PREDICTION OF DALY NEW COVID-19 CASES IN CHINAUSING ARTIFICIAL NEURAL NETWORKS

*Dr. Smartson. P. NYONI[1], Thabani NYONI[2], Tatenda. A. CHIHOHO[3]

[1]ZICHIRe Project, University of Zimbabwe, Harare, Zimbabwe
[2]Department of Economics, University of Zimbabwe, Harare, Zimbabwe
[3]Department of Economics, University of Zimbabwe, Harare, Zimbabwe
**Corresponding Author*

ABSTRACT

In this research article, the ANN approach was applied to analyze daily new COVID-19 cases in the Republic of China. The employed data covers the period 1 January 2020 to 31 October 2020 and the out-of-sample period ranges over the period November 2020 to 30April 2021. The residuals and forecast evaluation criteria (Error, MSE and MAE) of the applied model indicate that the model is stable in forecasting daily new COVID-19 infections in China. The results of the study indicate that daily COVID-19 cases in China; will continue to rise over the out-of-sample period. Despite all these measures the novel virus is still a public health challenge thus the Chinese government must continue to enforce WHO guidelines on prevention of COVID-19 in order to save precious lives.

Keywords: - *ANN, COVID-19, Forecasting*

INTRODUCTION

In December 2019, several pneumonia cases of unknown aetiology were discovered in Wuhan in the Hubei Province of China and it was later found by scientists that it was the novel coronavirus, COVID-19 (WHO, 2020; CDC, 2020). The virus was previously known to be found in animals but now it has slipped into humans causing a severe acute respiratory syndrome (Hu et al., 2020).The virus causes upper and lower respiratory infection with symptoms such as cough, fever, shortness of breath and occasionally gastro intestinal symptoms like diarrhea (Huang et al., 2020). The majority of COVID19 cases are mild cases. Patients with pre-existing chronic medical conditions tend to develop severe clinical disease. By 3 February 2020, 17238 cases of COVID-19 and 361 deaths had been reported (WHO, 2020). According to CDC the virus can be spread by air droplets from coughing and sneezing, close personal contact, touching contaminated surfaces or objects that contain viral particles and rarely from faecal contamination. The virus has an incubation period of about 14days during which it can spread to others (Guan et al., 2020; Wang et al, 2020; Cheng et al., 2020). The Chinese government responded aggressively to the COVID-19 outbreak on 23 January 2020. The whole cities were quarantined, the national holiday was extended, strict measures limiting travel and public gatherings were introduced, public spaces were closed and rigorous temperature monitoring was implemented nationwide (Yang et al., 2020; Li et al., 2020). Many studies have been done in order to understand the dynamics of COVID-19, of these, compartment models are on top in terms of being applied (Liu et al., 2020; Tuite & Fisman, 2020). In this study we applied the

Artificial neural network model in order to forecast daily new COVID-19 cases from November 1, 2020 to 30 April 2021. The most widely used ANN model is the multilayer Perceptron (MLP), with a single hidden layer feed forward network (Zhang, 1998, 2003, 2007; Jenkins, 1970). In this paper the ANN (12,12,1) model is used meaning that the model consists of 3 layers; input layer with 12 nodes, single hidden layer with 12 nodes and output layer with 1 node. The hyperbolic tangent function was chosen because it has a zero centered output and it does not have the vanishing gradient problem.

LITERATURE REVIEW

Table 1: Literature Review

Author(s)/year	Study Period	Method	Major Findings
Roosa et al (2020)	22June January 2020-9February 2020	Generalized logistic growth model, Richards growth model	The epidemic growth has slowed in recent days and the containment measures have successfully reduced transmission.
Koczkoda et al (2020).	30December to 09March 2020.	Exponential growth model (Heuristic)	Selfcare is a major factor to prevent the spread of COVID-19
Hu et al (2020).	11January 2020 to27 February 2020.	AI-based method	The model predicted that the epidemic of COVID-19 will be over by mid-April 2020.
Yang et al (2020)	10January 2020-3February 2020.	Simple linear regression model.	Findings suggest that the case fatality rate (CFR) of COVID-19 is lower than the previous corona virus epidemics caused by SARS COV and MERS COV.
Su et al (2020)	24January 2020-to 23 February 2020.	-SEIR model -Markov Chain Monte- Carlo	Public health interventions would reduce the risk of the spread of OVID-19 and that more rigorous control and prevention measures would effectively

