PREDICTION OF DAILY NEW CASES OF COVID-19 IN BRAZIL USING 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

The first case of COVID-19 in Brazil was first reported on 26February 2020. Understanding the dynamics of COVID-19 epidemic in Brazil is very critical in order to plan and allocate adequate resources for health. In this research article, the ANN approach was applied to analyze daily new COVID-19 cases. The employed data covers the period 26 February 2020 to 31October 2020 and the out-of-sample period ranges over the period November 2020 to April 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 cases in Brazil. The results of the study indicate that the pandemic may not go away anytime soon in Brazil: daily cases will remain very high over the out-of-sample period. Relevant authorities in the country ought to continue implementing control and preventive measures as outlined by WHO.

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

INTRODUCTION

Machine learning is a field of study which is concerned with construction of computer programs that automatically improve with experience (Alimadadi et al, 2020). Machine learning algorithms use computational methods to learn information directly from data without relying on a predetermined equation to model. It can be broadly divided into Supervised, Unsupervised and Semi-supervised learning. Supervised learning trains a model on known inputs and output data to predict future outputs. Unsupervised learning finds hidden patterns or intrinsic structures in the input data. Semi-supervised learning uses a mixture of both techniques (Weng et al, 2017). Machine learning algorithms are widely used in public health and are meant to solve 3 basic groups of problems which are classification, regression and density estimation. In this section we will briefly discuss some of the algorithms found in each group. Classification algorithms are meant to solve the classification problem and examples of these are logistic regression, K-nearest Neighbors, Support vector machine, neural networks, Naïve Bayes, discriminant analysis, decision trees and ensembles. Logistic regression fits a model that can predict the probability of a binary response belonging to one class or the other (Souza etal, 2020; Weng etal ,2017; Geron etal,2017). The support vector machine (SVM) classifies data by finding a linear decision boundary (hyperplane) that separates all data points of one class from another (Duda etal, 2000). The points on the wrong side of the hyperplane is penalized using a loss function (use of slack variables). It uses a kernel transformation to transform non-linearly separable data from the input

space, X, to a higher dimensional space H. (Bishop.C,2006, Weng etal,2017). Decision trees predict responses to data by following the decisions in the tree from the root to a leaf node. The number of branches and values of the weights are determined in the training process. The k-Nearest Neighbors algorithm is based on the concept of neighborhood in which the neighbors are similar (Norvig.S. R,2013). Categorizes objects based on the classes of their nearest neighbors. Distance metrics are used (Weng S. et al, 2017). Examples of regression algorithms include linear regression, nonlinear regression, support vector regression and Regression trees. In forecasting problems in public health artificial neural networks and the support vector machine algorithms are widely used. Support vector regression (SVR), DLSM-SVM and LSM-SVM are the commonly applied SVM algorithms. In this paper we applied the artificial neural network, the Multilayer perceptron with a single hidden layer. The model mimics the function of the human brain (Mitchelle, 1997; Norvig S. R,2013; Witten H. L, 2011). The results of this study will help the Brazilian government in policy and decision making and in the national response to the epidemic.

LITERATURE REVIEW

Table 1: Selected Papers for Review

Author(s)/year	Study period	Method	Major findings
Ribeiro etal (2020)	5March 2020-19 April 2020	ARIMA, Random forest, RIDGE, SVR, ensemble, CUBIST	All the models generate accurate forecasting and can be applied in forecasting and monitoring the growth of COVID-19 cases.
Crokidakis. N (2020)	5March 2020 -26April 2020	SIQR model	The implementation of social distancing policies changed the initial exponential growth to a sub-exponential one. the model predicts mitigation effect (flattening of the curve).
Silva R.G etal (2020).	26February 2020-28April 2020	Bayesian regression neural network, cubist, K-nearest neighbors, Random forest, SVR, VMD	The evaluated models were very efficient in forecasting cumulative COVID-19 cases up to 6days ahead.

