International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST207517

COVID-19: Detailed Analytics & Predictive Modelling using Deep Learning

Arjun Dutta[*1] , Aman Gupta[2] , Farhan Hai Khan[3]
[*1]R&D-Growth, Brickview Studios, West Bengal, India
[2]Business Head, Brickview Studios, West Bengal, India
[3]Student, WBUT, West Bengal, India

Corresponding Author E-mail:- : arjuncode47@gmail.com

Article Info
Volume 7, Issue 5
Page Number: 85-94
Publication Issue :
September-October-2020

Article History
Accepted : 10Sep 2020
Published :20 Sep2020

ABSTRACT

The rise of COVID-19 pandemicis alarmingly exponential in nature, and there is no certain predictability in the patterns of its future growth. This research aims at providing suitable insights into the current spread of the novel coronavirus (COVID 19) and conveys detailed analysis into current statistical growth rates of the infection frequency for diverse cases. We also provide schematic visualizations contrast and compare trends between various countries and India. We inspect closely and differentiate between the curves for the mortality rates globally. Also projected categorically most affected states of India.Finally,this research proposes methodologies for design and implementation of deep learning models for the predictive modelling of the various cases around the world and delivers predictions till March 2021 for confirmed and death cases.

Keywords:COVID-19 Predictions, Deep Learning, Exponential Growth, Infection Frequency Analysis, Coronavirus Trends

I. INTRODUCTION

The grievous situation due to zoonotic SARS-CoV-2 or COVID-19 pandemic started no sooner its outbreak in Wuhan, China during December 2019. COVID-19 pinpointed in early January and its genetic successionwas shared openly on 11-12th January [1]. SARS-CoV-2 exhibits distinctive epidemiological traits collated with SARSdurimg 2002-03 and MERSdurimg 2012. The virus rapidly spread a number of countries overseas, and was declared a Public Health Emergency of International Concern by the Director-General of the World Health Organization on 30 January 2020. Due to its contagious nature, the COVID 19 pandemic became the threat to mankind and changed the economic as well as social structure all over the world. Gradual and continuous increase in number of infected people and number of deaths forced to think to predict its effect in near future. There have been several studies on prediction or forecasting COVID 19 globally. In Egypt, a study has been performed to forecast about final size of the epidemic [2].

It was estimated in the study that the outbreak-spike could certainly reach at 22-June 2020 and rearmost

Copyright: © the author(s), publisher and licensee Technoscience Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited

time of the outbreak at 8-September 2020 with the count of, 1.6676E+05 cases in Egypt.

In another study, deep learning models of LSTM, GRU and Bi-LSTM were used to predict COVID 19. Without scaling, LSTM generates lowest MAE values for confirmed cases and deaths as 2.0463 and 0.0095, while RMSE values are 2.2428 and 0.0103, respectively, for China and Best value for re covered cases is for UK. The results from GRU gives best result for China with MAE values of 2.8553, 0.0321, and 7.04867 and RMSE values of 3.3158, 0.0402, and 8.4009 for confirmed cases, deaths and recoveries, respectively. Whereas among all countries, Bi-LSTM predicts best three cases for China with highest accuracy among all methods for all countries with lowest MAE and RMSE values of 0.0070 and 0.0077, respectively, for deaths in China [3]. Researchers have performed prediction on growth and trend of COVID-19 pandemic with the use of machine learning and cloud computing [4]. Their observation through using the iteratively weighted approach reveals that the Inverse Weibull functionfits the best to the COVID-19 dataset, rather than to the Beta (4-parameter),iterative versions of Gaussian, Log Normal functions or Fisher-Tippet (Extreme Value distribution). Iterative Weibull in their study showed an average MAPE of 12% lower than non-iterativelyweighted Weibull using the same dataset. In another study by Tomar A et al [5] prediction regarding spreading of COVID-19 in India was studied along with the effectiveness of preventive measures.

