



Time-Series Prediction for the Epidemic Trends of COVID-19 Using Conditional Generative Adversarial Networks Regression on Country-Wise Case Studies

Arnabi Bej¹ · Ujjwal Maulik¹ · Anasua Sarkar¹

Received: 19 December 2021 / Accepted: 20 May 2022 / Published online: 29 June 2022
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2022

Abstract

Probabilistic Regression is a statistical technique and a crucial problem in the machine learning domain which employs a set of machine learning methods to forecast a continuous target variable based on the value of one or multiple predictor variables. COVID-19 is a virulent virus that has brought the whole world to a standstill. The potential of the virus to cause inter human transmission makes the world a dangerous place. This article predicts the upcoming circumstances of the Corona virus to subside its action. We have performed Conditional GAN regression to anticipate the subsequent COVID-19 cases of five countries. The GAN variant CGAN is used to design the model and predict the COVID-19 cases for 3 months ahead with least error for the dataset provided. Each country is examined individually, due to their variation in population size, tradition, medical management and preventive measures. The analysis is based on confirmed data, as provided by the World Health Organization. This paper investigates how conditional Generative Adversarial Networks (GANs) can be used to accurately exhibit intricate conditional distributions. GANs have got spectacular achievement in producing convoluted high-dimensional data, but work done on their use for regression problems is minimal. This paper exhibits how conditional GANs can be employed in probabilistic regression. It is shown that conditional GANs can be used to evaluate a wide range of various distributions and be competitive with existing probabilistic regression models.

Keywords Conditional generative adversarial networks (CGAN) · COVID-19 · Time series · Forecasting · Deep learning

Introduction

Since genesis of 2020, COVID-19 has brought entire world to a standstill, infecting greater than 135 million civilians and 2.92M deaths. [1]

This paper gives forecasts for escalation of the COVID-19 virus for the widely infected countries like India, USA, Germany, Italy and Spain. Each country is individually inspected, because of their variations in population, habits, customs, health facilities, precautionary measures, etc. The analysis is dependent on actual data, as given by World Health Organization. The deployed deep learning model provides forecasts for new cases and death cases of the population for each country. COVID-19 has a high R0 value (denoting viral infectivity) [2, 3] of 3.25–3.4, represented by wide inter-human spread through the air, which means the flareup is uncontrollable. The main route of COVID-19 transmission is air-borne, thus curbing the source of infection, breaking off the infection route and prevention of endangered masses are the primary ways to discontinue the transmission of COVID-19. Because of this, many countries have taken measures like social distancing, night curfews, shutting down educational institutions, ceasing travel systems to prohibit more transmission of the epidemic [4][5]. Due to collected effort of authorities and the public, steps have given

This article is part of the topical collection “Machine Learning for Pandemic Prediction and Control” guest edited by Anand J Kulkarni, Akash Tayal, Patrick Siarry, Arun Solanki and Ali Husseinzadeh Kashan.

✉ Anasua Sarkar
anasua.sarkar@jadavpuruniversity.in

Arnabi Bej
arnabib057@gmail.com

Ujjwal Maulik
maulik@yahoo.com; ujjwal.maulik@jadavpuruniversity.in

¹ Department of Computer Science and Engineering, Jadavpur University, Kolkata, India

decreased spread in multiple countries. Therefore recently the epidemic is also decreasing to a large extent. Taking into account the seriousness of the pandemic, depicting when the epidemic will terminate, becomes extremely essential for the smooth running of systems of affected countries. Since COVID-19 cases are basically data on time series, so forecasting of future data is extremely needed. Machine Learning and Deep Learning techniques are really beneficial and commonly used in time series prediction. Figure 1 shows transmission and mitigation of COVID-19.

Related Works

From early centuries, the pandemics are prevalent before happening in year 2020. The pace of the entire globe has been stopped due to it. The SARS-COVID-19 virus is originated from Wuhan city, China in the year 2019 [6]. The first death is seen in the month of January 2020 that convinces that the virus spread through inter human transmission or very near contact with the person who is suffering from the disease [7]. The United States, Turkey, Russia, the UK and Germany—are the countries mostly affected in December 2020. 46 different countries also saw extreme cases in December 2020. Among them maximum deaths are seen by 31 countries. After fighting long, vaccines are finally available in most countries.

Arora et al. [8] show the prediction of COVID-19 using LSTM for all states of India with error of 3% using deep learning models. An ensemble of three models—ARIMA,

Prophet and Holt-Winters Exponential smoothing, has been designed by Abdulmajeed et al. [9] that takes realistic COVID-19 data from Nigeria Center of disease control for forecasting in a data constrained environment. Clément Massonnaud et al. [10] study the epidemic in France, and also evaluate the effects on healthcare resources like ICU beds for each area of France metropolitan. They produce a SEIR model based on attractive area of the hospitals. Cleo Anastassopoulou et al. [11] provide estimates of the main epidemiological parameters and simulation results in “effective” SIRD model.