Sousa. G.B etal (2020)	26February 2020-30March 2020	SIR model	There is rapid COVID-19 spread and high mortality. The epidemic peak is likely to be between late April and early May 2020.
Martinez E.Z etal (2020).	26February 2020-25April 2020	Holt's model	The model can be an adequate short-term forecasting method if their assumptions are adequately verified.
Crokidakis N (2020).	26February 2020-25March 2020.	SIQR model	The basic reproduction number is 5,25 and the epidemic doubling time is 2.72 days.
Bastos &Daniel (2020).	25Febrauary 2020-30March 2020	Modified SIR models.	Short-term and long-term forecasts indicate that social distancing measures are able to flatten the pattern of COVID-19.
Morato .M etal (2020).	25February 2020-16June 2020	SIRD model	Social isolation policy could save 100,000 lives.
Canabarro. A etal (2020).	25February 2020-30March 2020	Modified SIRD model	There is a high demand for ICU units. early isolation and quarantine may reduce this demand
Souza F etal (2020).	26February 2020-30May 2020	Machine learning models.	Machine learning algorithms can help in the prediction of prognosis.

METHOD

This paper applies the multi-layer perceptron neural network type of the ANN approach in order to predict daily new COVID-19 cases in Brazil. This study particularly 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 Covid19cases (referred to as Z series in this study) for all age groups in Brazil. The data covers the period 26 February 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
22231.	20647.	0.00000	69074.
Std. Dev.	C.V.	Skewness	Ex. kurtosis
18060.	0.81238	0.40660	-0.86854
5% Perc.	95% Perc.	IQ range	Missing obs.
1.0000	52761.	31844.	0

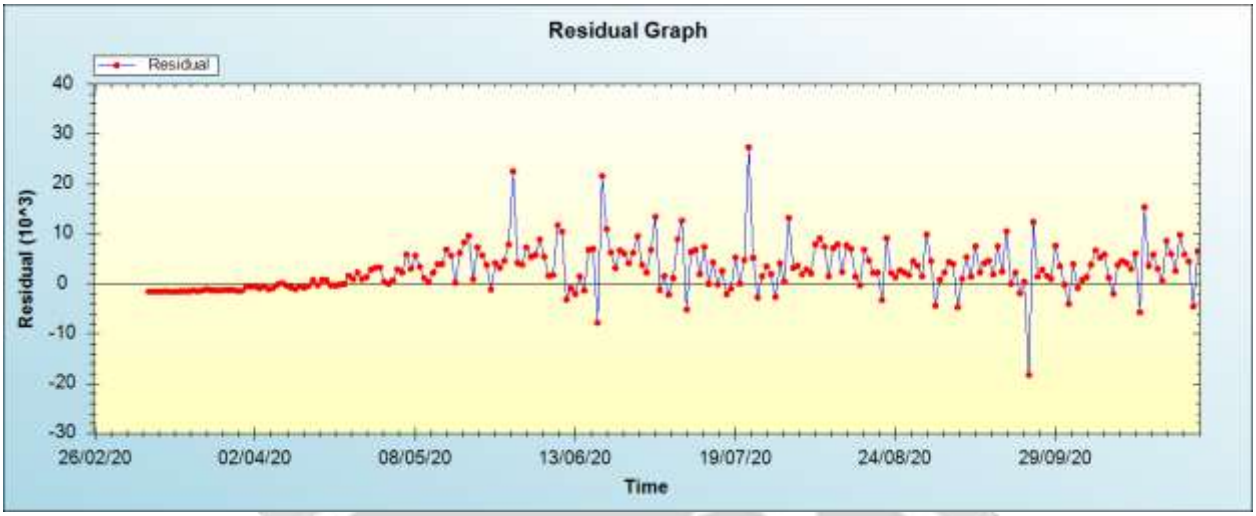
ANN MODEL SUMMARY FOR COVID-19 DAILY CASES IN BRAZIL

Table 3: ANN model summary

Variable	Z
Observations	237 (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	0.1443O5
MSE	30665185.216637
MAE	3926.638355

