NOVEL CORONAVIRUS (2019-NCOV): IMPACT OF GLOBAL PANDEMIC AND PREDICTIVE ANALYSIS USING MACHINE LEARNING

Kanu Goel
Department of Computer Science and Engineering
Thapar Institute of Engineering and Technology, Patiala, India

Shefali Arora
Division of Computer Engineering
Netaji Subhas Institute of Technology, Delhi, India

Abstract— **Novel coronavirus (2019-nCoV) is known for causing severe respiratory and system infections in a variety of hosts. Bats have been believed to be the source of this deadly outbreak. Since the beginning of its transmission, there have been more than 10 million confirmed cases of the coronavirus disease 2019 and around 544000 deaths have been reported to date. According to the World Health Organization (WHO) around 196 countries, areas and territories have been affected by this disease. Billions of people are under complete lockdown to contain its spread which has resulted into shutdown of various industries, loss of business, unemployment and global economic crisis. In this paper, we review and analyse the impact of coronavirus across communities of various countries while particularly highlighting its possible impact on India. We focus on the challenges and lessons India needs to learn from other countries so as to contain the transmission of this virus with some ideal preventive measures. From the available data on Covid-19, we analyze the patterns of deaths and recoveries for various countries using machine learning models like Support Vector Regression, Linear Regression and neural networks. Our models are quite effective in predicting the number of deaths and recoveries.**

Keywords— **Covid-19, Deep Learning, India, Regression**

I. INTRODUCTION

On 12th December, 2019, 27 cases of viral pneumonia were reported in Wuhan. Most of the people had exposure to wildlife animals at a wholesale market in China. It was gradually identified as the 2019 n-CoV virus by WHO.
The pandemic has been spreading widely since then. Rarely has the world witnessed such an unprecedented chaos in recent modern times. The chaos which finds very few parallel in history is being caused by something which one cannot even see i.e. COVID-19. The pathogen of severe acute respiratory syndrome coronavirus 2 or (SARS-CoV-2) is officially known to cause severe pneumonia and primarily attacks the respiratory system (World Health Organization,2020) in the later stages it might even cause a major multiple organ failure. This virus has four strains: alpha, beta, delta and gamma, out of which only alpha and beta can be transmitted to humans. This belongs to a category of virus which is transmitted from animals to humans (also known as Zoonotic) and has its origin in bats (Huang et al.,2020).

Cases of COVID-19 are no longer limited to Wuhan and the widening geographical spread is leading to major concerns regarding its future trajectory. The pandemic has led to the implementation of extreme measures so as to contain its spread. In this paper, we review the various challenges faced by different countries including India, and the preventive measures being taken. Section 1 explains the scenario of Covid-19 since 2019, and the challenges faced and tackling it by the Indian government. This is followed by the analysis of COVID-19 according to the global perspective in Section 2. Section 3 depicts the analysis to predict the number of deaths and recoveries corresponding to the number of cases in five different countries using three machine learning algorithms. Section 4 visualizes the results and Section 5 concludes the paper while stating the future work.

A. Transmission of Coronavirus since 2019

On 31st December, 2019 WHO was reported of several cases of unusual pneumonia in Wuhan province, which they suspected to be a repeat of SARS. On 5th January, 2020 this theory was ruled out and on 7th January, 2020 a new strain of virus was identified as COVID-19, belonging to Coronavirus family (Cheng and Shan,2020).
Coronaviruses are common and spread through being in proximity to an infected person and inhaling droplets generated when they cough or sneeze, or touching a surface where these droplets land and then touching one's face or nose (Centers for Disease Control and Prevention,2020). Later on January 11th the first death due to Coronavirus was reported. On January 13th a first case from outside China was reported in Thailand. On January 16th, a case was reported in Japan. Soon after that other countries of the world such as United States, Italy, France, Australia, Malaysia, Singapore, South Korea, Vietnam and Taiwan confirmed cases over the following
days. WHO declared this outbreak as a Public Health Emergency of International Concern on 30 January

2020(Malik et al. 2020) and a Global Pandemic on 11 March, 2020 when epicentre shifted from China to Europe. Since then, cases have been increasing exponentially and are leading to worldwide loss of human lives (Lopez-Martin et al., 2017). Further, we discuss the scenario of India in the midst of this crisis and the manner in which it is dealing with the containment of the virus.