			contain further spread.
Mohamed et al (2020).	21January 2020 to 18February 2020.	FPASSA-ANFIS	The model has a high ability to predict the number of confirmed cases within ten days.
Jia et al (2020).	12February 2020-11April 2020.	Uncertain SEIAR model.	The number of active cases on 15 April 2020 would be between 617-1200.
Yang et al (2020)	January 2020	-Modified SEIR model -AI methods	The modified SEIR model was effective in predicting COVID-19 Peaks and sizes
Ye &Yang (2020)	13February 2020 to 23March 2020	-ARIMA -Uncertain ARMA model	The uncertain time series analysis was more appropriate for predicting the cumulative number of confirmed COVID-19 cases in China.

METHOD

This paper applies the multi-layer perceptron neural network type of the ANN approach in order to predict daily new COVID-19 infections in China. The study specifically applies the ANN (12, 12, 1) model and chooses the more efficient hyperbolic tangent function as the activation function.

Data Issues

This study is based on daily new Covid-19 cases (referred to as Q series in this study) for all age groups in China. The data covers the period 1January 2020 to 31October 2020 while the out-of-sample forecast covers the period November 2020 to April 2021. All the data employed in this paper was gathered from the COVID-19 data repository prepared by the CSSE at JH University.

FINDINGS OF THE STUDY

DESCRIPTIVE STATISTICS

Table 2: Descriptive statistics

Mean	Median	Minimum	Maximum
299.41	30.000	0.00000	15136.
Std. Dev.	C.V.	Skewness	Ex. kurtosis
1140.6	3.8096	8.5455	96.547

5% Perc.	95% Perc.	IQ range	Missing obs.
0.00000	2054.7	79.500	0

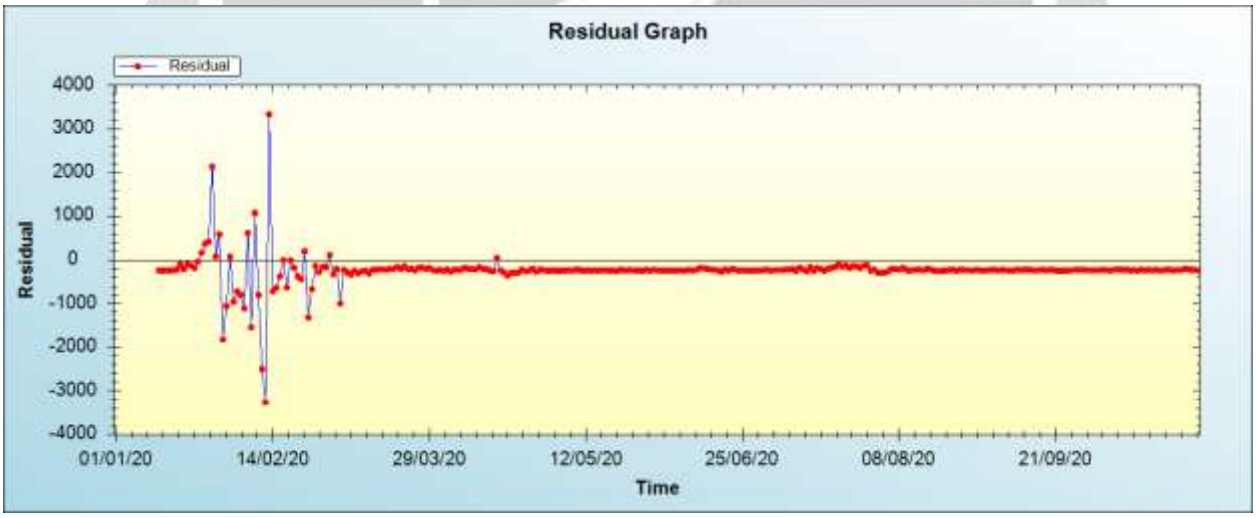
ANN MODEL SUMMARY FOR COVID-19 DAILY CASES IN CHINA

Table 3: ANN model summary

Variable	Q
Observations	293 (After Adjusting Endpoints)
Neural Network Architecture:	
Input Layer Neurons	12
Hidden Layer Neurons	12
Output Layer Neurons	1
Activation Function	Hyperbolic Tangent Function
Back Propagation Learning:	
Learning Rate	0.005
Momentum	0.05
Criteria:	
Error	0.055648
MSE	218962.163858
MAE	301.883043