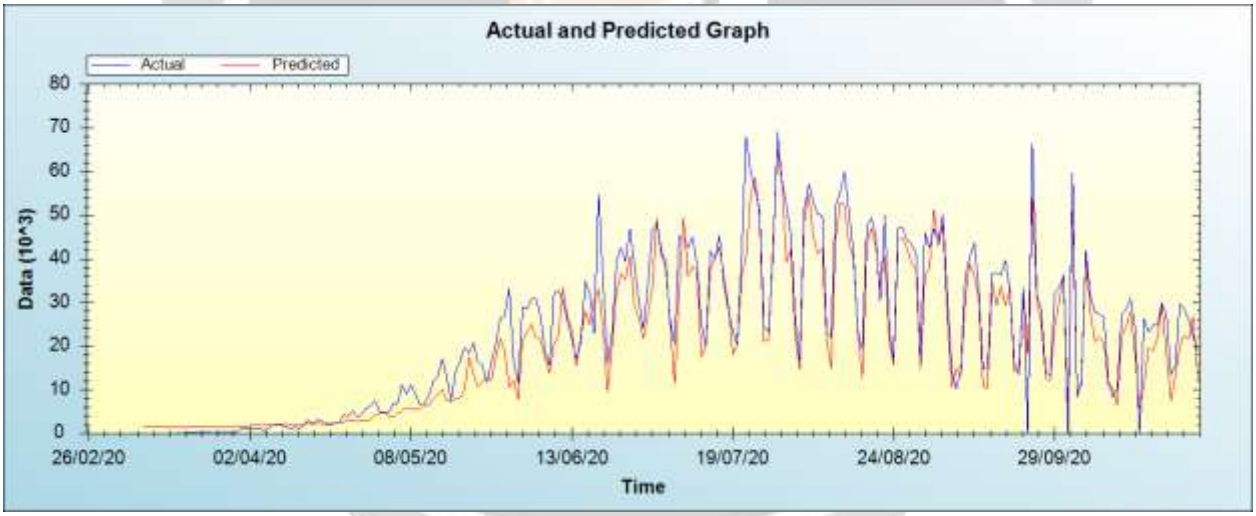
Residual Analysis for the ANN model

Figure 1: Residual analysis



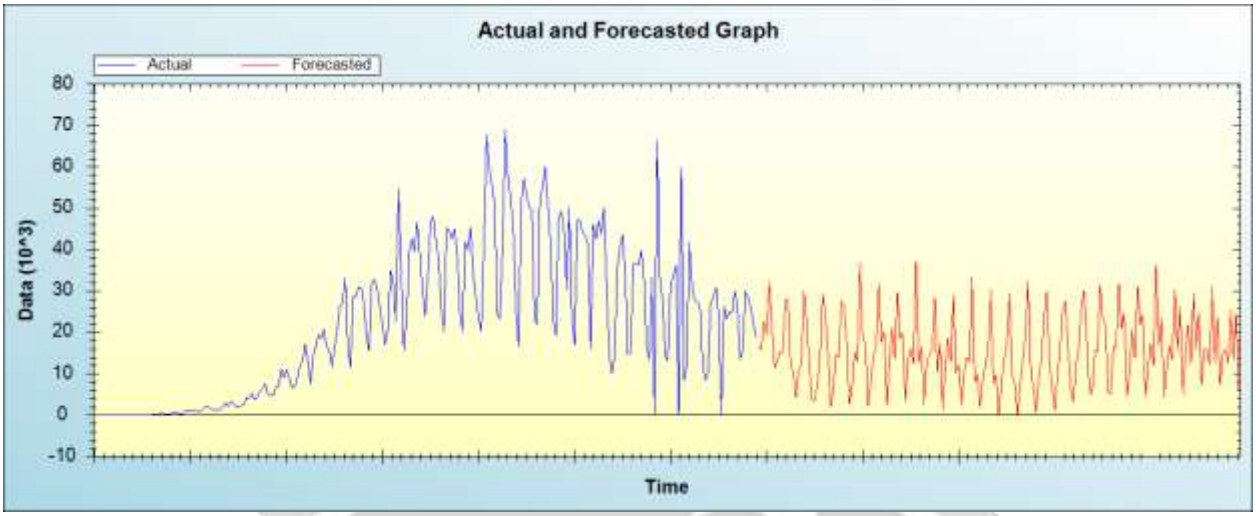
In-sample Forecast for Z

Figure 2: In-sample forecast for the Z series



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

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



Out-of-Sample Forecast for Z: Forecasts only

Table 4: Tabulated out-of-sample forecasts

Day/Month/Year	Forecasted daily new COVID-19 cases
01/11/20	16669.1638
02/11/20	15814.6019
03/11/20	22661.5710
04/11/20	19166.2829
05/11/20	32315.6341
06/11/20	20977.8281
07/11/20	11341.6246
08/11/20	13007.6298
09/11/20	15456.5635
10/11/20	15065.4746
11/11/20	27968.5308
12/11/20	27651.1555
13/11/20	12155.0333
14/11/20	10428.7939

15/11/20	4206.5751
16/11/20	10894.9603
17/11/20	12457.3403
18/11/20	30001.5344
19/11/20	22590.3216
20/11/20	11749.6894
21/11/20	3342.0768
22/11/20	3485.0386
23/11/20	7493.1095
24/11/20	14373.3289
25/11/20	29201.9268
26/11/20	24415.6568
27/11/20	14836.0586
28/11/20	1972.1370
29/11/20	6502.5093
30/11/20	14647.2595
01/12/20	13956.5206
02/12/20	27738.7877
03/12/20	26504.7240
04/12/20	17273.4754
05/12/20	2711.5341
06/12/20	6826.0178
07/12/20	15222.6071
08/12/20	13006.7143
09/12/20	36724.5914
10/12/20	18988.0976

11/12/20	17361.8098
12/12/20	2274.1419
13/12/20	7755.0078
14/12/20	14410.8737
15/12/20	16226.6177
16/12/20	31740.8023
17/12/20	17687.7568
