
e-mail: nandukowshik@gmail.com

B. Yamuna
e-mail: yamunabmsc@gmail.com

K. Ramesh
Department of Mathematics, FET, Jain (Deemed-to-be-University), Bangalore, Karnataka, India
e-mail: r.kokila@jainuniversity.ac.in

N. H. Shah and M. Mittal (eds.), *Mathematical Analysis for Transmission of COVID-19*, Mathematical Engineering, https://doi.org/10.1007/978-981-33-6264-2_10

Mathematics Subject Classification 60G25 · 62J05 · 62M10

Introduction

The newly discovered coronavirus disease (COVID-19) is infectious disease caused with common symptoms such as fever, tiredness, body aches, nasal congestion, runny nose, sore throat or diarrhoea and dry cough. The outbreak of coronavirus disease (COVID- 19) has happened in China in December 2019, the first case being found in Wuhan. In March, it was declared as pandemic by the World Health Organization (WHO) exposing the world to public health emergencies.

Several preventive measures have been taken by governments' of various countries across the globe to prevent the spread of COVID infection. These preventive measures include social distancing, wearing the masks, face shields, frequent sanitization, quarantine in case of travel or home isolation in case of suspected symptoms. In India, first confirmed case happened to be in Kerala on 30-Jan-2020. The man, who was studying in WUHAN University and had travelled to India, tested positive for the virus. India's first death was confirmed in Karnataka (Kalaburgi) on 12-Mar-2020. The man who returned from Saudi Arabia and had a history of Hypertension, Diabetes and Asthma succumbed to the disease.

Some of the worst affected cities in India are Delhi, Mumbai, Chennai and Bangalore. Bangalore's first case got confirmed on 08-Mar-2020, when a software engineer who returned from Austin, US along with his wife and daughter tested positive. The first death case was of a 70 year old woman from Chikkabalapur in Bangalore, on 24-Mar-2020. She had travelled to Mecca and arrived in India on 14 March.

Chennai's first COVID case was confirmed on 07-Mar-2020, when a man travelled to Chennai from Oman. He was admitted to Hospital on 05-Mar-2020 with complaints of fever and cough, and finally, the reports on 7 March showed that he was positive. The first death cases of Chennai confirmed on 5-Apr-2020, where a 71 year old man from Ramanathpuram and 60 year old man from Washermanpet died at Government hospital, Chennai.

Mumbai's first confirmed case was on 11-Mar-2020. A couple from Andheri had tested positive post their return from a Dubai and Abu Dhabi trip. The first death case was confirmed on 17-Mar-2020, where a 71 year old man with a history of high blood pressure returned from Dubai. He developed pneumonia and inflammation of heart muscles and increased heart rate leading to death.

A 45 years old person from East Delhi, with a history of travel from Italy, was the first confirmed case in Delhi on 02-Mar-2020. Delhi's first death was of a 68 years old woman who got the virus from her son who returns from Switzerland. It was confirmed on 14-Mar-2020.

Researchers are trying to contribute their bit in every possible way that can lead to some solution. While conventional methods are precise and deterministic, artificial intelligence (AI) techniques could give high-quality predictive models. In this study, authors use a publically available dataset from which contains information on positive cases, recovered cases, active cases and death cases in four metropolitan cities of India over 97 days (from 26th April to 31st July 2020). In the present study, a detailed statistical analysis has been performed on the data procured. It is observed that the data has a strong relationship with past days, an autoregressive model for active cases with 10 lag days data as input and an artificial neural network (ANN) model to extract the non-linearity between the data if it exists has been developed and verified both the models in the training as well as testing periods. Needless to say that the methods leading to reliable prediction of spreading of COVID-19 would be of big help in taking preventive measures to minimize its spread, deaths, active cases and maximizing recovery cases.

Literature

Ahmed [1] did an exhaustive review to understand the epidemiological evidences, clinical manifestations, investigations and treatment given to COVID cases who are admitted in various hospitals of Wuhan city and other parts of China.

Luo et al. [2] made an effort to expand screening capacity, reviewed advances and challenges in the rapid detection of COVID-19 by targeting nucleic acids, antigens or antibodies. They summarized some of the effective treatments and vaccines against COVID-19. They also discussed about possible reduction of viral progression post ongoing clinical trials of interventions.

Poletto et al. [3] published a comment, highlighting some of the important discoveries as a result of predictive modelling to diverse data sources. These results had an impact on clinical and policy decisions.

Through link [4], one will find expert, curated information on SARS-CoV-2 (the novel coronavirus) and COVID-19 (the disease), that will help the research and health community to work together. All these resources are free to access and include clear guidelines for clinicians and patients.

Shah et al. [5] proposed a generalized SEIR model of COVID-19 to study the behaviour of its transmission under different control strategies. This model considered all possible cases, where transmission happens from one human to another and formulated its reproduction number to analyze the accuracy of transmission dynamics of the coronavirus outbreak. Further, they applied optimal control theory to demonstrate the impact of various intervention strategies, people in quarantine and isolation of infected individuals, immunity boosters and hospitalization.

An epidemic model describing its spread in a population was formulated by Arino and Portet [6]. This model considered an Erlang distribution of times of sojourn in incubating, symptomatically and asymptomatically infectious compartments.