B. Steps taken by India to combat Covid-19

- Government of India has set up 24*7 helpline to answer any queries and provide assistance in regards to COVID-19. Along with this central and state wise helpline numbers were also issued.
- India started airport screening of passengers with any travel history from countries like China, South Korea, Japan, Singapore etc.
- When the virus started spreading to Iran and Europe, then passengers from Iran, Germany, Spain, France were also been screened at the airport itself to identify the infected persons. Approximately 1.52 million passengers were screened at various airports in India.
- Thereby the infected persons were kept in isolation for 14 days and those with travel history from COVID-19 affected countries were advised to self-quarantine themselves for 14 days.
- As the number of infected cases began to arise and crossed 0.25 million globally and 200 in India, Prime Minister of India urged the citizens of India to observe ``Janta Curfew'' on 22 March, 2020 from 7 am to 9 pm. This was the curfew imposed by the citizens themselves without any enforcement of police/army aimed towards breaking the chain of transmission of the virus.
- As the citizens of the world's second most populous nation observed a complete shutdown, making the ``Janta Curfew'' very successful. This lockdown was extended till 3rd May, 2020 by majority of Indian states. However, during these curfew essential services like grocery, milk, medical facilities, transportation of essential needs was permitted. To support the livelihood of farmers, daily wagers, construction workers ,Govt. of India has introduced a relief package of Rs 170,000 crores (equivalent to 22.53 billion USD).
- Meanwhile all passenger rail services, including Delhi metro and Inter-state bus transport system was shutdown w.e.f. 23 March, 2020.
- To make personal care, sanitation and hygiene products more affordable for public, Government has capped the prices of hand sanitizers and masks on 21st March, 2020 till 30 June, 2020.
- To cater to the increasing number of people getting infected by the novel Coronavirus, Government has also roped in private laboratories to conduct tests for COVID-19 while issuing strict guidelines for testing and setting price caps for the same. With this capacity to undertake tests increased to 60,000 to 70,000 per week.
- International passenger flight service has been discontinued w.e.f. 22 March, 2020 till 31 May, 2020 while domestic flights have also been stopped from 25 March, 2020.

C. Challenges faced by the government

- India faces several hurdles in fighting the pandemic COVID-19 due to its large population, high population density and low testing rate due to limited health care infrastructure.
- Casual attitude of some section of society towards the gravity of situation and lack of realization of the impact it would have.
- As about 15% of population lives below poverty line (BPL), with most of them being daily wagers; any lockdown would hamper their earnings which may make them reluctant to support lockdown.
- It is advised that Govt. should test all the people who came to India from abroad in March, 2020 to ensure 100\% coverage of infected persons.

D. What India can learn from other countries

- All the states of India should identify health facilities earmarked exclusively for COVID-19 infected patients.
- All states should assess the current availability and production of the face masks, protective suits required for doctors, health workers and ventilators for critical patients. If need arises, production capacity should be increased with the help of prominent industrialists. In Wuhan, various automakers were roped in for the production of masks and hand sanitizers which increased its production by ten times. Automakers in US were asked to manufacture the ventilators to cater to the growing need for infected persons.
- Pharmaceutical companies should be asked to increase the production of medicines used by doctors and hospitals for the recovery of patients as per requirement.
- It must be ensured that contacts of COVID-19 positive patients must be traced and tested at the earliest to break the chain of transmission. For example in Germany, the testing rate was approximately 42,000 cases per week, efficiently covering most of the COVID-19 suspects. This has proved to be the reason behind low mortality rate of just 0.4% in contrast to Italy which is close to 9%.

- If situation improves in coming one or two months even then the foreign travel restrictions and screening of passengers must be kept continued with strict measures to prevent COVID-19 from resurfacing just the way some such cases have been reported from China.
- Shutdown should be imposed effectively; any casual attitude, even if by small number of citizens, must be dealt with utmost seriousness.

In Table 1, we analyse the global situation, in terms of the number of cases, deaths and recoveries due to COVID-19.