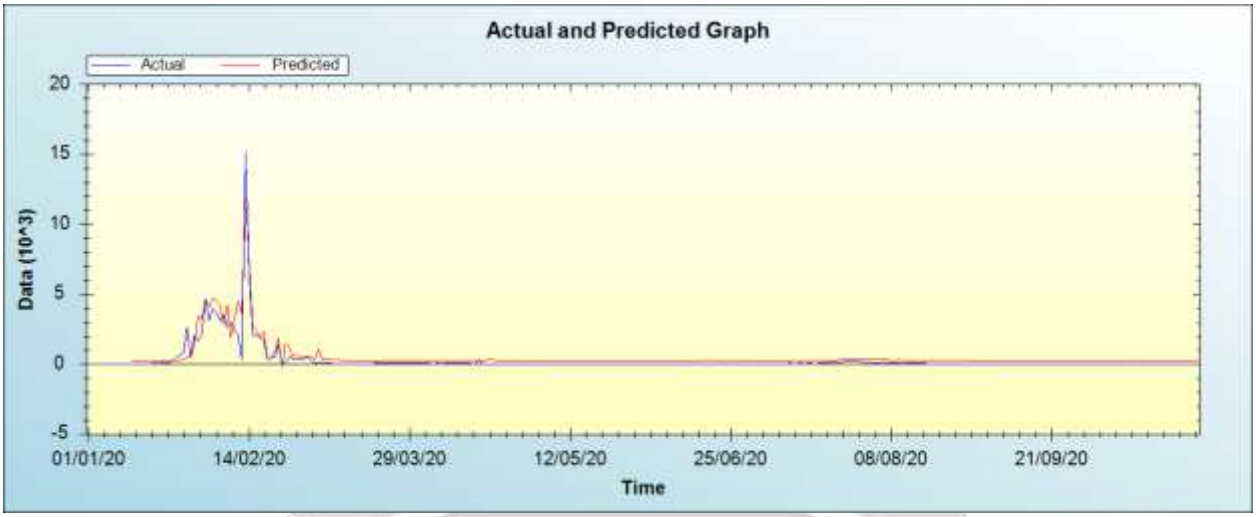
Residual Analysis for the ANN model

Figure 1: Residual analysis



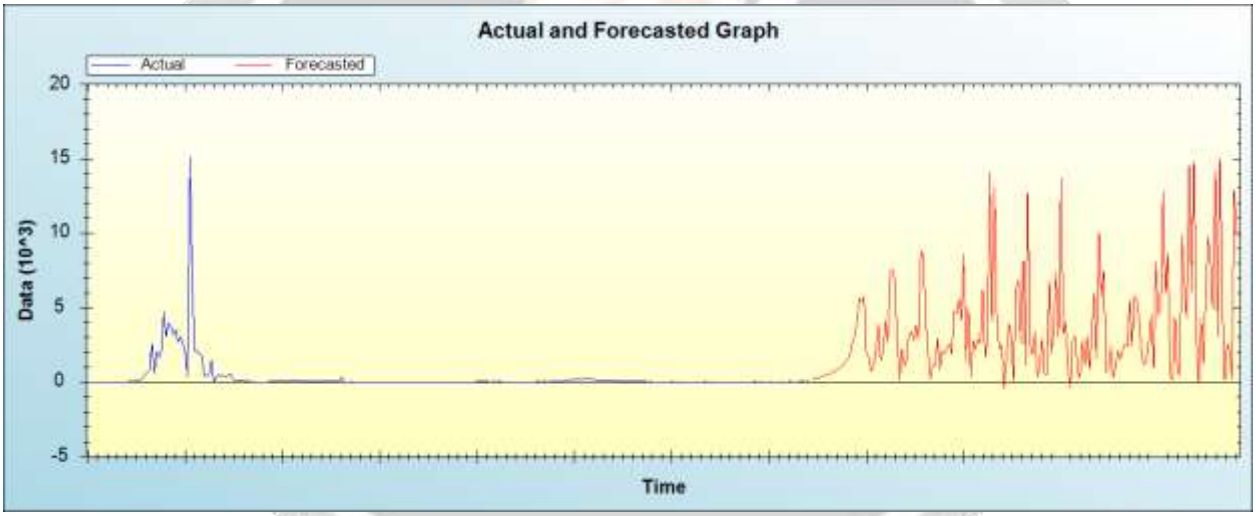
In-sample Forecast for Q

Figure 2: In-sample forecast for the Q series



Out-of-Sample Forecast for Q: Actual and Forecasted Graph

Figure 3: Out-of-sample forecast for Q: actual and forecasted graph



Out-of-Sample Forecast for Q: Forecasts only

Table 4: Tabulated out-of-sample forecasts

Day/Month/Year	Forecasted daily new COVID-19 cases
01/11/20	266.1495
02/11/20	247.8917
03/11/20	259.9013
04/11/20	323.5088
05/11/20	374.3798

06/11/20	421.9584
07/11/20	516.9904
08/11/20	535.9048
09/11/20	616.8100
10/11/20	735.2183
11/11/20	702.3572
12/11/20	925.1917
13/11/20	1000.8854
14/11/20	1068.9366
15/11/20	1367.9631
16/11/20	1584.4620
17/11/20	1772.1365
18/11/20	2593.7460
19/11/20	3003.2020
20/11/20	4003.6667
21/11/20	5632.8575
22/11/20	5372.7458
23/11/20	5735.1512
24/11/20	1988.6707
25/11/20	1653.2725
26/11/20	703.2296
27/11/20	1047.9968
28/11/20	2103.4695
29/11/20	3922.7244
30/11/20	1427.1929
01/12/20	1981.3921

02/12/20	4168.1815
03/12/20	2624.9471
04/12/20	7474.1606
05/12/20	7540.1746
06/12/20	6879.7956
07/12/20	2862.8577
08/12/20	102.3544
09/12/20	2224.2121
10/12/20	1043.7300
11/12/20	1574.9053
12/12/20	3235.4316
