PREDICTION OF DAILY NEW COVID-19 CASES IN THE UNITED STATES OF AMERICA 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

Time series forecasting models are broadly classified into 2 groups; Linear and nonlinear models. Box-Jenkins (ARIMA) models are linear models which are simple, easy to understand and popular in forecasting problems in various fields such as Medicine, Engineering, Agriculture and Economics. Examples of stochastic nonlinear models are the non-linear moving average model (NMA), ARCH, GARCH and EGARCH models. ARCH and GARCH models are popularly known for modeling volatility in financial time series. Machine learning methods have also been applied by many researchers and they have been found to be good in modeling complex data especially in Public health. Some of the methods include Artificial neural networks (ANN), Support Vector Machine (SVM), Ensembles, Decision trees and Graphical Models (Bayesian Networks). In ANN model, appropriate selection of the number of inputs, hidden, output nodes and activation function are critical. In this study the ANN (12,12,1) model with the hyperbolic tangent function as the ideal activation function is applied, the findings of the study show that the daily new corona virus infections will be ranging from 75,000-80 000 cases from November 1,2020 to April 30 ,2021. However, daily new cases may go above 80,000 because of a negative shock from events associated with US elections. A significant drop in daily new cases is expected if an effective vaccine is availed early to the American people. The study recommends that the US government should tighten lockdown restrictions, enforce wearing of face masks, regular hand washing with soap or use of hand sanitizers, physical distancing, massive mobile testing, contact tracing, early treatment of cases and continuous education among communities.

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

INTRODUCTION

The United States of America (USA) has been hard hit by the COVID-19 pandemic and is currently facing its worst economic meltdown and Health crisis. The first case of the novel corona virus was reported on 20th of January 2020 in Snohomish county, Washington DC (Holshue et al, 2020; Moghadas et al, 2020; Harcourt et al, 2020). COVID-19 is a communicable disease caused by the new virus SARS-CoV-2 which was first reported in Wuhan in the Hubei Province of China at the end of December 2019 (Yang et al, 2020). According to the World

meter, on the 20th of April 2020,USA had reported 792,759 confirmed cases of COVID-19 (Worldometer, 2020). As of 24 April 2020 the SARS-CoV-2 had resulted in over 830,000 confirmed infections (Miller et al, 2020). The United States surpassed the 2million corona virus cases on the 15th 0f June 2020 in less than 5months after the first case was identified (Pham, 2020). According to statistics from Johns Hopkins University 115 people had succumbed to the disease by 15th June 2020. Several studies have been done to try and model the incidence of Covid -19 in China however the trends of daily incidence and deaths in USA are still poorly understood (Yuan et al, 2020). COVID-19 prediction models can be classified into 1) mathematical models which simulate disease within a population 2) statistical models that fit a function to the observed data and extrapolate forward into the future (Gregory et al, 2020). The widely used mathematical models are the compartmental models. These divide the population into compartments that denote disease status and supply a set of differential equations that define the flow of the population between compartments. The compartments are named as follows; the SIR, susceptible, infectious, and Recovered and SEIR; susceptible, exposed, infectious and Recovered. However, in this study the MLP is applied. ANN has been applied in many public health researches and the forecast results have been observed to be very accurate and reliable (Mollalo et al, 2020). After applying this machine learning algorithm, the study results will help the US government in understanding the trends of covid -19 in the out of sample period. This will guide planning, policy making and the execution of appropriate scientific and evidenced based National response to the epidemic.

LITERATURE REVIEW

Table 1: Literature Review

Author(s)/year	Study period	Method	Major findings
Velasquez &Lara (2020)	21January 2020 to 12 April 2020.	Gaussian Process Regression model	The peak number of deaths in USA will be around June 14,2020 with a peak number of 132,074 deaths and infected individuals of about 1,157,796 and the number of deaths at the end of the epidemic of about 132,800
Ahmar & Boj (2020)	15Feb 2020 to 2July 2020.	Sutte ARIMA	COVID-19 surging in the US
Pham (2020)	29Feb 2020 to 13 June 2020.	Novel generalized Mathematical Model	The model predicted that 128,500-140,000 people in US will have died of COVID-19 by July 4, 2020.Between