The data was analysed till April 2020. In another study the estimated pandemic life cycle along with the real data or history to date was demonstrated [6]. The proposed model monitors susceptible (S), asymptomatic (A), recovered (R), infected (I), isolated infected (Iq) and quarantined susceptible (Sq), collectively expressed SARIIqSq and are estimated from the real data on the COVID-19 pandemic in India till September 2020.Some renowned works also used AI tools to predict the spread of Covid-19.[8-11]. There has been a further need of such kind of study for future as the effect of pandemic is still of great concern. In the present study, Data related to COVID-19 has been thoroughly analyzed by giving crucial insights like the variation in recovery and mortality rate across different countries and globally.

II. PRESENT STUDY

Initially we discuss the currently top infected countries and gain insights in multiplicative trends being following in various curves using deep learning method. More specifically we observe the approximate time period for the values to get doubled up with a stark focus on cases in India. We display global heat-maps and geographical concentration maps that display the current severity of conditions across the continents.

In addition to these lucid visualizations, we have integrated AI predictive tools and neural network modelling techniques to foresee the possible surge in the number of COVID-19 patients in the future. Our findings reveal that the virus, under the absence of drugs and proper isolative measures will infect half of the global population by late January 2021, following current trends in the confirmed cases. Also, this research displays that the death cases would rise to above a million by mid-weeks of November 2020 and would reach over 6 million by mid- February 2021. These highly disturbing findings call for a major step towards virus control and explain the severity of the situations globally.

III.COVID 19 CASES-CATEGORIES

Here suggested recovered and death cases according to the data globally and in India.

A. Global

Figure 1 gives visualizations of summary of the current COVID-19 situation in the leading ten countries in the fields of Active, Recovered, Confirmed and Death Cases. USA, India, and Brazil are the worst affected countries currently(22/01/2020-07/09/2020)[7].

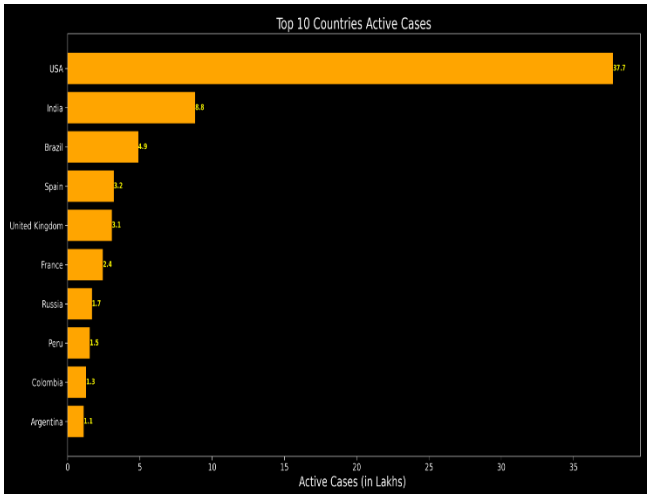

Figure 1 : Top 10 Countries Active Cases (in Lakhs)

USA is by far the worst hit country due to absence of precautionary measures in the early and later stages of the spread. India and Brazil follow, with mismanaged exercise of lockdown application.

Figure 2 shows the recovered cases in top ten infected countries. Brazil is the fastest recovering country due to its well-equipped medical facilities and advanced pharmaceutical technologies. India and USA follow, with a high population still struggling to be cured between (22/01/2020-07/09/2020).[7]

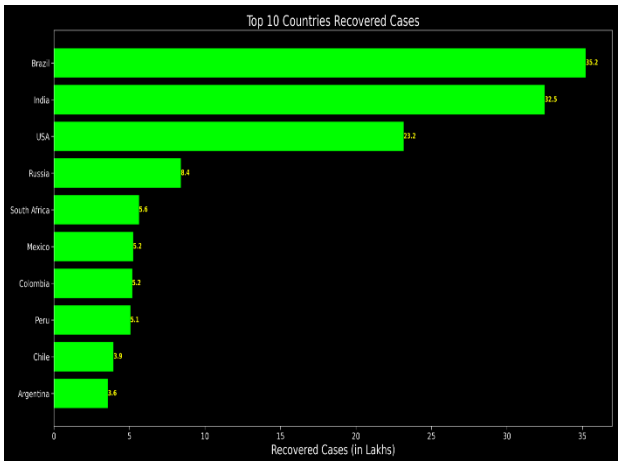

Figure 2 : Top 10 Countries Recovered Cases (in Lakhs)