Table 1. Summary of global situation of Covid-19 uptil May 2020

S.no	Country	No of cases	Deaths	% Deaths	Recovered	% recovered
1	China	83966	463	5.52	78792	93.84
2	Italy	211938	29079	13.72	82879	39.11
3	United States	1180634	68394	5.84	187180	15.85
4	Germany	166152	6993	4.21	135100	81.31
5	India	46437	1566	3.37	12847	27.67
6	Iran	98647	6277	6.36	79379	80.49
7	South Korea	10804	254	2.35	9283	85.92

Figures 1 and 2 can be used to visualize the recovery and death patterns respectively in the specified countries upto the month of May.
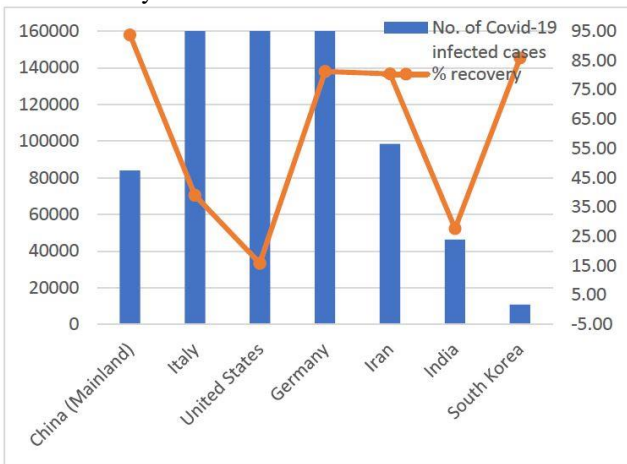

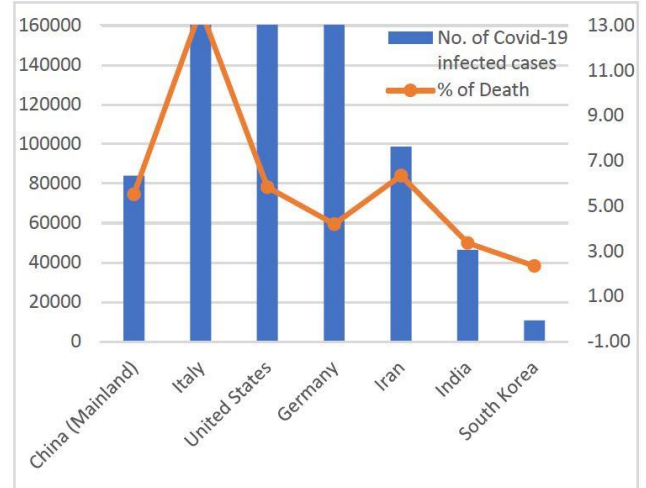Figure 1. Number of Covid-19 infected patients vs a)\% No. of patients recovered

Figure 2. Number of Covid-19 infected patients vs a)\% No. of patients deaths

II.   GLOBAL ANALYSIS OF THE COVID-19 SPREAD

From the available data, we analyze the crisis faces by various countries of the world. Seven countries (namely China, Italy, United States, Germany, Iran, India and South Korea) have been chosen for analysis as all these seven countries have diverse pattern in terms of number of infected cases, recovery rate, death rate and other demographic factors. Global scenario of death rate and recovery rate due to COVID-19 is given in Table 1 Figure 1 and Figure 2 depict the analysis of number of patients recovered and death rate respectively for COVID cases. According to this analysis, the highest mortality rate as far is of Italy with 13.72% deaths. It can be majorly attributed to the delayed response system of the local government. Apart from this, high percentage of old age population (>60 years) amongst the total infected. Old citizens accounted nearly 96% of Italy's total mortality rate. In Germany, though initially COVID-19 was declared as very low health risk and no travel advisory was issued till 22 January, 2020 but it had a huge testing rate. 42,000 tests per week right from the beginning. Due to this the death rate is quite low as all the suspected cases were identified timely. Next, highest recovery rate is found in China which is about 94%. As most of the infections were found about 1.5-2 months back, it may be assumed that it generally takes six-weeks time for recovery. South Korea introduced largest and best organized program in the world to screen the population for the virus and isolate infected people. They developed drive through test centres and entire procedure of collecting a sample takes less than ten minutes while the suspected patients remain in the vehicle itself. It resulted in checking the spread of pandemic successfully. Iran announced its first confirmed on 19 January, 2020. It was the first middle-east country to report coronavirus. As the first case was reported very late and mortality rate is high, it may indicate that the first outbreak might have happened earlier and there was delayed reporting. In India as the number of

confirmed cases is not very high as compared to other countries, it still has time to learn from the mistakes of others and to adopt best practices if suitable as per the local conditions to limit the spread of COVID-19.

In the next section, we highlight the contributions utilize the data obtained on a global level and predict the number of deaths and recoveries corresponding to the number of cases in five different countries.