			137,900-154,000 people will have died of covid-19 by 31July ,2020 and 148,500-169,700 will have died by end of August 2020.
Seyed etal (2020).		Compartmental based mathematical model; Transition model under different scenarios	Self-isolation within 24hrs substantially reduces the peak number of ICU beds by 75% required for covid -19 infected patients. Covid-19 outbreak will most likely overwhelm current hospital capacity and that expanding the number of ICU beds is an urgent issue.
Yuan et al (2020)	1March 2020 to 7April 2020	General linear model and Pearson correlation test	There were 636,282 new cases and 28325 deaths from COVID-19 in USA from 1March 2020 to 15April 2020 with crude Mortality of 4,45%. Search terms related to COVID-19 are highly correlated to the daily covid-19 new cases and deaths in USA.
Mollalo et al (2020)	22Jan 2020 to 25April 2020.	ANN(MLP) ANN-with 1hiddenlayer, ANN - with 2 hidden layers Logistic regression, Geospatial analysis with Global Moran's I	Spatial analysis showed that COVID-19 incidence rate is clustered in the continental US. The ANN with 1hidden layer performed better than other models with an

			accuracy of 65%.
Naid et al (2020)	20January 2020 to 22 April 2020	A machine learning algorithm was developed.	By 22 April 2020the US may have had 1,5 -2,029times the number of reported infections.
Gregory et al (2020)	20Jan 2020 to 30 April 2020.	Bayesian nonlinear model and Random Forest Algorithm were fused.	The model predicts COVID-19cases and deaths with an absolute scaled error of 0,40 for cases and 0,32 for deaths throughout the 2week period.

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 the United States of America. This piece of work 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 A series in this study) for all age groups the United States of America (USA). The data covers the period 24 January 2020 to 31 October 2020 while the out-of-sample forecast covers the period November 2020 to April 2021. All the data employed in this research paper was gathered Johns Hopkins University (USA).

FINDINGS OF THE STUDY

DESCRIPTIVE STATISTICS

Table 2: Descriptive statistics

Mean	Median	Minimum	Maximum
32360.	31362.	0.00000	99321.
Std. Dev.	C.V.	Skewness	Ex. Kurtosis
22524.	0.69603	0.20034	-0.62634
5% Perc.	95% Perc.	IQ range	Missing obs.
0.00000	69154.	28826.	0

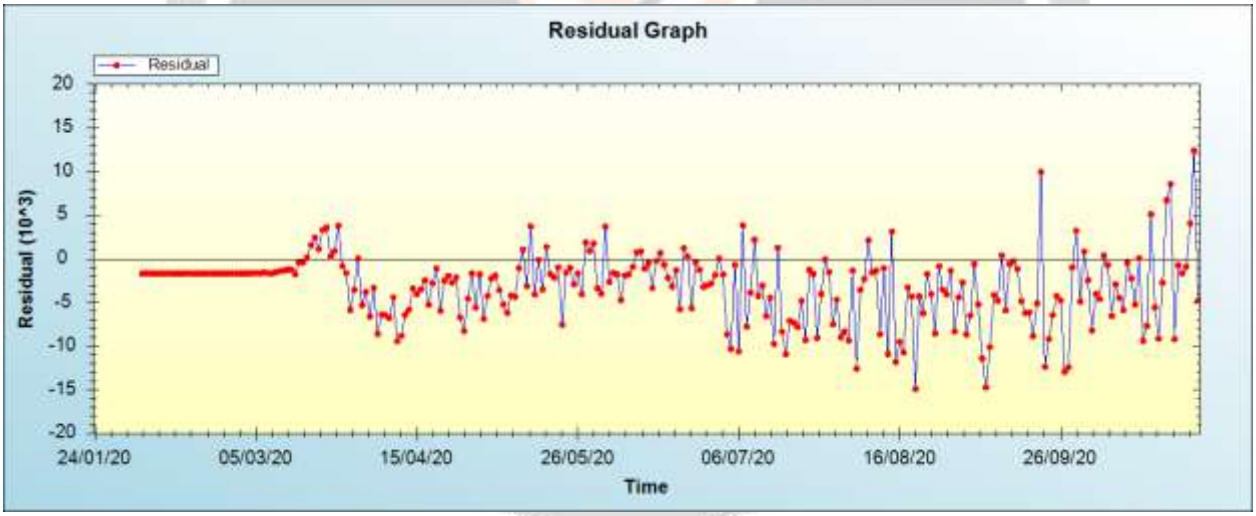
ANN MODEL SUMMARY FOR COVID-19 DAILY CASES IN USA

Table 3: ANN model summary

Variable	A
Observations	270 (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	0.091846
MSE	25683682.776166
MAE	3930.682535