Figure 3 shows confirmed case data in top ten infected countries. Recent surges in testing kits in the USA have led to a drastic rise in the number of confirmed patients. India and Brazil follow, with some increase in the same.[7]

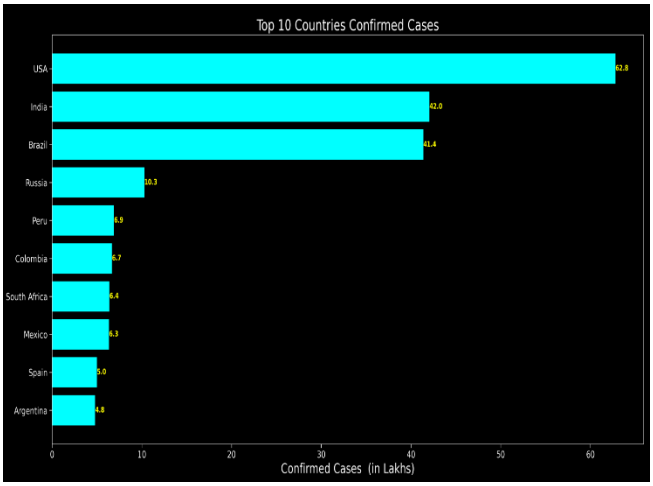

Figure 3 : Top 10 Countries Confirmed Cases (in Lakhs)

Figure 4 shows death rates in top ten infected countries. Massive deaths each day are being encountered due to lack of the presence of a drug. As the pandemic growth continues, death cases increase wildly. [7]

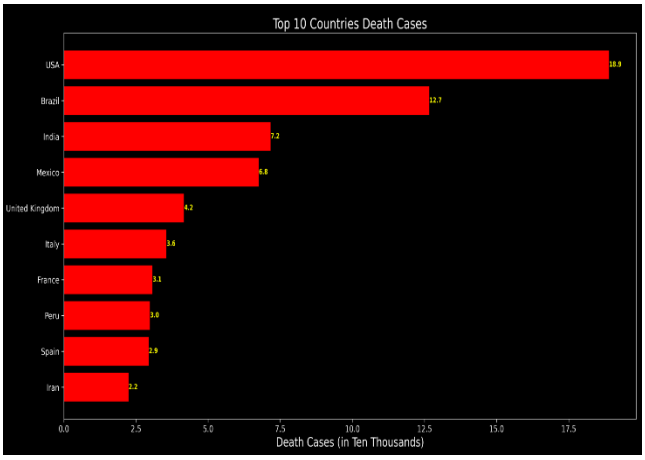

Figure 4 : Top 10 Countries Death Cases (in Thousands)

Table 1 shows the Summary of COVID-19 cases globally between (22/01/2020-07/09/2020). USA, Brazil and India are the worst affected as earlier visualized. [7]

Table 1: Global Covid 19 cases

CountryName	Confirmed	Deaths	Recovered	Active
Russia	1027334	17818	840997	168519
Peru	689977	29838	506422	153717
India	4204613	71642	3250429	882542
Mexico	634023	67558	524669	41796
France	362892	30707	88088	244097
U.K	349500	41640	1824	306036
Chile	422510	11592	394399	16519
Iran	388810	22410	335572	30828
Italy	277634	35541	210015	32078
Colombia	666521	21412	518229	126880
S.A	638517	14889	563891	59737
Argentina	478792	9912	357388	111492
Brazil	4137521	126650	3522155	488716
USA	6277994	188942	2315995	3773057
Spain	498989	29418	150376	319195

Upon the consideration of the threshold value-10 lakhs, the red boxes in the table indicates the alarming stage as it contains the most number of confirmed cases of corona virus, whereas among mass population in India, Brazil and USA the high recovery rate is also highlighted.