13/12/20	3398.8072
14/12/20	2736.0174
15/12/20	3874.1924
16/12/20	2862.7611
17/12/20	8860.1322
18/12/20	8537.7467
19/12/20	4104.4220
20/12/20	1903.9223
21/12/20	199.1962
22/12/20	1160.0380
23/12/20	1102.7506
24/12/20	2956.7701
25/12/20	820.3899
26/12/20	2117.7775
27/12/20	1872.1335

28/12/20	2288.8982
29/12/20	2608.0208
30/12/20	1903.9167
31/12/20	4804.9953
01/01/21	4620.7669
02/01/21	5642.6668
03/01/21	4152.9015
04/01/21	8582.5183
05/01/21	1365.9641
06/01/21	5007.3158
07/01/21	379.6056
08/01/21	2814.3374
09/01/21	2213.9624
10/01/21	2851.5239
11/01/21	2662.5598
12/01/21	6170.2550
13/01/21	1629.9606
14/01/21	2768.6596
15/01/21	14131.4795
16/01/21	4063.6117
17/01/21	13062.7191
18/01/21	3951.2008
19/01/21	2246.3473
20/01/21	2463.0996
21/01/21	-479.1250
22/01/21	1523.3080

23/01/21	3916.4094
24/01/21	3106.0314
25/01/21	43.1904
26/01/21	6333.1199
27/01/21	6869.3301
28/01/21	2628.9428
29/01/21	8162.0820
30/01/21	1158.1454
31/01/21	12655.1944
01/02/21	2544.4360
02/02/21	1853.3925
03/02/21	3425.7627
04/02/21	379.5091
05/02/21	872.9299
06/02/21	2920.3171
07/02/21	649.4860
08/02/21	467.9781
09/02/21	6835.4741
10/02/21	1970.5782
11/02/21	3537.4557
12/02/21	7377.5467
13/02/21	3083.8032
14/02/21	13741.1619
15/02/21	3312.8851
16/02/21	4010.0259
17/02/21	908.7873