Fong et al. [7] proposed a methodology that embraces three virtues, (1) augmentation of existing data, (2) selecting a panel to pick the best forecasting model from many existing models and (3) tweaking the parameters of an individual forecasting model so that the accuracy of data mining is highest possible.

Shah et al. [8] further made an attempt to assess the impact of inter-state, foreign travel and public health interventions imposed by the US Government in response to the COVID-19 pandemic. They developed a disjoint mutually exclusive compartmental to study the transmission dynamics of the coronavirus. Formulation of system of non-linear differential equations, computation of basic reproduction number R_0 and stability of the model at the equilibrium points was discussed in detail.

A visionary perspective on data usage and management for infectious diseases is provided by Wong et al. [9]. They highlighted that there is ample opportunity for researchers to make use of artificial intelligence methods to enable reliable and data-oriented disease monitoring in this information age. It is concluded that together with reliable data management platforms AI methods will enable effective analysis of infectious disease. It will also provide surveillance data to support risk and resource analysis for government agencies, healthcare service providers and medical professionals in the future.

Dey [10] developed a time series model for number of total infected cases in India, considering data from 3rd to 7th March 2020. They had developed two models during the initial days of COVID which were discarded because they lost their statistical validity. But later on they developed another model as a third degree polynomial that has remained stable since 8 Apr , with $R^2 > 0.998$ consistently. This model is used for forecasting total number of confirmed COVID cases after cautionary discussion of triggers that would invalidate the model.

Hu et al. [11] also proposed the artificial intelligence (AI)-inspired methods for real-time forecasting of COVID-19 for estimating the size, lengths and ending time of COVID-19 across China. They developed a modified stacked auto-encoder for modelling the transmission dynamics of the epidemics and applied this model for forecasting the real-time confirmed cases of COVID-19 across China. The data collected for this study varied from 11 January to 27 February, 2020 from WHO.

Car et al. [12] transformed a time series dataset into a regression dataset and used it in training a multilayer perceptron (MLP) artificial neural network (ANN). By training this dataset, they tried to achieve a worldwide model of the maximal number of patients across all locations in each time unit. Hyper parameters were varied using a grid search algorithm, and a total of 48,384 ANNs were trained. Their study models showed high robustness of the deceased patient model, good robustness for confirmed and low robustness for recovered patient model.

For our present study, we collected data from [13–15].

Data

There are various sources that are tracking the coronavirus data. They are updated at different times and are gathered in different ways, so the data might differ from source to source. As on 21st August 2020, WHO website quotes 21,294,845 confirmed cases, 761,779 confirmed deaths, total of 216 countries/territories affected with this respiratory disease. According to revised guidelines on public health surveillance for

COVID-19 by WHO (on 13-08-2020), emphasis should be on information on the importance of the collection of metadata for analysis and interpretation of surveillance data. The data has been collected from [13–15], for the purpose of studying the trend in four urban cities Bangalore, Delhi, Chennai and Mumbai along with India. Four major parameters, namely confirmed, recovered, active and death cases have been considered.

The basic statistics of the data for the period of 113 days (26 April to 17 August) for four metropolitan cities of India and India as a whole which is given in Table 10.1. The data has been plotted to see the pattern if visible in figure and also for the visual appreciation of the distribution in Fig. 10.1.

It is clearly visible from the basic statistics mentioned in Table 10.1 that data considered for the present study is non-Gaussian in nature as in all the cases, the skewness is not nearly or equal to 0 except for Indian case, but even in this case, the Kurtosis is not near 3. The most important property of COVID-19 data is its non-Gaussian nature. Hence, even though the mean and the standard deviation are valuable descriptors, when questioned about the severity of the spread, the assumption of normality (Gaussianity) will not be applicable. This can further be seen in the form of the probability distributions. The data of the four cases are normalized using the relation $D_i = (d_i - m_i)/s_i$ where D_i is the normalized data, d_i is the actual data, m_i is the average of the given sample length, and s_i is the standard deviation of the

Table 10.1 Basic statistics of COVID-19 data for all different types such as positive, active, recovered and death cases of all India and its four important metropolitan cities for the period of 113 days from 26 April to 17 August (*Source* Own)

S. no.		1	2	3	4	5
Place		India	Bengaluru	Chennai	Mumbai	Delhi
Average	Positive cases	484,809	9151	39,060	56,590	55,464
	Active cases	184,427	6753	11,571	21,021	14,817
	Recovered cases	287,561	2219	26,867	32,605	37,976
	Death cases	12,821	180	626	2876	1651
Standard deviation	Positive cases	456,229.05	15,259.24	33,246.22	34,419.53	47,213.22
	Active cases	145,474.97	11,104.84	7018.75	7142.94	9065.65
	Recovered cases	300,728.24	3928.90	27,734.52	28,105.36	40,847.39
	Death cases	10,593.51	294.95	660.42	2148.90	1462.78
Skewness	Positive cases	0.07	1.72	−1.31	−1.32	−1.45
	Active cases	−0.02	1.72	−1.14	−0.15	−1.35
	Recovered cases	0.15	3.68	−0.81	−1.17	−0.95
	Death cases	−0.80	1.73	−0.40	−1.59	−1.61
Kurtosis	Positive cases	1.05	1.72	0.44	0.05	0.41
	Active cases	0.95	1.77	0.03	−0.99	0.16
	Recovered cases	1.10	2.11	0.79	0.47	0.79
	Death cases	0.67	2.36	0.91	0.24	0.27