B. India

Figure 5 ,shows Top 10 Affected States in India. These visualizations summarize the current COVID-19 situation in the Indian Subcontinent in the fields of Confirmed and Death Cases. Maharashtra, & Tamil Nadu are the worst affected states currently. As testing centers are established in Maharashtra, the number of cases continually rise through, leading Andhra Pradesh & Tamil Nadu. Fig 6 shows death cases in India (state wise). The uncontrolled initial spread of the pandemic in Maharashtra has led to the surmounting deaths in its provinces leaving behind all states by a huge margin.

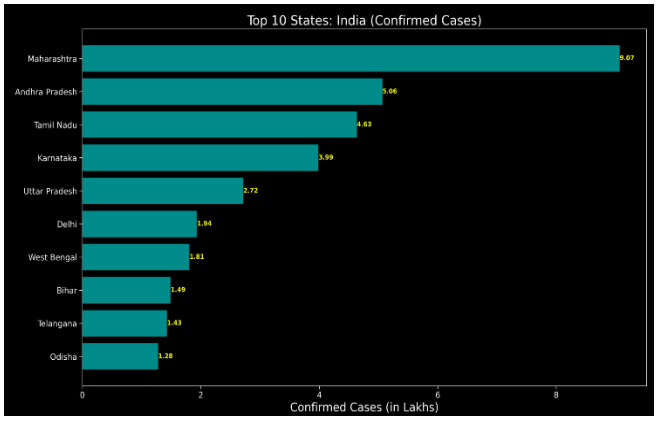

Figure 5 : Top 10 States India: Confirmed Cases (in Lakhs)

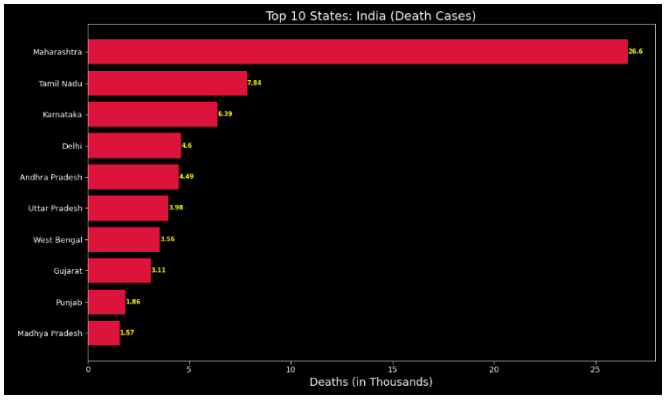

Figure 6 : Top 10 States India : Death Cases (in Thousands)

Table 2 summarizes the present (12/03/2020-07/09/2020)[14].COVID-19 situation in India statehouse. Maharashtra and Tamil Nadu are the worst affected as earlier visualized.

Table 2: Summary of COVID-19 India[14]

State Name	Confirmed	Deaths	Recovered	Active
Punjab	63473	1862	45455	16156
M.P	73574	1572	55887	16115
Delhi	193526	4599	168384	20543
Telangana	142771	895	110241	31635
Gujarat	104341	3106	84858	16377
Karnataka	398551	6393	292873	99266
Maharashtra	907212	26604	644400	235857
T.N	463480	7839	404186	51455
A.P	506493	4487	404074	97932
W.B	180788	3562	154008	23218
Odisha	127892	609	96364	30919
Bihar	149027	761	132145	16120
U.P	271851	3976	205731	62144

IV.COVID-19 Global Cases

The following plot that shows the exponential growth of the pandemic on the basis of daily confirmed cases globally. Here, we contrast and compare the total deaths vs the confirmed cases across the globe and can clearly see that the pandemic has a very low mortality rate. Figure 7 provides a basic understanding of the mortality rate increment of the pandemic per day. As clearly visible the mortality rate compared to the coronavirus spread is very low. Table 3 shows the overall exponential growth of the COVID-19 pandemic in the world indicating low mortality rates. (approximately 3.5%).