A. Contribution

The main highlights of our research work are:

- We analyze the growing trend of coronavirus cases in different parts of the world, especially countries which are the most affected.
- Using various machine learning and deep algorithms, we predict the number of deaths and recoveries corresponding to the number of cases. We focus on the rising rates in India in the midst of these prevailing conditions.
- Our proposed deep learning architecture is efficient in predicting the number of deaths and recoveries by learning from the information corresponding to the historical data.

III. ANALYSIS: PREDICTION OF NUMBER OF DEATHS AND RECOVERY RATE USING MACHINE LEARNING

Machine learning (ML) has given us the answer to a number of daily life questions. Using ML algorithms, we can predict unforeseen problems and challenges. On analyzing the data obtained from different countries, we can predict the mortality rate with the rise in number of cases. Thus governments can estimate unforeseen challenges ahead of them. Linear Regression, Support Vector Regression and neural networks (Multilayer Perceptron) (Lopez-Martn et al.,2017)are three popular algorithms which are used to predict continuous target values, if an input value is available(Kavitha et al.,2016). Given the number of cases every day, we use these three machine learning algorithms to predict the number of deaths and recoveries proportional to this number in five different countries. The working of these algorithms has been explained ahead.

A. Linear regression (LR)

Given data with n-dimensional variables and a target variable, Eq. 1 is used to represent
the feature space where $x \, \varepsilon \, Rn$, $y \, \varepsilon \, R$.

$$(x1; y1); (x2; y2)…..(xn; yn) \qquad (1)$$

The objective of the function used is to find the best fit by minimizing the sum of squared errors(Browne,1975). In linear regression, the relationship with x and y is considered to be linear and the model is represented as Eq. 2:

$$f(w1….wn)= y= w.x + b \qquad (2)$$

The best fit can be found by using Eq 3 which finds the sum of squared errors between real and predicted values.

$$min \sum_{i=1}^{m}(y_i - \hat{y_i})^2 = min \sum_{i=1}^{m}(y_i - (wx + b))^2 \qquad (3)$$

B. Support Vector Regression(SVR)

SVR finds a function f(x) such that it is at a deviation of atmost $\in$ from the target variable y. This can be treated a sconvex optimization problem represented in Eq. 4-6.

$$minimize \frac{1}{2} ||w||^2 \qquad (4)$$

subject to:

$$y_i - (w, x_i) - b \leq \in \qquad (5)$$
$$(w, x_i) + b - y_i \leq \in \qquad (6)$$

The assumption was that such a function f actually exists in which convex optimization problem is feasible with all the given variable and target pairs.

C. Proposed Multilayer perceptron network

A deep neural network can be used to predict the value of target variable y from input variable x. It consists of various hidden layers, and the use of deep networks improves the process of regression. A neuron can be represented as:

$$y=wx+b \qquad (7)$$

Where x is the input variable, w is the weight associated with the neuron and b is the bias used. y is the value of the target variable. An activation function a is applied to the output, such as tanh or Rectified Linear Unit to produce the final output , as presented in studies (Choubin et al.,2016) (Singh and Banerjee,2019)

$$a=g(y) \qquad (8)$$

The application of ReLu is the creation of a non-linear boundary by disregarding all the negative values sampled during learning of features from the data.

$$y=max(0,y) \qquad (9)$$

The multilayer perceptron framework proposed in this paper is composed of the following layers as shown in Table 2. The first hidden layer consists of 128 neurons, followed by two layers of 256 neurons and one layer of 512 neurons respectively. The final layer comprising of a single neuron predicts the number of deaths and recoveries by learning from the previously extracted information.

Table 2. Deep neural network architecture

Layer	Number of neurons	Activation
1	128	ReLu
2	256	ReLu
3	256	ReLu
4	512	ReLu
5	1	NA

D. Experimental setup

We have pre-processed the data obtained for following five countries, followed by predictive analysis using support vector regression, linear regression and deep neural networks. The pandemic data has been collected from Wikipedia. The countries, on which analysis of Covid-19 has been performed, are as follows:

- India
- Italy
- China
- USA
- Iran

The data is split into two parts: training and testing in a ratio of 60:40. The trained models are tested on unforeseen data to predict the number of deaths and recoveries corresponding to the number of cases. We conduct the following experiments in Python, where the pre-processed data is loaded into models trained on Google Colab (Colaboratory,2020).