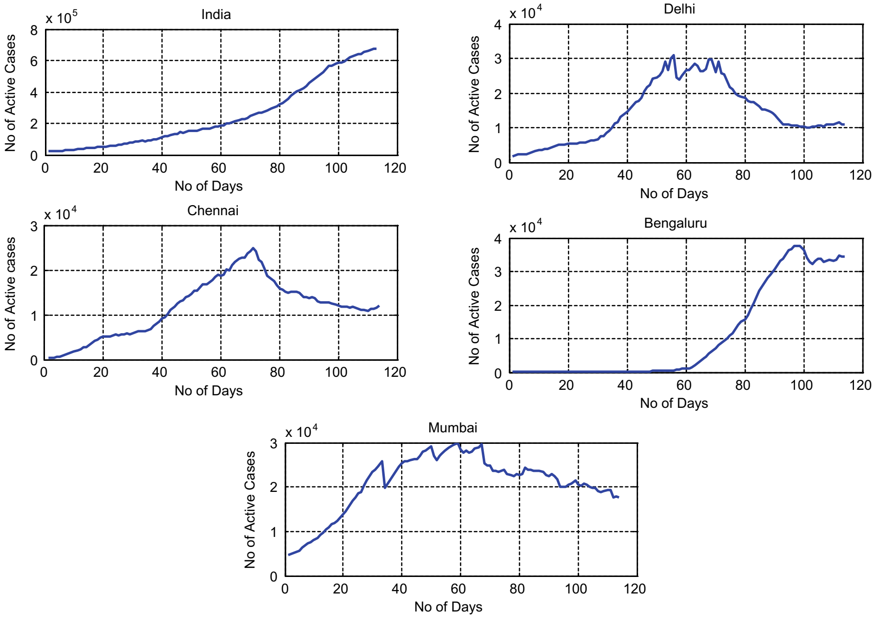

Fig. 10.1 Observed data plot of active cases in India and four metropolitan cities, namely Bengaluru, Chennai, Mumbai and Delhi for the period of 113 days (*Source* Own)

given sample length for each *i* for the period of 113 days from 26 April to 17 August. Even after normalizing the data, the skewness and kurtosis remain the same. The data distribution of the normalized data has been plotted as a histogram to see patterns in the distribution so as to choose the model appropriately and is shown in Fig. 10.2.

A common assumption that is popular with any time series data is to consider it to be a stationary random process. This helps in defining the long-term average and long-term deviation which remain as reference values in modelling and forecasting exercises. For a stationary process, basic statistical parameters such as mean and standard deviation of the long period remain time independent. However, if they vary widely over a period of time, then the stationary assumption will not be valid. In Fig. 10.3, the non-stationarity of long-term average and long-term deviation of COVID-19 active cases data for all India, and its metropolitan cities is shown by changing the sample length. In all cases, the number of samples does not lead to a constant value of the average. The data considered for modelling purpose in the present study focuses on active cases of all India and the four metropolitan cities mentioned in Table 10.1.

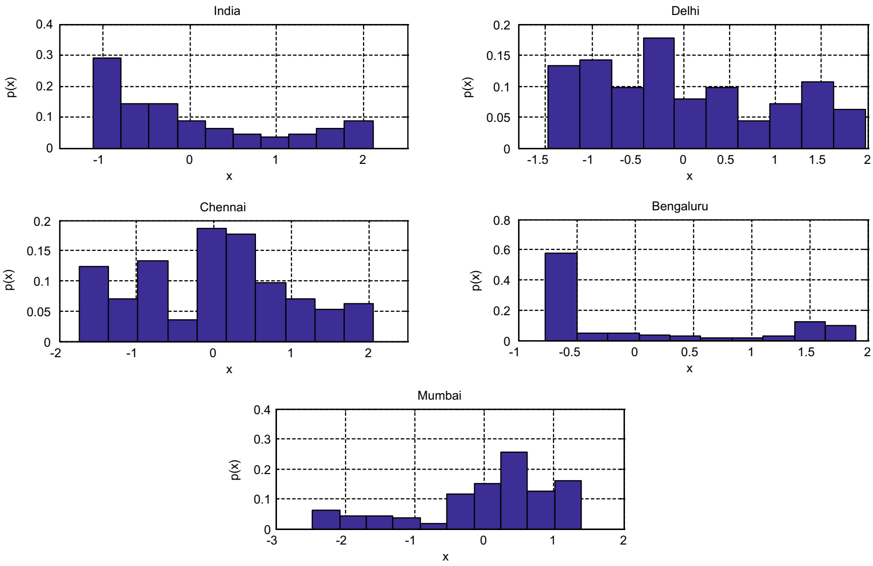

Fig. 10.2 Histogram plot of COVID-19 active cases of all India and four metropolitan cities mentioned in the diagram (*Source* Own)