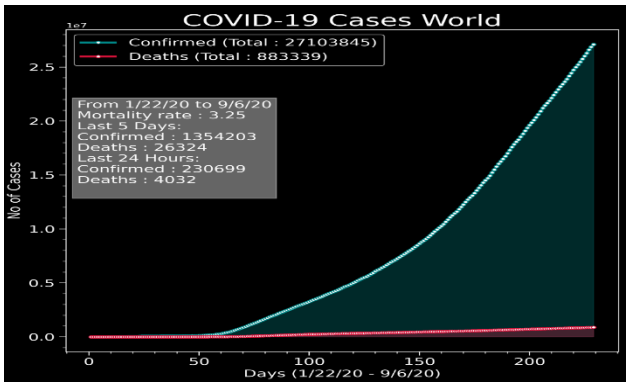

Figure 7 : Number of cases COVID-19 Cases globally

Table 3: Summary of COVID-19 Global[7]

Month (22nd day)	Confirmed	Deaths
January	555	17
February	78601	2459
March	342341	14853
April	2631027	189266
May	5217705	340178
June	9072474	471950
July	15228469	619821
August	23203532	804416

(22/01/2020-07/09/2020)

According to 22nd day of every month the data is collected and the global cases has been plotted in Figure 7.

A. Statistical Global Trends in Confirmed & Mortality Rate

Figure 8 shows the plotclaiming that the number of confirmed cases doubles approximately between 5 to 6 days &the number of mortality rates doubles approximately every week. This has been calculated using the line slopes and the respective gradients. These Trend Comparison Charts provide a calculative spread measure for the plots of the Continents, Countries and specifically provide comparison of other countries with India.

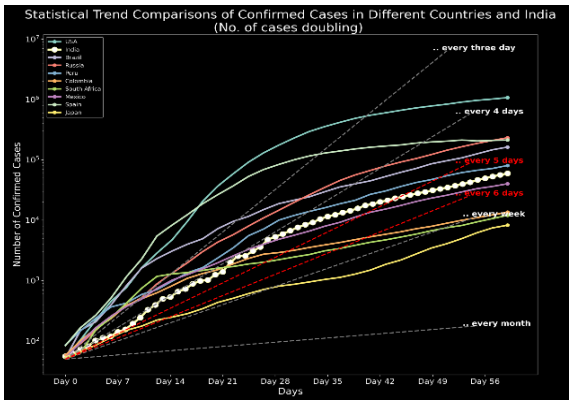

Figure 8 : Statistical Trend Comparisons of Confirmed Cases in Different Countries & India (No. of cases doubling) initially : The initial increment of the confirmed cases for different countries is shown for the first 60 days from when the virus infestation occurred.

For India the number of confirmed cases is rising to double of its value at an approximate interval of every 5-6 days. USA has a highly positive steep slope which evaluates its doubling capacity at each 3-4 days. The distributions display the hypercritical situation of the confirmed cases curve, which is enlarging by a double value in such a short amount of time.

The curve for USA has a greater exponential influence than other countries currently leading to its more threatened future. India is intermediate for this category and attempts should be made to flatten the curve.

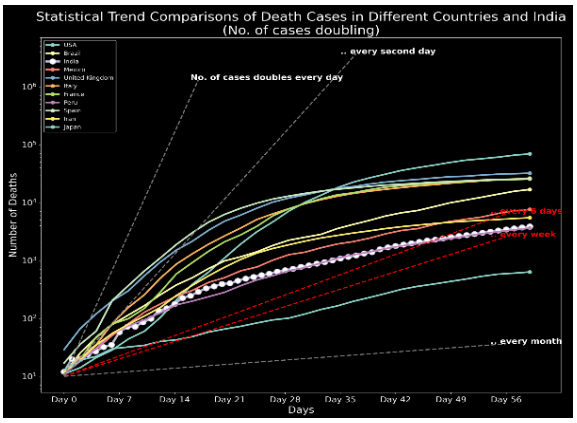

Figure 9 : Statistical Trend Comparisons of Death Cases in Different Countries & India (No. of cases doubling) : The initial increment of the death cases in different countries is shown for the first 60 days when the virus infestation first occurred.