18/12/20	20459.3003
19/12/20	2177.7536
20/12/20	13370.4519
21/12/20	21182.6456
22/12/20	13551.6429
23/12/20	29923.5208
24/12/20	18525.6750
25/12/20	20004.7949
26/12/20	3124.0916
27/12/20	13899.8448
28/12/20	16132.7514
29/12/20	12491.6564
30/12/20	37104.6842
31/12/20	12847.7929
01/01/21	16669.9148
02/01/21	2812.6332
03/01/21	12643.8545
04/01/21	14616.9322
05/01/21	15604.7521

06/01/21	28483.3272
07/01/21	10522.3930
08/01/21	17285.1029
09/01/21	1451.0612
10/01/21	11821.0271
11/01/21	18679.3060
12/01/21	12840.8060
13/01/21	28934.9088
14/01/21	10262.9957
15/01/21	12371.3020
16/01/21	2364.2307
17/01/21	10272.6584
18/01/21	13916.2893
19/01/21	12618.4778
20/01/21	33347.6133
21/01/21	8003.5395
22/01/21	9935.0891
23/01/21	1791.1109
24/01/21	8465.7060
25/01/21	12105.4955
26/01/21	16543.4652
27/01/21	30409.2083
28/01/21	7867.5155
29/01/21	9727.8468
30/01/21	-9.4987
31/01/21	7494.2997

01/02/21	13558.1594
02/02/21	17428.6007
03/02/21	29608.4109
04/02/21	9608.6637
05/02/21	7508.9574
06/02/21	89.5743
07/02/21	7756.6330
08/02/21	13186.6874
09/02/21	19115.2095
10/02/21	32461.1828
11/02/21	10762.4549
12/02/21	6423.7976
13/02/21	733.9271
14/02/21	8988.0826
15/02/21	13600.2735
16/02/21	24141.5627
17/02/21	30106.1138
18/02/21	12210.2911
19/02/21	7028.9382
20/02/21	1168.0533
21/02/21	9805.3776
22/02/21	16339.8105
23/02/21	25800.1350
24/02/21	27513.9472
25/02/21	15703.5676
26/02/21	6237.3021