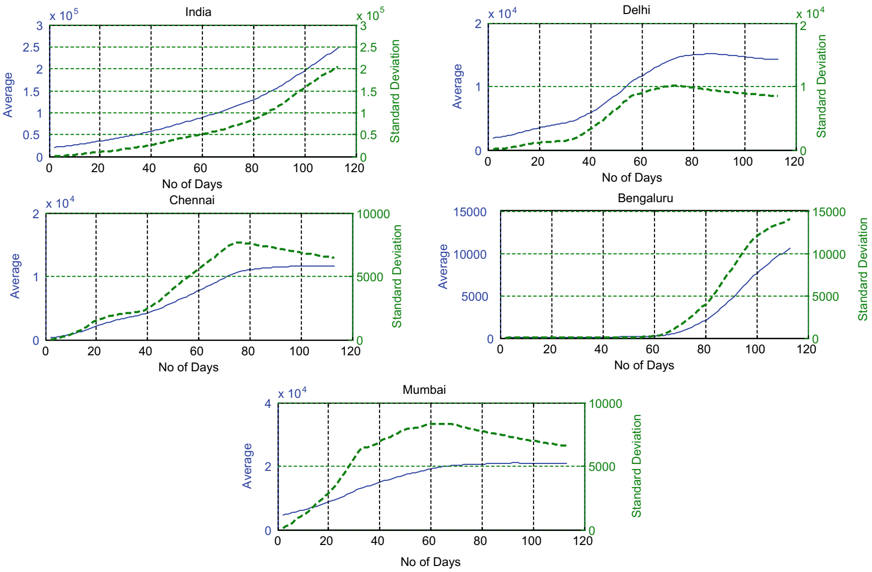

Fig. 10.3 Running mean and the standard deviation of COVID-19 data of the active cases for the data from 26 April to 17 August of 2020 for all India and four metropolitan cities of India (*Source* Own)

Methodology

To model any data, one has to understand the hidden structure or pattern in the data. In order to understand the data well, specifically for the active cases, data has been normalized using its own data and standard deviation. Also, detailed analysis of the same will help in understanding the pattern. The autocorrelation function indicates that a strong connection is there in data lags at least until 10 days lag as shown in Fig. 10.4. The data has been divided into training period (26 April to 31 July) with the sample size 97 and testing period (1 August to 31 August) with the sample size 31. In the training period, data will be trained for a particular model with the appropriate parameters with the number of parameters being less than 50% of the sample size; otherwise, it will be a polynomial fit for the entire length. Model will be validated using the measures such as the root mean square error and the coefficient of determination. If all the measures stay well within the confidence bands, then it will be implemented and checked again in the testing period mentioned.

Model 1: (Auto-Regressive Model)

Based on the autocorrelation function plotted in Fig. 10.4, an auto-regressive (AR) model considering the past 10 lag days in the regression equation as the variable is constructed for the present active cases. This is due to a strong correlation which exists in 10 lag days, beyond which it starts reducing. For the sample size considered for the

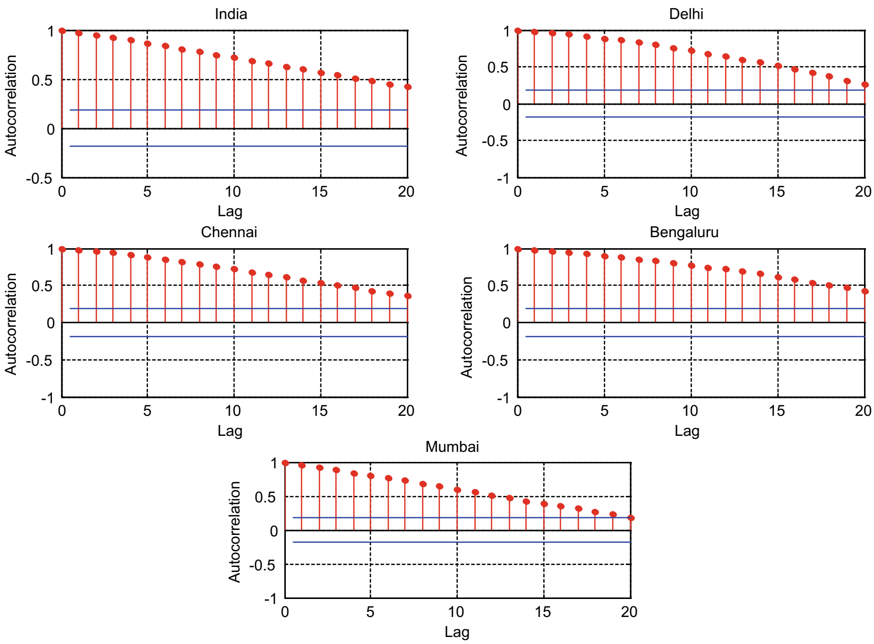

Fig. 10.4 Sample Autocorrelation function of COVID-19 active cases of all India and its four important metropolitan cities (*Source* Own)

modelling purpose, a significant correlation is 0.6. Hence, until 10 lag information has been incorporated in modelling the active cases of all India and the four metropolitan cities considered for the present study. As mentioned in the previous section, the data has been trained using this regression model for the sample size of 97 days (from 26 April to 31 July) using the following equation:

$$\begin{aligned} A_t = {} & B0 + B1A_{t-1} + B2A_{t-2} + B3A_{t-3} + B4A_{t-4} + B5A_{t-5} \\ & + B6A_{t-6} + B7A_{t-7} + B8A_{t-8} + B9A_{t-9} + B10A_{t-10} \end{aligned} \tag{10.1}$$

Here, A represents active cases and t in days. The parameters in equation for all India and four metropolitan cities during training the model have been given in Table 10.2.