The white dotted mark in Figure 9, shows for India the number of death cases is rising to double of its value at an approximate interval of every 6-7 days. USA and most other countries have a higher positive steep slope which evaluates in their doubling capacity at less than 6 days. The number of deaths is also increasing, have entered the high value ranges due to lack of the presence of a drug and more importantly lack of suitable precautionary measures, especially in the USA.

B. Map Visualizations

This part of the research displays an interactive map that helps to visualize the geographical distribution of the pandemic globally and in India, visually in the form of color mappings and circle radiuses as referred in Figures (10,11).

Figure 10: Color Mapping-Global

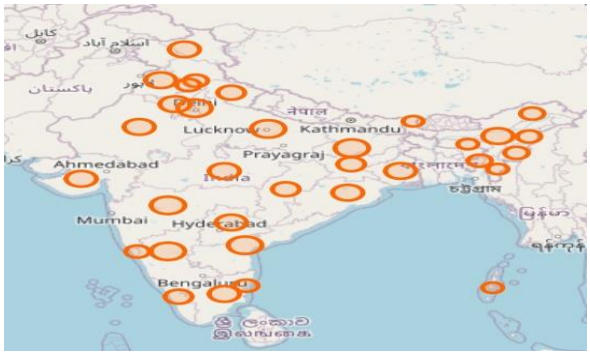

Figure 11 : Color Mapping-India

V. COVID-19 Predictive Tools

In the present study, an attempt to predict the Confirmed Cases and the Deaths all around the globe using AI tools has been performed. The model architecture followed is discussed as follows.

A. Model Architecture

The model architecture that we choose here (as shown in Figure 12) consists of recurring units of a Linear Dense Layer combined with a Leaky ReLU layer.

The Model Hypothesis Function used in the present study as follows:

Using the Dense layer [13],

$D_n(i) = f(w_n * i + b_n)$

where,

i is the layer input, D is the layer output, w are the network weights, n is the number of parameters, b is the network bias, & f is the activation function.

In our case the activation function is linear, i.e.,

$f(\phi)=\phi$ (Adaline, Linear Regression)
This simplifies the dense layer equation to,

$D_n(i) = w_n * i + b_n$

The LeakyReLU [12] layer,

$L(z)=z\alpha$, when $z>0$ &

z , when $z<=0$
where,

z is the layer input, L is the layer output, & α is a small, non-zero, constant gradient.
Here we used the value of $\alpha=0.3$ as defined by default in the Keras API.

Following the Neural Network Architecture, the model hypothesis over the scalar input m can be written as,