Current Machine learning research is based on deploying models that produces distributions similar to underlying data. Most generative models presume a particular variant. CGAN [13] have exhibited a higher capability to produce authentic samples. Kernel methods in deep neural networks form the non-linear function. Intricate noise forms such as heteroscedastic noise have to be produced using specialized models [12]. Nowadays, research is with image data with a huge organized output space. The categories of various issues like regression are given minimal attention.

This paper gives an efficient distribution using GAN model to predict COVID-19 infection. Conditional Generative adversarial networks [13] regression is used here. Using rolling update method within CGAN, our approach shows an effective prediction for COVID-19 pandemic data. This implementation shows a new pathway to apply deep learning regression methods with rolling update for pandemic forecasting.



Fig. 1 Transmission and mitigation of COVID-19

Methodology

GAN is a model which predicts actual data and has two neural networks, generator and discriminator. Generator produces fake samples. The discriminator has to distinguish whether the input it is getting is G versus real data set. The discriminator will output a logic 1 if it finds that its input is more realistic. [14]. GANs are continuously evolving domain that provides generative models, which produces real like data across different areas of problems for example in image translation tasks for translating pictures of young face to aging face, etc. The generator produces a group of samples. These samples along with real-like data are given to the discriminator and categorized as fake or real. The discriminator thus gets better at distinguishing original and duplicate samples in the subsequent round. After that, the generator will get updated depending upon how perfectly the generator model fooled the discriminator.

Conditional GAN (CGAN) [13] differs from original GAN in the way that it puts an extra condition to manage the

production method of the data by concatenation of a vector y to both the generator as well as the discriminator. The latent vector z and vector y are inputs to generator. The samples and the information are fed to the discriminator. Figures 2 and 3 show GAN and CGAN architectures, respectively. The CGAN objective function is shown in Eq. 1. Here D denotes discriminative model while G denotes generative model for the conditional model with y . $p_z(z)$ denotes prior input noise.

$$\min_G \max_D V(D, G) = E_{x \sim p_d(x)} [\log D(x|y)] + E_{z \sim p_z(z)} [\log(1 - D(G(z|y)))] \tag{1}$$

In probabilistic regression, the relation between feature variable and target variable is given as conditional probability distribution. Therefore in this work, Conditional GAN method has been used to successfully approximate complex conditional multimodal distributions. In this case, y is very high-dimensional considering x with low dimension. With the help of heteroskedastic noise distributions, the proposed