Comparison between the actual data, i.e. number of active cases and the AR model fit is shown in Fig. 10.5. Also the basic statistics such as average, standard deviation have been matched with the model, and the measures such as correlation coefficient (CC) have been found and listed in Table 10.3. In Fig. 10.5, it can be clearly seen that the model exactly matches with the observed data, and hence, the same be tested in the testing period data for using it in forecasting.

Model 2: Artificial Neural Network (ANN)

The non-Gaussianness of the data cannot be ignored at this moment as there can be some non-linearity hidden in the data; it is seen in the data distribution plotted as histograms. Hence, the artificial neural network (ANN) model has been used here with 6 days lag as the inputs in the input layer with one hidden layer and one output in the output layer. Network has been trained in the training period using back propagation algorithm using sigmoid function. There are totally 49 weights used while training the network. Network used for modelling is shown in Fig. 10.6. In the diagram, I represents the input layer, H represents the hidden layer, and O represents the output layer.

The network has been experimented only for two regions active cases data, i.e. for all India and Bengaluru active cases to compare with the regression model used in the previous section. The comparison between the actual data and the network model for these two cases is shown in Fig. 10.7.

The moment comparison is shown in Table 10.4 for ANN model. It can be clearly observed that the comparison between the actual data and the network fit is really appreciable even though only 6 lag information was used in the network for training the entire length data. Both model 1, i.e. Auto-regressive model and ANN model are performing good in the training period; check has to be in the testing period. The model which performs better in the testing period can be considered for future forecasting.

Table 10.2 Parameters of the regression Eq. (10.1) for all India and four metropolitan cities of India (*Source* Own)

Place	B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10
India	0.0248	1.1170	0.0765	−0.1507	0.0041	0.0203	−0.0860	0.3402	−0.4715	0.1052	0.0648
Delhi	0.0029	0.8634	0.0651	0.2501	−0.1855	0.0184	−0.0266	0.2158	−0.0411	−0.0268	−0.1653
Chennai	0.0069	1.3064	−0.1226	−0.0258	−0.1571	0.1233	−0.1875	0.0108	0.1465	−0.2535	0.1467
Bengaluru	−0.0027	1.7168	−0.7592	0.3016	−0.3415	0.0848	0.0030	0.6354	−1.2750	0.8317	−0.2066
Mumbai	0.02287	0.9299	0.04812	−0.0023	−0.0754	0.0611	−0.1051	0.0853	0.2495	−0.0926	−0.1512

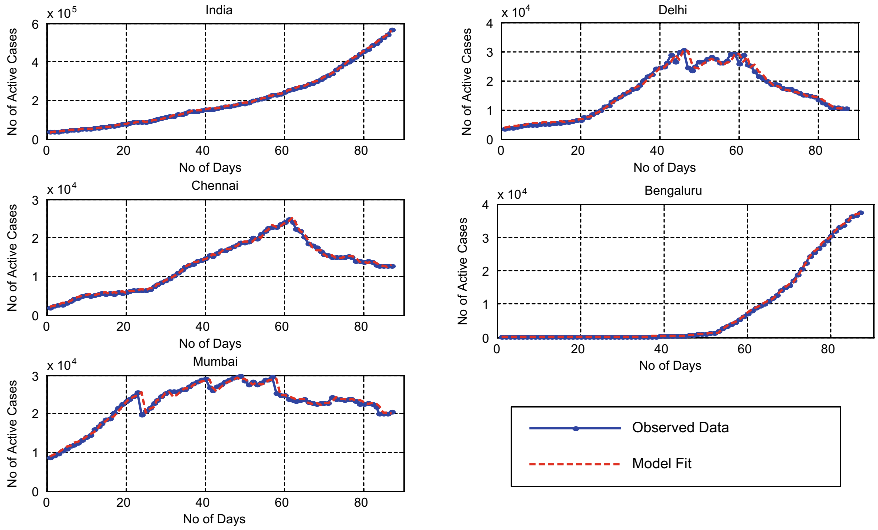

Fig. 10.5 Comparison between the observed data and the mode fit of active cases of all India and four metropolitan cities (*Source* Own)

Table 10.3 Comparison between the observed data and the model parameters for the training period (*Source* Own)

Parameters		India	Delhi	Chennai	Bengaluru	Mumbai
Average	Actual	202,624	16,238	12,802	7522	22,728
	Model fit	202,624	16,238	12,802	7522	22,728
Standard deviation	Actual	142,731.8576	8479.2639	6332.0540	11,483.2192	5307.9156
	Model fit	142,715.1130	8395.2372	6318.1238	11,481.7111	5218.2834
Correlation coefficient	Between actual and fit	0.99	0.98	0.98	0.99	0.98