Once the models are trained, we use them against a particular test case to foresee the number of deaths. We consider a particular test case and observe the number of deaths predicted vs the actual number of deaths observed in a particular country corresponding to thoe number of cases. Table 3 depicts the estimated number of deaths using linear regression, vis a vis the actual number of deaths. Similarly, Table 4 foresees the number of recoveries using SVR for a particular test case for three different countries India, Italy and China vs the actual number of recoveries. It is observed that SVR performs better as compared to LR.

This is because SVR deploys the use of kernels to solve non-linear fitting problems. Whereas, linear regression follows a linear approach while modeling the relationship between independent and dependent variables.

Table 3. Actual vs Predicted deaths for countries using LR

Country	Cases	Actual deaths	Predicted deaths
India	283	4	4
USA	1598	42	31
Italy	12462	827	816
China	80409	3012	3023
Iran	3513	107	80

Table 4. Actual vs Predicted recoveries for countries using SVR

Country	Cases	Actual recoveries	Predicted recoveries
India	41	4	5
USA	9172	622	596
China	80409	52045	47659

We use Mean Squared Error (MSE) as a metric to estimate the performance of the model. The value of MSE for our trained Support Vector Regression (SVR) model is 0.6. T MSE value is 1.5 for the linear regression model. This proves that the

analysis obtained using SVR gives better insights as compared to the LR model.

Further, we have elaborated the scenario of India in the ongoing crisis of Covid-19. We apply the SVR model to predict the number of deaths and recoveries in India vis a vis the total number of deaths and recoveries corresponding to the number of cases.

Table 5 depicts the number of predicted deaths and recoveries corresponding to the actual values in India using SVR.

Table 5. Actual vs Predicted recoveries for countries using SVR

Cases	Actual deaths	Predicted deaths	Actual recoveries	Predicted ecoveries
97	2	1	13	12
107	2	1	14	14
118	2	2	14	15
137	3	2	14	18
151	3	3	22	19
173	4	3	22	21
223	4	3	22	21
283	4	4	23	22
360	7	6	23	22
434	9	7	24	23
519	9	9	41	41
606	12	11	43	42
16116	508	800	2302	4000

E. Advantage of using Deep Learning for Analysis

Whereas machine learning algorithms like SVR and LR are efficient in learning from the data, deep neural networks outperform the use of these algorithms. This is because of the presence of hidden layers in the perceptrons helps to extract features from the data efficiently. The use of non-activation functions helps to model the relationship between the independent and dependent variables in a non-linear manner. Tables 6 and 7 depict the number of deaths and recoveries vs the actual number in India, respectively. These have been analysed with the help of our deep learning model explained in Table 2.

Table 6. Actual vs Predicted deaths for India using Deep Learning(DL)

Cases	Actual deaths	Predicted deaths
3	3	2
283	23	21
23452	723	604
40623	1306	1035
12759	420	328
16116	519	415
31787	1008	820
42836	1389	1120

Table 7. Actual vs Predicted recoveries for India using Deep Learning(DL)

Cases	Actual recoveries	Predicted recoveries
5865	468	760
6761	516	719
23452	4814	4123
28380	6362	5063
40263	10887	7330

It is observed that the mean squared loss achieved is below 0.5, which is a better estimate as compared to the SVR model. The model is trained for 50 epochs and Adam optimizer is used to minimize the difference between the actual and predicted values.