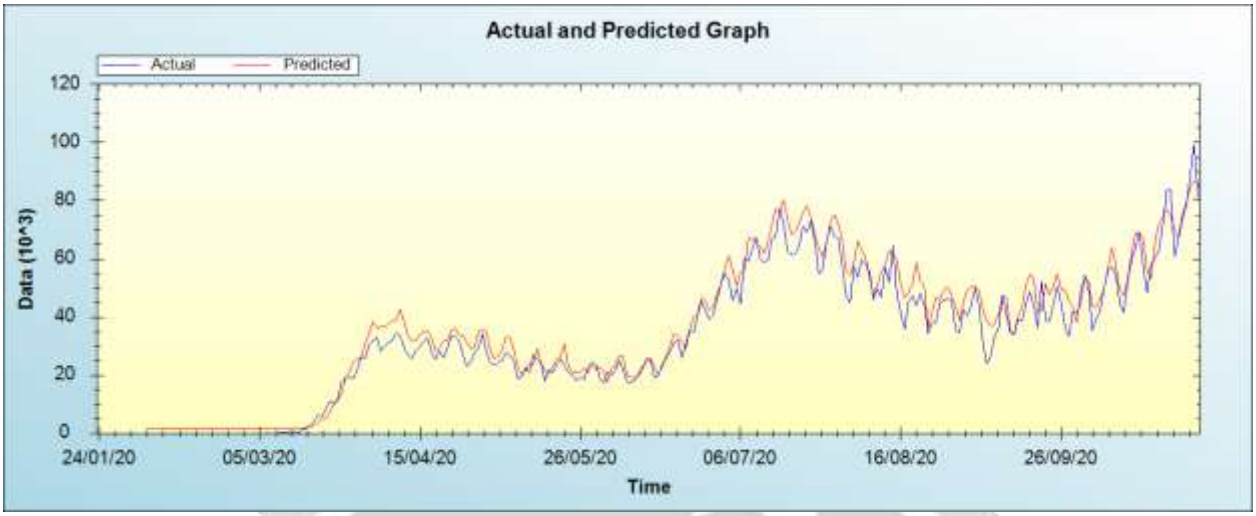
Residual Analysis for the ANN model

Figure 1: Residual analysis



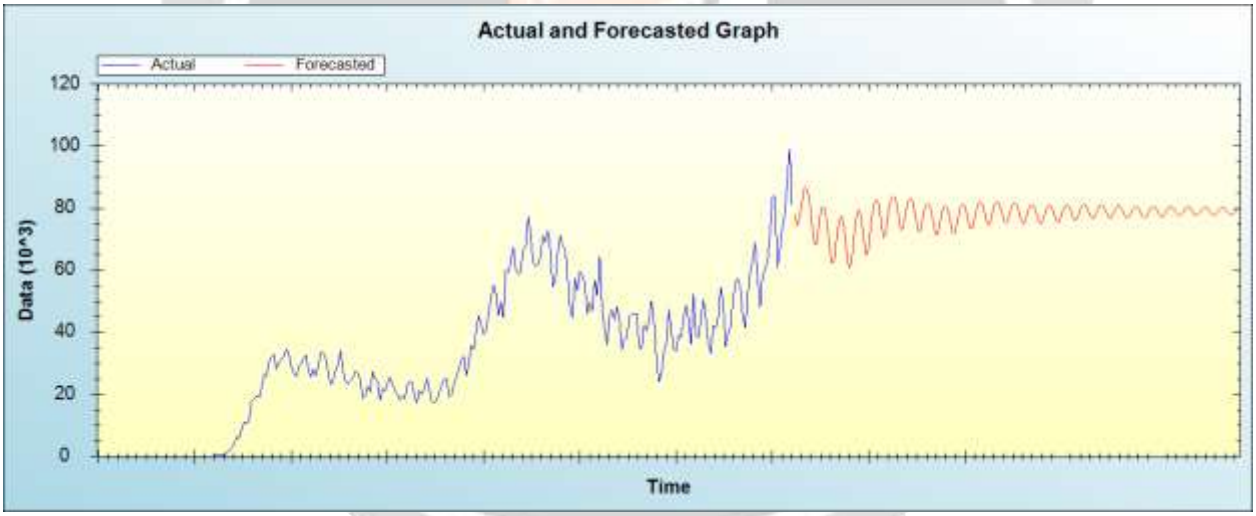
In-sample Forecast for A

Figure 2: In-sample forecast for the A series



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

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



Out-of-Sample Forecast for A: Forecasts only

Table 4: Tabulated out-of-sample forecasts

Day/Month/Year	Forecasts
01/11/20	78271.2683
02/11/20	74395.5417
03/11/20	77508.6314
04/11/20	81694.3825