Results and Discussion

It is observed that both AR model and ANN model have performed nearly same in the modelling or training period. In this section, both the models will be compared in the testing period; both the models have its own advantages and their disadvantages; one would be interested in the model which performs better in the testing period of 31 days (1 Aug to 31 Aug). AR model has been tested for all the five regions considered in the present study, whereas ANN model is performed only for two regions (All India and Bengaluru). In Table 10.5, the comparison between the AR model and ANN model for the two regions is shown:

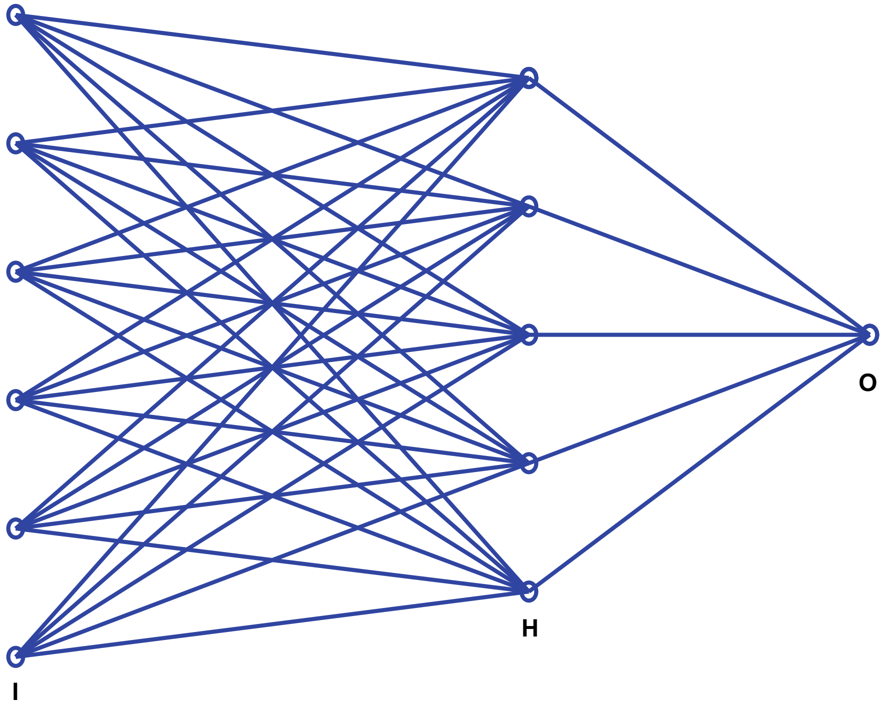

Fig. 10.6 Network used with six inputs in the input layer (I), one hidden layer with six neurons (H) and one output in output layer (O) (*Source* Own)

Day-wise forecast comparison has been listed in Table 10.5 for number of active cases comparison with the observed data so as to check each day whether it is nearly matching with the actual data or not. This may not be the correct measure to show the comparison. Hence, comparison in terms of root mean square error (RMSE) between observed data and the forecast by AR and ANN models, correlation coefficient (CC) between actual data and forecast and the performance parameter (PP) between the actual data and forecast has been given in Table 10.6 which is considered to be the best parameters for the comparison.

It is clearly visible that training period and testing periods will not have the same relations as in terms of the measures which are used to compare the model in the forecast period as well as the training periods. Hence, one has to definitely check the measures in both the periods. But in the present cases, specifically in two cases, i.e. in India and Bengaluru, both methods perform similar as the result shows with minor variations in the measures, so any model can be used to forecast the present situation of the active cases. Similarly, in the remaining cases, only AR model has been tried and tested for the parameters which are well within the significance bands of their nature. Hence, the models can be accepted for modelling as well as forecasting

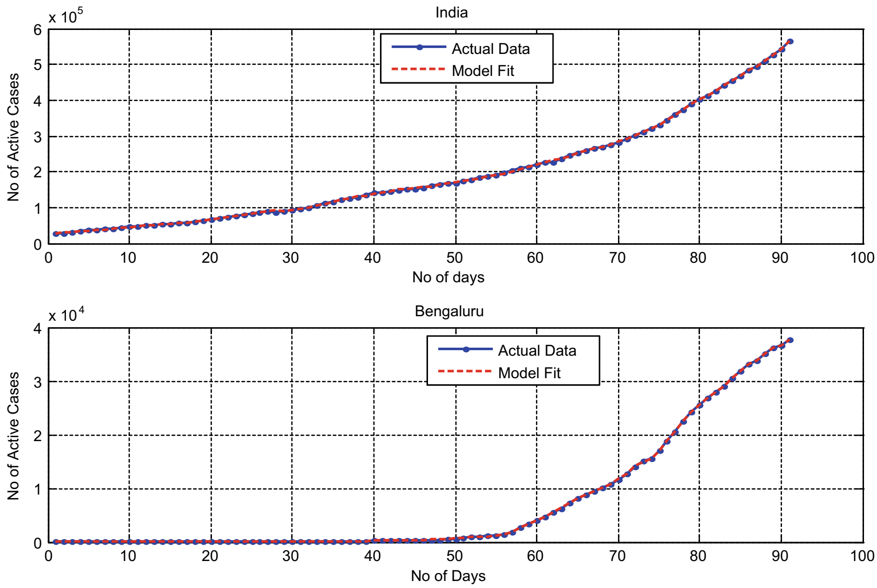

Fig. 10.7 Comparison between the actual data and the model fit using ANN model

Table 10.4 Comparison between the observed data and the model parameters for the training period (*Source* Own)

Parameters		India	Bengaluru
Average	Actual	195,063	7526
	Model fit	195,023	7520
Standard deviation	Actual	143,959.27	11,709.33
	Model fit	143,907.42	11,709.42
Correlation coefficient	Between actual and fit	0.99	0.99
Performance parameter	Between actual and fit	0.99	0.99

purposes of COVID-19 data. Also, due to limitation of time and also due to restriction of the work, model has been tried only on the active cases.