18/02/21	-364.7261
19/02/21	2922.8428
20/02/21	3112.4900
21/02/21	628.0202
22/02/21	283.0686
23/02/21	2765.8528
24/02/21	1112.8564
25/02/21	3074.8852
26/02/21	844.9301
27/02/21	3638.9220
28/02/21	5979.6080
01/03/21	1554.9610
02/03/21	10030.8976
03/03/21	5699.6648
04/03/21	7483.3094
05/03/21	675.5674
06/03/21	785.4439
07/03/21	2431.2687
08/03/21	295.0798
09/03/21	1012.2456
10/03/21	2100.0466
11/03/21	1502.5618
12/03/21	2132.4206
13/03/21	2601.5125
14/03/21	2380.1303
15/03/21	5565.1743

16/03/21	2466.1246
17/03/21	5752.8570
18/03/21	5491.2197
19/03/21	4051.0340
20/03/21	1753.5256
21/03/21	1145.8280
22/03/21	1557.6095
23/03/21	2749.3816
24/03/21	4515.8738
25/03/21	916.4149
26/03/21	8076.3606
27/03/21	4455.1048
28/03/21	4894.7697
29/03/21	12773.5732
30/03/21	5978.4487
31/03/21	8656.9678
01/04/21	431.4899
02/04/21	192.7163
03/04/21	4386.1432
04/04/21	773.5756
05/04/21	386.0026
06/04/21	9892.5948
07/04/21	6502.5533
08/04/21	4250.9072
09/04/21	14491.3120
10/04/21	6131.3689

11/04/21	14796.2874
12/04/21	3024.5154
13/04/21	-76.9361
14/04/21	4310.6053
15/04/21	1208.5059
16/04/21	6088.2608
17/04/21	9760.2023
18/04/21	8130.9650
19/04/21	4891.7146
20/04/21	14179.7344
21/04/21	3098.2235
22/04/21	15067.0266
23/04/21	5000.8529
24/04/21	141.7559
25/04/21	2597.7426
26/04/21	2085.2300
27/04/21	235.7622
28/04/21	12887.0252
29/04/21	10060.7156
30/04/21	9818.0944