05/11/20	86682.6878
06/11/20	86096.7465
07/11/20	82706.8924
08/11/20	74929.6045
09/11/20	68268.5650
10/11/20	68589.1135
11/11/20	76375.2989
12/11/20	80316.1631
13/11/20	80256.8134
14/11/20	76244.4746
15/11/20	68705.6795
16/11/20	62297.5395
17/11/20	62832.7950
18/11/20	70353.1368
19/11/20	76343.4805
20/11/20	77617.9541
21/11/20	74194.1604
22/11/20	66476.1028
23/11/20	60987.5447
24/11/20	62277.0744
25/11/20	70757.8517
26/11/20	77573.4541
27/11/20	79519.1052
28/11/20	77324.2433
29/11/20	70483.5333
30/11/20	65002.1219

01/12/20	67009.8001
02/12/20	75881.7245
03/12/20	81258.4492
04/12/20	82670.5049
05/12/20	81500.5045
06/12/20	76048.6274
07/12/20	70555.0118
08/12/20	72149.6044
09/12/20	79406.4054
10/12/20	83016.9195
11/12/20	83962.9494
12/12/20	82947.3577
13/12/20	78190.8179
14/12/20	73119.3856
15/12/20	73725.9094
16/12/20	79194.2382
17/12/20	82362.0035
18/12/20	83187.3435
19/12/20	81893.3233
20/12/20	77382.1150
21/12/20	72759.8404
22/12/20	72758.5105
23/12/20	77284.0808
24/12/20	80739.8472
25/12/20	81730.3963
26/12/20	80356.4800

27/12/20	76102.3055
28/12/20	71820.0313
29/12/20	71679.1408
30/12/20	75957.6986
31/12/20	79799.9578
01/01/21	81069.1450
02/01/21	79891.2263
03/01/21	76040.8803
04/01/21	72080.8687
05/01/21	72009.3135
06/01/21	76239.5186
07/01/21	80156.3261
08/01/21	81529.7911
09/01/21	80617.3408
10/01/21	77243.9767
11/01/21	73601.9975
12/01/21	73529.8183
13/01/21	77445.1208
14/01/21	81000.3989
15/01/21	82242.1608
16/01/21	81471.1847
17/01/21	78474.3519
18/01/21	75109.9045
19/01/21	74873.7083
20/01/21	78189.9290
21/01/21	81268.1286

22/01/21	82347.9959
23/01/21	81591.2824
24/01/21	78816.1844
25/01/21	75681.9433
26/01/21	75256.5388
27/01/21	78018.1055
28/01/21	80804.5502
29/01/21	81839.3132
30/01/21	81093.8673
31/01/21	78493.2971
01/02/21	75578.4340
02/02/21	75066.6932
03/02/21	77495.1664
04/02/21	80179.2676
05/02/21	81273.6943
06/02/21	80614.1445
07/02/21	78228.4595
08/02/21	75552.0194
09/02/21	75056.1130
10/02/21	77309.8241
11/02/21	79938.2747
12/02/21	81097.0747
13/02/21	80564.1620
14/02/21	78425.4079
15/02/21	75993.5622
16/02/21	75536.3762

17/02/21	77621.1230
18/02/21	80109.1693
19/02/21	81252.0421
20/02/21	80812.7312
21/02/21	78892.9424
22/02/21	76672.4717
23/02/21	76209.6460
24/02/21	78050.9467
25/02/21	80304.8514
26/02/21	81362.9454
27/02/21	80961.9530
28/02/21	79201.1481
01/03/21	77144.8185
02/03/21	76639.5764
03/03/21	78212.1756
04/03/21	80230.6921
05/03/21	81207.3211
06/03/21	80827.8559
07/03/21	79192.5483
08/03/21	77279.0139
09/03/21	76745.4978
10/03/21	78105.0120
11/03/21	79955.0509
12/03/21	80895.5987
13/03/21	80560.7289
14/03/21	79057.0944