Conclusion

In the present study, AR and ANN models have been tried on the present situation of COVID-19 pandemic, specifically on the active cases. Before applying these models, an exhaustive statistical analysis has been performed on the data due to understand the nature and pattern of the data. Stationarity test has been performed by plotting

Table 10.5 Day-wise comparison of the forecast for the period of 31 days (1 Aug to 31 Aug) using AR and ANN model with observed data (*Source* Own)

Day	India			Bengaluru			Chennai		Mumbai		Delhi	
	Actual data	AR forecast	ANN forecast	Actual data	AR forecast	ANN forecast	Actual data	AR forecast	Actual data	AR forecast	Actual data	AR forecast
1 Aug	567,730	581,773	570,306	37,760	38,499	37,129	12,439	12,662	20,731	20,692	10,596	10,476
2 Aug	579,357	586,262	576,822	37,513	37,847	37,912	12,193	12,438	21,394	21,032	10,356	10,316
3 Aug	586,298	593,057	583,934	36,290	37,232	37,702	11,986	12,063	20,528	21,697	10,207	10,225
4 Aug	586,244	602,699	598,355	34,021	35,780	35,939	11,859	11,832	20,309	20,512	9897	10,014
5 Aug	595,501	602,137	598,516	32,757	32,143	28,487	11,814	11,852	20,679	19,945	10,072	10,055
6 Aug	607,384	608,994	595,085	32,314	31,118	32,782	11,723	11,753	20,546	20,513	10,348	9866
7 Aug	619,088	625,355	610,965	33,308	32,013	33,141	11,609	11,712	20,124	20,802	10,409	10,156
8 Aug	628,747	630,759	624,741	33,726	33,056	33,134	11,737	11,581	19,914	20,389	10,667	10,368
9 Aug	634,945	646,127	637,378	33,815	33,631	33,892	11,657	11,755	19,700	20,279	10,729	10,541
10 Aug	639,929	650,812	637,362	32,985	33,234	34,600	11,331	11,713	19,172	20,141	10,346	10,761
11 Aug	643,948	652,994	641,410	33,070	31,369	33,093	11,133	11,291	18,887	19,303	10,868	10,794
12 Aug	653,622	660,845	656,412	33,489	33,236	33,822	10,956	11,055	19,047	18,974	10,946	10,485
13 Aug	661,595	670,546	662,099	33,148	33,344	32,779	10,871	10,900	19,314	19,391	10,975	10,966
14 Aug	667,950	678,252	669,630	33,432	33,373	33,048	11,212	10,812	19,337	19,626	11,366	11,048
15 Aug	677,714	683,705	674,768	34,858	33,039	34,700	11,324	11,279	17,591	19,504	11,489	11,323
16 Aug	676,900	693,145	679,620	34,584	35,470	34,267	11,501	11,484	17,825	17,896	10,823	11,593
17 Aug	673,213	692,458	675,657	34,408	33,657	33,548	12,006	11,657	17,704	18,034	10,852	11,693
18 Aug	676,549	686,458	670,446	33,081	34,968	33,974	12,106	12,250	17,693	17,830	11,068	11,104
19 Aug	686,395	691,832	683,454	33,280	31,726	33,598	12,259	12,395	17,914	17,942	11,137	11,126
20 Aug	692,030	702,187	691,463	34,186	32,729	33,575	12,290	12,467	18,172	18,189	11,271	11,127

(continued)

Table 10.5 (continued)

Day	India			Bengaluru			Chennai		Mumbai		Delhi	
	Actual data	AR forecast	ANN forecast	Actual data	AR forecast	ANN forecast	Actual data	AR forecast	Actual data	AR forecast	Actual data	AR forecast
21 Aug	697,330	707,541	689,907	34,532	34,720	34,556	12,711	12,415	18,299	18,744	11,426	11,376
22 Aug	707,668	714,209	699,734	34,224	35,297	34,534	12,965	13,017	18,301	18,651	11,594	11,524
23 Aug	710,771	722,043	712,653	34,877	32,913	33,606	13,226	13,109	18,567	18,195	11,778	11,659
24 Aug	704,348	726,006	701,456	34,735	35,440	33,797	13,258	13,457	18,267	18,644	11,626	11,674
25 Aug	706,851	720,222	700,865	35,430	33,426	34,828	13,374	13,415	17,938	18,572	11,998	11,818
26 Aug	725,991	723,062	718,567	36,053	36,789	36,510	13,520	13,477	18,979	18,223	12,520	11,756
27 Aug	742,023	742,754	739,876	35,989	36,295	35,916	13,453	13,635	19,463	19,239	13,208	12,197
28 Aug	752,424	761,294	754,576	36,521	35,529	36,533	13,533	13,491	19,407	19,835	13,550	12,614
29 Aug	765,302	773,116	771,234	37,315	36,507	36,998	13,656	13,616	19,971	19,757	14,040	13,353
30 Aug	781,975	783,666	783,456	37,703	38,201	37,359	13,475	13,692	20,321	20,183	14,793	13,761
31 Aug	785,996	799,731	786,543	37,116	37,251	37,153	13,227	13,481	20,511	20,619	14,626	14,299