$h\theta(m)=L(D_1(L(D_{80}(L(D_{80}(L(D_{80}(L(D_{80}(m))))))))))$

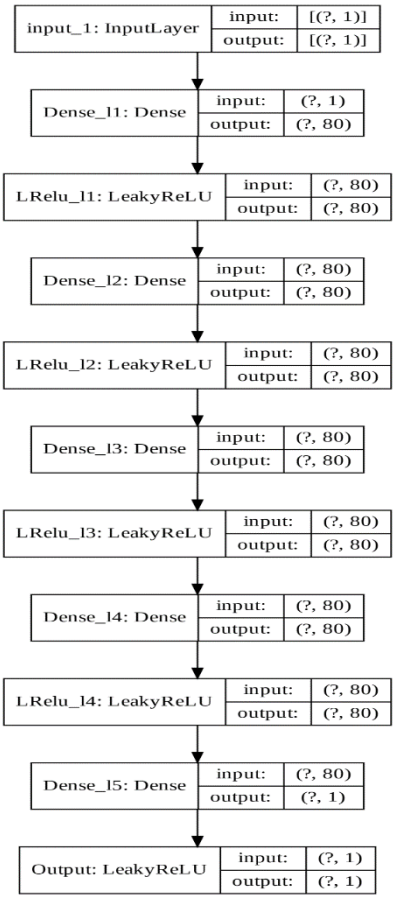

Figure 12: Model Architecture of the present study

B. Model Training Pipeline

Let the input value (date) be represented by x and the output value(no of global death cases/no of global confirmed cases) be represented as y.
Since the curve follows an exponential nature ($y=e^x$), we first take the logarithm to simplify

the hypothesis and fit a relatively simple model to the problem than using the original unmodified output values.
Here we use the value of
Let,
$y=10^t$ or $t(y)=\log_{10}(y)$

We will feed x as input and t as output into the neural network and find the hypothesis function,
$t(y)=h\theta(x)$

Simplifying, $\log_{10}(y)=h\theta(x)$

And finally predict using the following equation,

$y=10^{h\theta(x)}$
Hence the final equation becomes,
$y=10^{L(D_1(L(D_{80}(L(D_{80}(L(D_{80}(L(D_{80}(m))))))))))}$

C. Model Predictions

The Global Confirmed Cases Visualizations obtained after the predictions are shown in Figure 11. The exponential curve obtained using our model hypothesis postulates that the number of confirmed cases will continue to grow at a large scale in the absence of a drug. These values are alarming as the predictions highlighted in red are the predictions exceeding half of the global population itself. Our study, therefore indicates that half of the whole world population (39 Crores) will get infected by 22nd January, 2021 if the current trends. Therefore the need for isolation and the creation of a drug is vital to stopping the spread the novel coronavirus.

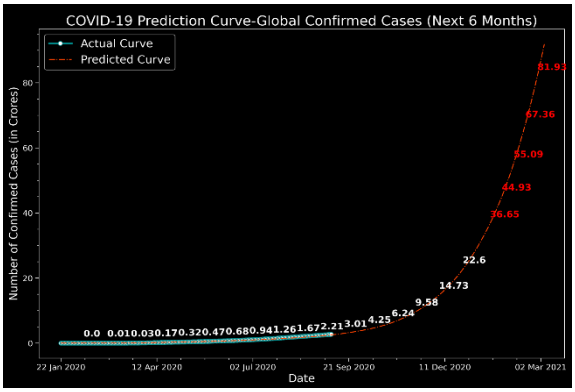

Figure 13 : COVID-19 Prediction Curve-Global Confirmed Cases (Next 6 Months 07/09/2020-05/03/2021)

Also disturbing are the final prediction values, for 5th March 2021, which comes out to be 91.91 Crores! This abnormally high value surely poses some problems and displays that the novel coronavirus is a threat to all of humanity. Figure 12 shows The same model hypothesis was applied to get these figures which demonstrate the steep exponential nature of the coronavirus pandemic spread.

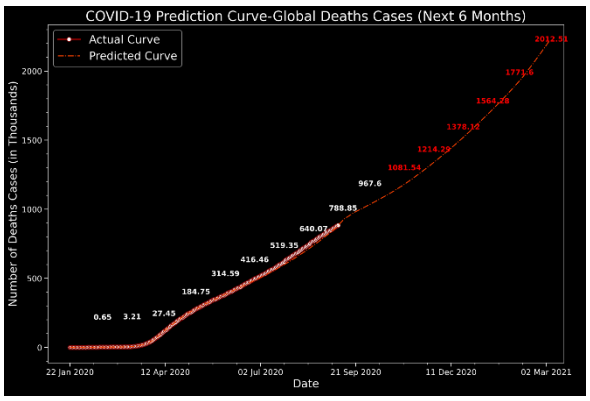

Figure 14 : COVID-19 Prediction Curve-Global Death Cases (Next 6 Months 07/09/2020-05/03/2021)

The COVID-19 death cases will reach a million and rise (red) after 25th September 2020 as predicted by our study (shown in Figure 13,14). Hence substantial measures must be taken by the government to treat patients and find a drug/ vaccine before such

speculated dates. Here are some of the numerical results obtained, after applying the model inference after training the model and running the inferences.The crucial findings display that the pandemic will reach unbound values without proper control, taking over the whole planet by February.