15/03/21	77293.2438
16/03/21	76771.6046
17/03/21	77991.4799
18/03/21	79730.2151
19/03/21	80664.4323
20/03/21	80396.7919
21/03/21	79038.1699
22/03/21	77431.2787
23/03/21	76944.5578
24/03/21	78057.9058
25/03/21	79691.9344
26/03/21	80608.7332
27/03/21	80401.9389
28/03/21	79179.6157
29/03/21	77716.2792
30/03/21	77255.5663
31/03/21	78252.5583
01/04/21	79755.9420
02/04/21	80624.2999
03/04/21	80452.6318
04/04/21	79339.3983
05/04/21	77992.4223
06/04/21	77537.5319
07/04/21	78404.3192
08/04/21	79763.0077
09/04/21	80567.9117

10/04/21	80414.5219
11/04/21	79388.6930
12/04/21	78138.8685
13/04/21	77684.6431
14/04/21	78432.6417
15/04/21	79663.1768
16/04/21	80415.5146
17/04/21	80283.6989
18/04/21	79341.4741
19/04/21	78186.1370
20/04/21	77744.8167
21/04/21	78403.1733
22/04/21	79534.7590
23/04/21	80253.0085
24/04/21	80153.5053
25/04/21	79299.3065
26/04/21	78241.6281
27/04/21	77826.2161
28/04/21	78417.8782
29/04/21	79467.8374
30/04/21	80158.3147

Figure 1 shows that over the study period the minimum and maximum number of daily new COVID-19 cases is 0 and 99321 respectively. The average daily new cases are 32 360 infections. The distribution of the data over the study period is positively skewed and platykurtic, meaning that the data is not normally distributed. Table 4 and Figure 4 show the out of sample forecasts over the period November 1, 2020 to 30April, 2020.The forecasts indicate that daily new coronavirus infections will be generally constant (at equilibrium) and will be ranging from 75,000-80,000 daily new cases throughout the entire period from November 1, 2020 to 30 April, 2021. However, the daily new cases are likely to experience a negative shock in November from the events associated with United States Elections and its not surprising to see daily new cases

which are even higher than 80,000 in the out of sample period. Results indicate that the corona virus community transmissions are very high and this trend is going to continue in the out of sample period. Our findings are consistent with previous studies such as Seyed et al (2020) who concluded that the COVID-19 outbreak will most likely overwhelm the current hospital capacity and that expanding the number of ICU beds is an urgent issue.

CONCLUSION & RECOMMENDATIONS

The United States is currently battling the COVID-19 out break and facing numerous challenges in the control of the epidemic including, economic meltdown, rapid spread of the virus associated with community transmissions, failure by some communities to follow WHO guidelines to prevent the spread of the virus and the overwhelmed health delivery system. The findings of the study indicate that daily new corona virus cases will be ranging from 75,000-80,000 cases from November 1, 2020 to April 30 2021. However daily new cases may increase beyond 80,000 because of a negative shock from events associated with the US elections. The study recommends massive mobile COVID -19 testing, early isolation and treatments of infected cases ,continuous education among communities, regular hand washing with soap or use of hand sanitizers and wearing of face masks, tightening lockdown restrictions. If an effective vaccine is availed to the majority of the citizens, daily new COVID-19 will be expected to drop as a response to the positive shock of the vaccinations.

REFERENCES

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

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

[3] Gregory etal (2020). Fusing a Bayesian case velocity model with random forest for predicting COVID-19 in the US.Manuscript.pp1-51.

[4] Harcourt etal (20200.Isolation and characterization of SARS-COV-2from the first US COVID-19 case patient. BioRxivdol:101101/2020.03.02.972935.

[5] Holshue et al (2020). First case of 2019 novel corona virus in the United states.N.Engl.J.Med.382,929-936.

[6] Miller etal (2020). Disease and healthcare burden of COVID-19 in the United States. Nature Medicine .26.1212-1217.

[7] Moghadas et al (2020). Projecting hospital utilization during the COVID-19 outbreaks in the United States.Proc.Natl.Acad.Sci.117,9122-9126.

[8] Mollalo et al (2020). Artificial neural network modeling of Novel Corona virus (COVID-19) Incidence rates across the continental United States. International Journal of Environmental Research and Public health .17,4204,1-13.

[9] Pham. H (2020). Predictive Modeling on the number of COVID-19 Death toll in the United states considering the effects of corona virus -related changes and COVID-19 recovered cases. medRxiv pp1-19.

[10] Yuan et al (2020). Trends and Prediction in daily new cases and deaths of COVID-19 in the United States: An Internet Search -interest Based Model. Exploratory Research and Hypothesis in Medicine.000,000,1-16.