Table 10.6 Comparison in terms of RMSE between observed and forecast values for a period of 31 days, from 1 August to 31 August (*Source* Own)

Place	Forecast model	Root mean square error (RMSE)	CC	PP
India	AR	5.3321e+003	0.99	0.99
	ANN	4.8385e+003	0.99	0.99
Bengaluru	AR	1.0637e+003	0.87	0.62
	ANN	1.0345e+003	0.88	0.64
Chennai	AR	173.7424	0.98	0.95
Mumbai	AR	524.1394	0.88	0.76
Delhi	AR	420.9962	0.95	0.89

running mean and standard deviation in the same plot to see if they converge to a particular value and concluded that it is non-stationary. Hence, AR model with 10 days lag having strong correlation with the data is modelled, and the same model has been tested in the testing period or forecast period also. In order to understand the non-linear structure if it exists in the data, ANN model is constructed, and both the models are compared so as to identify the best model for further forecasting of the other cases such as death cases and recovered cases. Both models have outperformed in terms of the parameters used for measuring the same. Hence, any model can be used for the remaining cases. For future study of the same link between active cases, recovered cases and death cases has to be found, and if possible, then a combined ANN model with three outputs have to be developed which will forecast all three cases at a time. The future work will be concentrated on this combined study.

References

1. Ahmed, S. S. (2020). The coronavirus disease 2019 (COVID-19): A review. *JAMMR, 32*(4), 1–9. https://doi.org/10.9734/jammr/2020/v32i430393
2. Luo, Z., Ang, M. J. Y., Chan, S. Y., Yi, Z., Goh, Y. Y., Yan, S., et al. (2020). Combating the coronavirus pandemic: early detection, medical treatment, and a concerted effort by the global community. *Research* https://doi.org/10.34133/2020/6925296.
3. Poletto, C., Scarpino, S. V., & Volz, E. M. (2020). Applications of predictive modelling early in the COVID-19 epidemic. *Lancet Digital Health, 2*(10), Published Online August 10, https://doi.org/10.1016/S2589-7500(20)30196-5
4. Elsevier, Novel Coronavirus Information Center. (2020). https://www.elsevier.com/connect/coronavirus-information-center.
5. Shah, N. H., Suthar, A. H., & Jayswal, E. N. (2020). Control strategies to curtail transmission of COVID-19. *Hindawi International Journal of Mathematics and Mathematical Sciences*, Article ID 2649514. https://doi.org/10.1155/2020/2649514.
6. Arino, J., & Portet, S. (2020). A simple model for COVID-19. *Infectious Disease Modelling, 5*, 309–315. www.keaipublishing.com/idm, Production & Hosting by Elsevier B V on behalf of KeAi Communications Co. Ltd, an open access journal under CC-BY-NC-ND 4.0 International license.

7. Fong, S. J., Li, G., Dey, N., Crespo, R. G., & Herrera-Viedma, E. (2020). Finding an accurate early forecasting model from small dataset: a case of 2019-nCoV novel coronavirus outbreak. *Int Journal of Interactive Multimedia and Artificial Intelligence, 6*(1), 132–140.
8. Shah, N. H., Sheoran, N., Jayswal, E., Shukla D,, Shukla, N., Shukla, J. & Shah Y. (2020) Modelling COVID-19 transmission in the united states through interstate and foreign travels and evaluating impact of governmental public health interventions. medRxiv preprint the copyright holder for this preprint this version posted. https://doi.org/10.1101/2020.05.23.20110999
9. Wong, Z. S., Zhou, J., & Zhang, Q. (2019). Artificial intelligence for infectious disease big data analytics. *Infection, Disease & Health, 24*(1), 44–48.
10. Smarajit D. E. Y. (2020). Modeling Covid19 In India (Mar 3-May 7, 2020): How flat is flat, and other hard facts. medRxiv preprint, https://doi.org/10.1101/2020.05.11.20097865.
11. Hu, Z., Ge, Q., Jin, L., & Xiong, M. (2020). Artificial intelligence forecasting of covid-19 in China. http://arxiv.org/abs/2002.07112.
12. C, Z., Šegota, S. B., Anđelić, N., Lorencin, I., & Mrzljak, V. (2020). Modeling the spread of COVID-19 infection using a multilayer perceptron. *Computational and Mathematical Methods in Medicine*. https://doi.org/10.1155/2020/5714714.
13. https://www.accuweather.com/en/in/national/covid-19.
14. https://www.bing.com/covid/local/delhi_india?vert=graph.
15. Ixigo train application—Coronavirus Live Tracker.