IV. VISUALIZATION OF ANALYSIS

The visualization of predicted deaths vs actual deaths using SVR is shown in Figures 3,4,5,6,7. The x-label depicts the number of cases and y-label predicts the number of deaths or recoveries. The actual values are normalized during training.
Whereas solid(blue) line is used to visualize the actual number of cases, dots(red) are used to visualize the predicted values. These figures show the efficiency of SVR in prediction of deaths corresponding to the number of cases in these five countries. Whereas Figures 8,9,10 visualize the same prediction using LR.
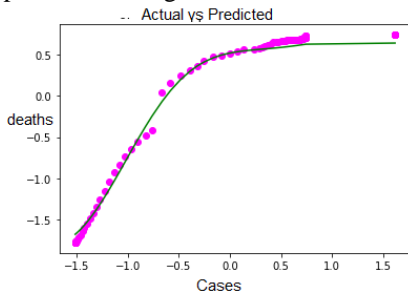
Fig.3. Actual vs Predicted deaths in China using SVR
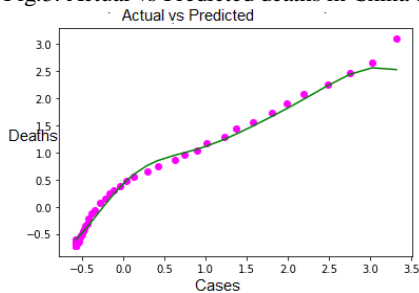
Fig.4 Actual vs Predicted deaths in India using SVR
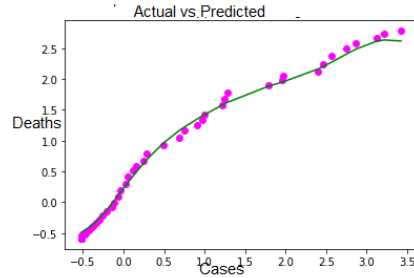
Fig.5. Actual vs Predicted deaths in Italy using SVR
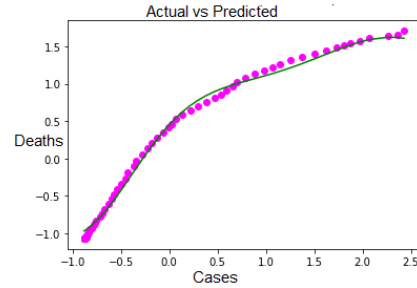
Fig. 6. Actual vs Predicted deaths in USA using SVR
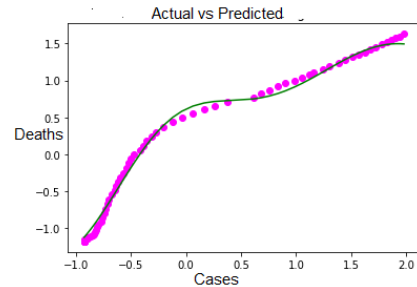
Fig. 7. Actual vs Predicted deaths in Iran using SVR
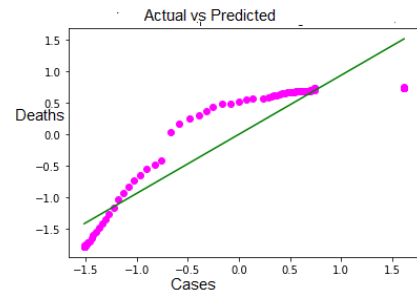
Fig. 8. Actual vs Predicted deaths in China using LR
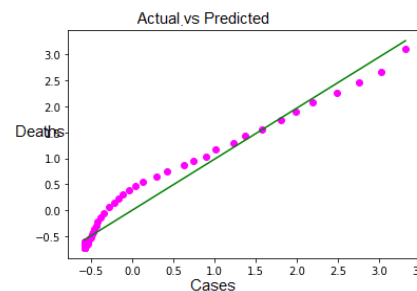
Fig. 9. Actual vs Predicted deaths in India using LR
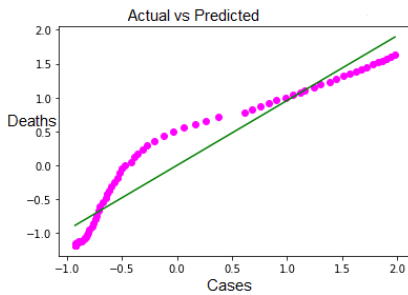
Fig. 10. Actual vs Predicted deaths in Iran using LR

Figures 11,12,13,14,15 analyze the number of deaths vs actual deaths for these countries using deep networks.
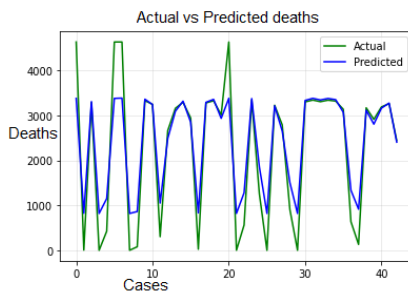

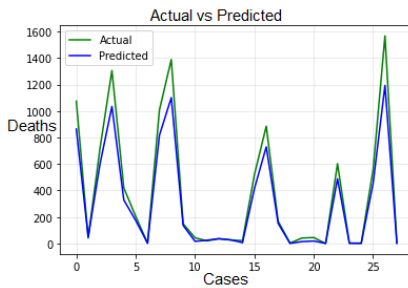Fig. 11. Actual vs Predicted deaths in China using DL