27/02/21	2980.7405
28/02/21	12493.4900
01/03/21	16079.8766
02/03/21	26849.0142
03/03/21	30114.5317
04/03/21	18275.4794
05/03/21	5775.4228
06/03/21	5069.9332
07/03/21	15724.3802
08/03/21	15402.1837
09/03/21	31653.1819
10/03/21	23456.8165
11/03/21	21777.0873
12/03/21	5535.7807
13/03/21	5089.9054
14/03/21	16608.9101
15/03/21	16418.9832
16/03/21	31576.2050
17/03/21	21004.9286
18/03/21	24742.9951
19/03/21	4609.4070
20/03/21	10257.1429
21/03/21	20718.9474
22/03/21	13733.0863
23/03/21	31156.4495
24/03/21	22001.5943

25/03/21	24454.1606
26/03/21	4645.2662
27/03/21	12006.1914
28/03/21	17457.3520
29/03/21	11999.5617
30/03/21	36475.7262
31/03/21	16711.0569
01/04/21	23461.2501
02/04/21	4410.0784
03/04/21	11637.5859
04/04/21	16695.2691
05/04/21	13359.9945
06/04/21	30403.7081
07/04/21	15754.5186
08/04/21	26097.0317
09/04/21	5047.2655
10/04/21	15218.6583
11/04/21	22016.3911
12/04/21	12509.4200
13/04/21	29499.4018
14/04/21	15906.8513
15/04/21	24916.4861
16/04/21	7195.9320
17/04/21	16125.0606
18/04/21	15838.3666
19/04/21	12133.9272

20/04/21	31085.8442
21/04/21	12779.8316
22/04/21	23739.5523
23/04/21	7249.4059
24/04/21	14771.2550
25/04/21	16001.1179
26/04/21	12489.0352
27/04/21	25417.7421
28/04/21	13486.4924
29/04/21	24518.1631
30/04/21	5863.1997

Figure 1 shows that over the study period, the minimum and Maximum daily new cases are 0 and 69074 cases. The average daily new cases are 22231 cases. The residual graph and model evaluation criteria indicate that the applied model is stable and suitable for forecasting daily new COVID-19 cases in Brazil. Figure 3 shows the in-sample forecasts which indicate that the applied model simulates the observed data very well. The out-of sample forecasts indicate that COVID-19 daily cases volume will remain very high in the country, between approximately 500 cases per day and about 35000 cases per day.

CONCLUSION & RECOMMENDATIONS

The COVID-19 virus is rapidly spreading and associated with high mortality in Brazil. The health delivery system is overwhelmed with managing both outpatients and admitted patients in the various health care facilities. The increased demand for ICU units can be significantly reduced by early isolation and quarantining of COVID-19 contacts. The government must carry out massive mobile COVID-19 tests to detect infected cases and treat them early. Furthermore, the Authorities should continue enforcing WHO guidelines on prevention and control of the Novel Corona virus in order to reduce morbidity and mortality related to the deadly COVID-19 virus.

REFERENCES

[1] Alimadadi etal (2020) "Artificial intelligence and machine learning to fight covid-19," Physiological Genomics, vol. 52, 03 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] Duda etal (2000), P. E. Hart, and D. G. Stork, Pattern Classification (2nd Edition), 2nd ed. WileyInterscience

[5] Friedman etal (1997). "Bayesian network classifiers," Machine learning, vol. 29, no. 2-3, pp. 131–163.

[6] G´eron A (2017). Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. Sebastopol, CA: O'Reilly Media.

[7] Mitchell T.M (1997), Machine Learning. McGraw-Hill Education.

[8] P. Bruce and A. Bruce, Practical Statistics for Data Scientists. Sebastopol, CA: O'Reilly

[9] S. R. Peter Norvig (2013), Intelig^encia Artificial, 3rd ed. Elsevier Editora, iSBN 8535237011, 9788535237016.

[10] Souza etal (2020). Predicting the distance outcome in COVID-19 positive patients through Machine learning: a retrospective cohort study with Brazilian data. medRxiv pp1-20.