VI. CONCLUSION

In this study we have performed a research upon the various categories of COVID-19 cases globally and visualized the countries being worst affected by the pandemic. State-wise visualizations over the Indian subcontinent was also done, followed by the mortality rate calculations of the curve globally. Hereafter, we contrast and compare the doubling of values of the confirmed and death cases with various countries and India and deliver the approximate time frequency for the COVID-19 curve values to reach twice it's heights. We perform preliminary observations using data analytics and gain important insights into rates and periods of growth of the coronavirus trends.

Finally, we predict Global Prediction Curves for the Confirmed Cases and Death Cases using a simple deep learning model highly adapted for exponentially incremented values. The resulting figures were disturbing and lead us to think about stabilizing the pandemic condition.

VII. REFERENCES

[1]. Coronavirus disease 2019(COVID-19) Situation Report-94,www. who.int/docsdefaultsource/coronaviruse/situation-reports/20200423-sitrep-94-covid-19.pdf,Accessed 10/09/2020.

[2]. Amar L.A., Taha A.A. & Mohamed M.Y., Prediction of the final size for COVID-19 epidemic using machine learning: A case study of Egypt, Infectious Disease Modelling (2020), doi: https://doi.org/10.1016/j.idm.2020.08.008.

[3]. F. Shahid, A. Zameer and M. Muneeb, Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM, Chaos, Solitons and Fractals 140 (2020) 110212

[4]. S. Tuli, S. Tuli and R. Tuli et al. / Internet of Things 11 (2020) 100222 Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing

[5]. A. Tomar, N. Gupta, Prediction for the spread of COVID-19 in India and effectiveness of preventive measures, Science of the Total Environment 728 (2020) 138762

[6]. K. Sarkar, S. Khajanchi and J.J. Nieto, Modeling and forecasting the COVID-19 pandemic in India, Chaos, Solitons and Fractals 139 (2020) 110049

[7]. Dataset Used for Research : Center for Systems Science and Engineering (CSSE) at Johns Hopkins University , COVID-19DataRepository,github.com /CSSEGISandData/ COVID-19 ,Accessed 07/20/20.

[8]. Prediction of Spreads of COVID-19 in India from Current Trend, Himanshu Shekhar , doi.org/10. 1101/2020.05.01.20087460

[9]. Forecasting the novel coronavirus COVID-19, Fotios Petropoulos ,Spyros Makridakis , doi.org/10.1371 /journal.pone.0231236

[10]. Coronavirus (COVID-19) Visualization & Prediction ,XingyuBian, kaggle.com /therealcyber lord/ coronavirus-covid-19-visualization-prediction

[11]. COVID-19 Outbreak Prediction with Machine Learning, SinaF. Ardabili,Amir MOSAVI,Pedram Ghamisi,doi: https://doi.org/10.1101/2020.04.17.20070094

[12]. Empirical Evaluation of Rectified Activations in Convolutional Network, Bing Xu, Naiyan

Wang, Tianqi Chen, Mu Li, https://arxiv.org/abs/1505.00853

[13]. Densely Connected Convolutional Networks, Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger, https://arxiv.org/abs/1608.06993

[14]. India JSON Dataset: COVID-19 REST API for India ,api.rootnet.in/covid19-in/stats/testing/history , Accessed 07/20/20.

Cite this article as :

Arjun Dutta, Aman Gupta, Farhan Hai Khan, "COVID-19 : Detailed Analytics & Predictive Modelling using Deep Learning", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 7 Issue 5, pp. 85-94, September-October 2020. Available at doi : https://doi.org/10.32628/IJSRST207517
Journal URL : http://ijsrst.com/IJSRST207517