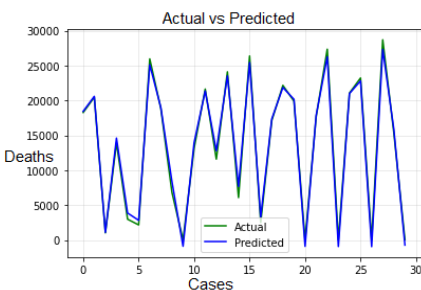Fig. 12. Actual vs Predicted deaths in India using DL
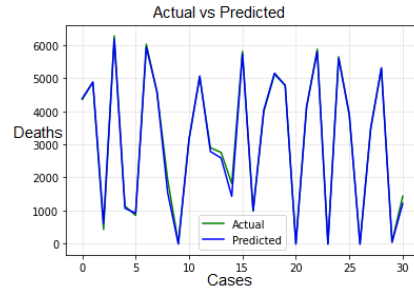

Fig.13. Actual vs Predicted deaths in Italy using DL

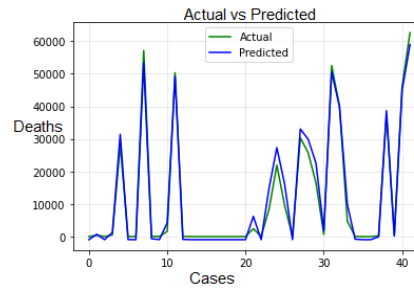Fig. 14. Actual vs Predicted deaths in Iran using DL

Fig. 15. Actual vs Predicted deaths in USA using DL

Similarly, the visualization of predicted recoveries vs actual recoveries using SVR is shown in Figures 16,17,18. Visualization of predicted recoveries vs actual recoveries using Deep learning is shown in Figures 19,20,21. These figures show the efficiency of SVR in prediction of recoveries corresponding to the number of cases in these four countries. It is observed that the performance of deep neural network is better than that of SVR thus we are considering Deep learning for the prediction of recoveries.
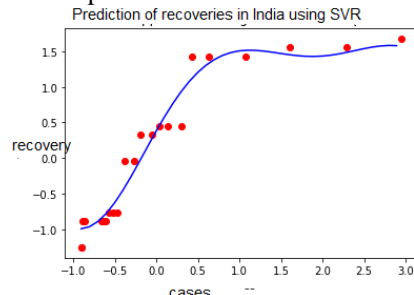

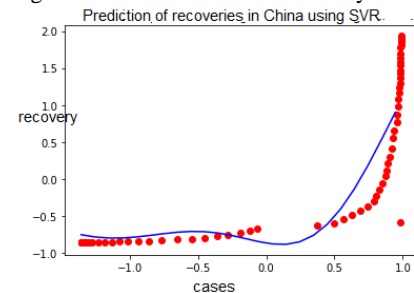Fig. 16. Actual vs Predicted recovery in India using SVR

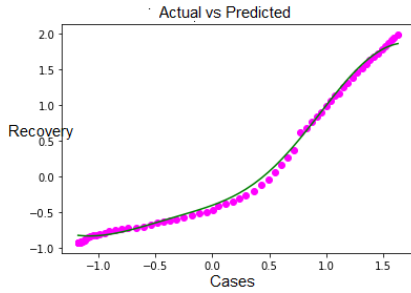Fig. 17. Actual vs Predicted recovery in China using SVR

374

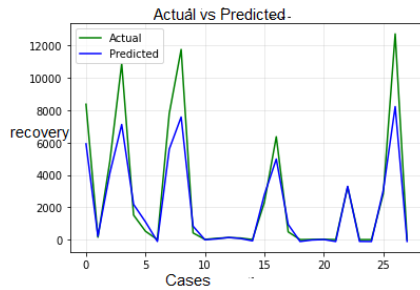Fig. 18. Actual vs Predicted recovery in Italy using SVR

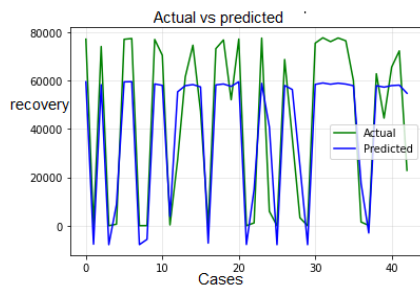Fig. 19. Actual vs Predicted recovery in India using DL

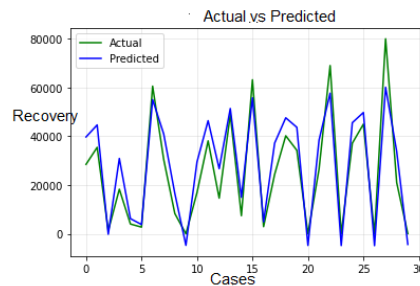Fig 20. Actual vs Predicted recovery in China using DL