Figure 4: Graphical presentation of out-of-sample forecasts

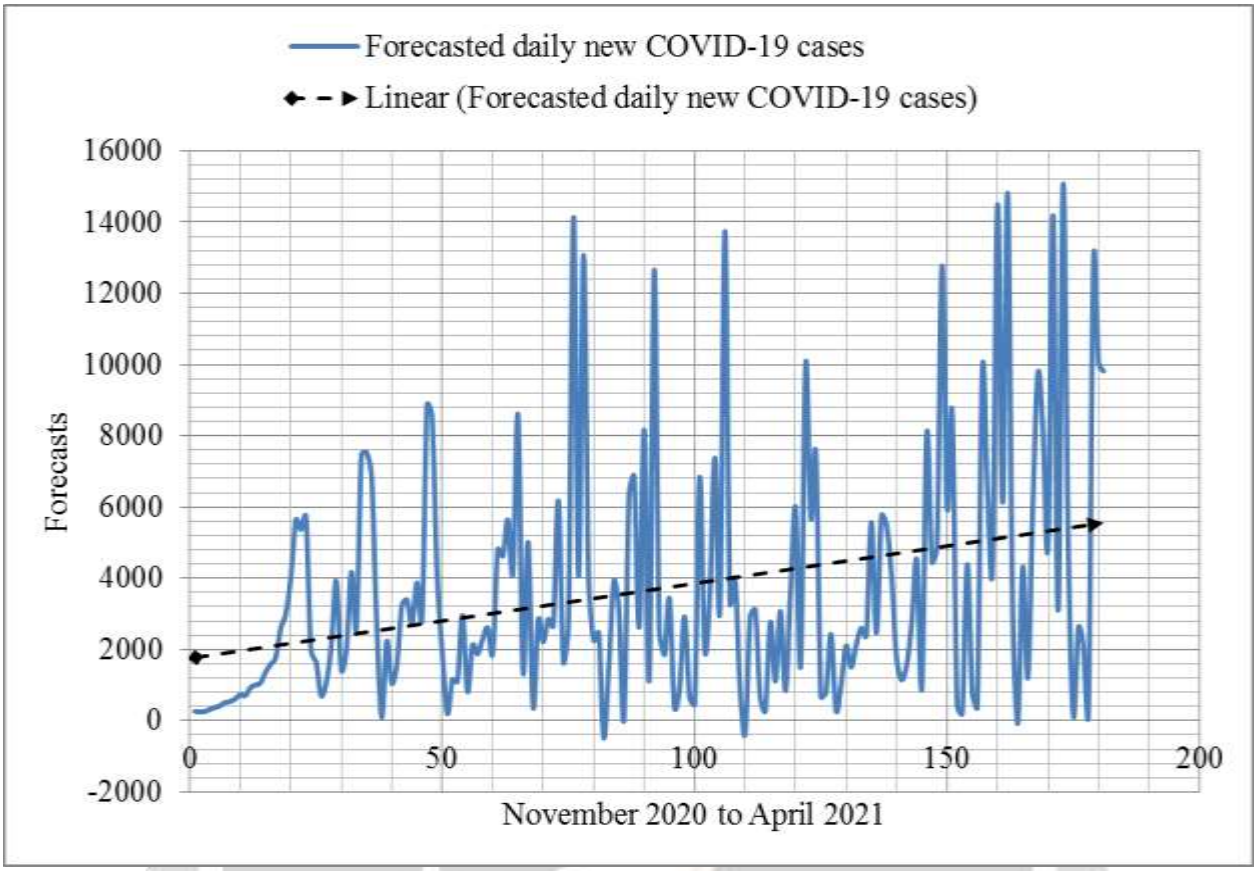


Figure 1 shows that over the study period, the minimum and maximum daily new COVID-19 cases are 0 and 15136 respectively. The average daily new cases are 299 per day over the study period. The ANN (12,12,1) model simulates the observed data very well as shown in Figure 3 and table 4. The residual graph and model evaluation criteria indicate that the applied mode is stable and suitable for forecasting daily new coronavirus cases in China. The study establishes that daily COVID-19 cases in China, will continue to rise over the out-of-sample period.

CONCLUSION & RECOMMENDATIONS

The COVID-19 virus was first reported in Wuhan China and it has rapidly spread across the whole world with varying extent of the impact on the economies of different countries. The western European, India, Brazil, Mexico and the US are among the hardest hit nations. The Chinese government responded aggressively to the out break by imposing travel restrictions, closing schools and Universities, temporary ban on public gatherings and other related activities. Despite all these measures the novel virus is still a public health challenge thus the Chinese government must continue to enforce WHO guidelines on prevention of COVID-19 in order to save precious lives. If an effective coronavirus vaccine is availed to many people in the out-of sample period daily new cases are expected to drop sharply.

REFERENCES

[1] Cheng &Shan (2020) .2019 Novel Coronavirus: Where We are and What We Know. Infection 2020.

[2] COVID-19 Repository By the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University.

[3] Dong, E., *et al*. (2020). An Interactive Web-based Dashboard to Track COVID-19 in Real Time, *Lancet Infectious Diseases*, 20 (5): 533 – 534.

[4] G.E.P. Box, G. Jenkins, "Time Series Analysis, Forecasting and Control", Holden-Day, San Francisco, CA, 1970.

[5] G.P. Zhang, "A neural network ensemble method with jittered training data for time series forecasting", Information Sciences 177 (2007), pages: 5329–5346.

[6] Guan et al (2020). Clinical characteristics of 2019 novel coronavirus infection in China.

[7] Hu et al (2020). Artificial intelligence Forecasting of COVID-19 in China. pp 1-20.

[8] Huang C et al (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395:497-506.

[9] Li Q et al (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. N Engl J Med. doi: 10.1056/NEJMoa2001316.

[10] Liu, Y., & Liu, B. (2020). Estimating unknown parameters in uncertain differential equation by maximum likelihood estimation, Technical Report.

[11] Tuite & Fisman (2020). Reporting, epidemic growth, and reproduction numbers for the 2019 novel coronavirus (2019-nCoV) epidemic. Ann Intern Med. doi: 10.7326/M20

[12] World Health Organization (2020). Novel Coronavirus (2019nCoV) Situation Report – 14: https:// www.who.int/docs/default-source/coronaviruse/situationreports/20200203-sitrep-14-ncov.

[13] Wu et al (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet .6736,20,30260-9.

[14] Yang et al (2020). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions 1-18.

[15] Zhang et al (2020). "Forecasting with artificial neural networks: The state of the art", International Journal of Forecasting 14 (1998), pages: 35-62.

[16] Zhang, "Time series forecasting using a hybrid ARIMA and neural network model", Neurocomputing 50 (2003), pages: 159–175.