Fig 19. Actual vs Predicted recovery in Italy using DL

V.   CONCLUSION AND FUTURE WORK

This paper reviews the challenges imposed on countries throughout the world due to the outbreak of COVID-19. It is the need of the hour to practice tough measures so that we can contain its spread. The paper analyzes data obtained from different countries and the use of machine learning models to estimate the rise in number of deaths, if preventive measures are not taken. The number of deaths and recoveries is predicted on the given dataset using Linear Regression, Support Vector Machine and deep networks consisting of multi-layer perceptrons.

Thus our trained models can predict the result of new and unforeseen data. It is observed that the efficiency of neural network in prediction of number of deaths and recoveries is more as compared to SVR and LR. If the precautions and measures are not followed, we might observe lakhs of cases within very less time as it happened in case of Italy. India needs to implement measures and devise a more efficient strategy for tracing, testing and treating every single infected patient while strengthening its health-care facilities, so as to minimize the impact of Coronavirus. In the future, we shall work on the prediction of number of cases which countries might see on daily basis, using other machine learning algorithms and time-series forecasting.

VI.   REFERENCE

[1] World Health Organization(2020). Coronavirus disease (COVID-19) Pandemic 2020. https://www.

who.int/emergencies/diseases/novel-coronavirus-2019. Online; accessed

1 May 2020.

[2] Huang C., Wang Y., Li, X., Ren, L., Zhao, J., Hu, Y.,Zhang, L., Fan, G., Xu, J., Gu, X. et al.(2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, The Lancet, 395(pp. 1497–1506).

[3] Cheng, Z. J. and Shan, J. (2020). 2019 novel coronavirus: where we are and what we know, Infection (pp. 1–9).

[4] Centers for Disease Control and Prevention(2020). Coronavirus symptoms and diagnosis 2020. https://www.cdc.gov/coronavirus/about/symptoms.html. Online; accessed

1 May 2020.

[5] Novel coronavirus (2019-nCoV) situation report-1 2020. https://www.who.org.

[6] Malik, Y. S., Sircar, S., Bhat, S., Sharun, K., Dhama, K., Dadar, M., Tiwari, R. and Chaicumpa, W. (2020). Emerging novel coronavirus (2019-ncov)-current scenario, evolutionary perspective based on genome

analysis and recent developments, Veterinary Quarterly (pp. 1-12).

[7] Lopez-Martin, C., Banitaan, S., Garcia-Floriano, A. and Yanez-Marquez, C. (2017). Support

vector regression for predicting the enhancement duration of software projects, 2017 16th

IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, (pp. 562–567).

[8] Kavitha, S., Varuna, S. and Ramya, R.(2016). A comparative analysis on linear regression and

support vector regression, 2016 Online International Conference on Green Engineering

and Technologies (IC-GET), IEEE(pp. 1–5).

[9] Browne, M.(1975). Predictive validity of a linear regression equation, British Journal of Mathematical

and Statistical Psychology 28(1),pp. (79–87).

[10]Choubin, B., Khalighi-Sigaroodi, S., Malekian, A. and Kis¸i, O¨ . 2016. Multiple linear regression,

multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting

precipitation based on large-scale climate signals, Hydrological Sciences Journal

61(6): 1001–1009.

[11] Singh, J. and Banerjee, R.(2019). A study on single and multi-layer perceptron neural network,

2019 3rd International Conference on Computing Methodologies and Communication (ICCMC),

IEEE, (pp. 35–40).

[12]   Colaboratory   2020. https://colab.research.google.com. Online; accessed 1 May

2020.

[13]Novel coronavirus (2019-nCoV) situation report-11 2020. https://www.

who.int/docs/default-source/coronaviruse/situation-reports/

20200131-sitrep-11-ncov.pdf?sfvrsn=de7c0f7_4. Online; accessed 3

June 2020.

[14] Van Der Hoek, L., Pyrc, K., Jebbink, M.F., Vermeulen-Oost, W., Berkhout, R.J., Wolthers, K.C., Wertheim-van Dillen, P.M., Kaandorp, J., Spaargaren, J. and Berkhout, B. (2004). Identification of a new human coronavirus, Nature medicine, 10(4), (pp.368-373).