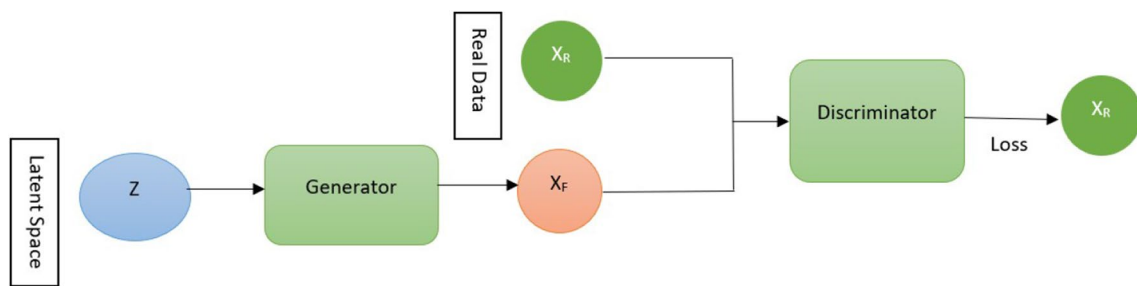


Fig. 2 GAN architecture

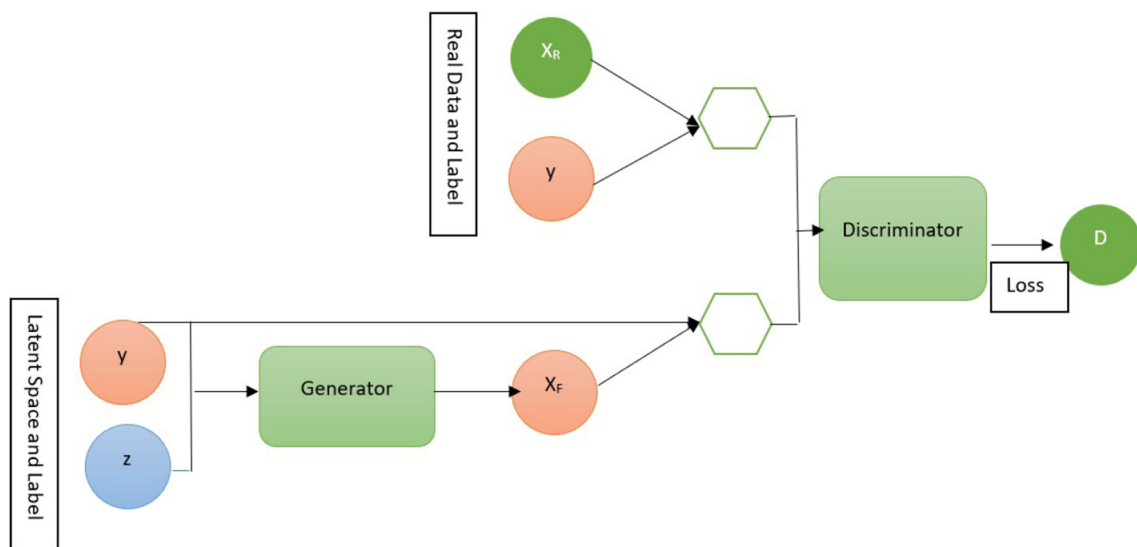


Fig. 3 CGAN architecture

CGAN regressor model estimates a conditional distribution in high-dimensional space. As dimension of y for this problem, cannot be well represented by limited, parametric families of distributions, the usage of implicit probability density function in CGAN method becomes able to predict the pandemic data and can quantify the uncertainty within data. The proposed model targets to learn a generator G , in which marginal generator distribution $p_g(y)$ is close to the marginal distribution of the data $p_d(y)$. Although MSE is used generally as loss function, it is not added in training process of GAN here.

Experimental Setup and Implementation

In this paper, we employ the CGAN [13] where conditioning is performed on the generator and discriminator at the time of training by adding some extra information. This information could be anything for example, class labels. Wasserstein GAN [15], can be used also for minimizing loss function [12]. They optimize another loss function, that could affect the performance. We report Root Mean Square Error (RMSE) metric with the experiments. It is one of the important metrics, as inspecting importance of the guess of unpredictability is of huge concern.

RMSE

Given a test dataset, $(x_i, y_i) i=1..N$, to find the Root Mean Squared Error, we first need to evaluate the residuals. Residuals are the subtracted value between the actual values and the predicted values. It is denoted by $(y_i - \hat{y}_i)$ where y_i is the observed value for the i th observation and \hat{y}_i is the predicted value. RMSE can have any values—both positive as well as nonpositive as the depicted value remains lesser or higher rates than the actual value. After performing the residuals square, the squares are averaged, and the square root is taken giving us the RMSE. After that the RMSE is deployed as a quantity of the spread of the y values around the depicted y value. This loss function is not any part of the CGAN training process.

Dataset

We have used the real-world WHO—COVID-19 dataset taken from <https://covid19.who.int/table>. The dataset consists of new and cumulative cases, as well as new and cumulative deaths, which are to be forecasted. The data used for training the model are from 1st January 2020 to 6th April

Table 1 Choosing data used for experiment

| Country | Confirmed cases | Death cases |
|---------|---------------------|---------------------|
| India | 1.1.2020–06.04.2021 | 1.1.2020–06.04.2021 |
| Germany | 1.1.2020–06.04.2021 | 1.1.2020–06.04.2021 |
| USA | 1.1.2020–06.04.2021 | 1.1.2020–06.04.2021 |
| Italy | 1.1.2020–06.04.2021 | 1.1.2020–06.04.2021 |
| Spain | 1.1.2020–06.04.2021 | 1.1.2020–06.04.2021 |

2021. We have tested the model from 7th April 2021 to 31st July 2021 (Table 1).

Preprocessing

On real-world COVID-19 data, training set was preprocessed using minmax scaler preprocessing.

Dataset Statistics

After dividing our dataset to training, test groups, we find 80% of dataset to be training and 20% of the dataset as test data.

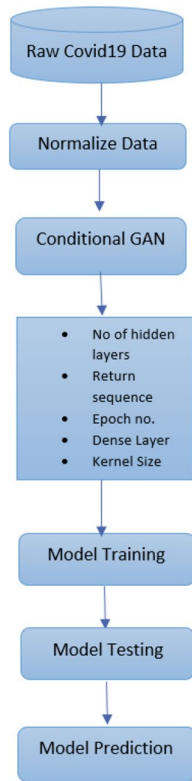
Model Architecture

We have used DNN for CGAN. A six layered network is used for the generator part. The three-layered multi-layer Perceptron gets as input x and noise z . On the other hand, the output illustrations are chained to a three-layered network. Linear activation is used for last layer while Rectified linear unit (RELU) [16] is activation function for remaining layers. RELU performed much better activations such as Leaky-RELU or selu. A four-layered network is used for discriminator. The single non-linear layer takes as input x and y . After that, a three layered multi layered perceptron that employs sigmoid activation function in the final layer takes the concatenated output. The optimizer used is Adam and learning rates used are of values 10^{-2} , 10^{-3} , 10^{-4} . The number of epochs used was 500. The discriminator and generator training steps ratio was set to 1. Our model used a batch size of 100. Noise z dimension was taken as 1. The input dimension was 3. Test was performed on a computer having 16GB RAM. CGAN captures very intricate conditional distributions, while different regression methods require various changes.

Heteroscedastic Noise

Covid19 real-world dataset shows more complex noise. This dataset employs heteroscedastic noise. CGAN generates more realistic samples and captures the

Fig. 4 Proposed methodology framework



het-eroscedastic noise structure very well. CGAN can learn to model any type of noise very easily.

Increasing the Dimensionality of Noise

Employing a larger dimensional noise can assist get enhanced behaviour for problems with complex noise. The proposed framework of the model is shown in Fig. 4. The generator network structure and discriminator network structure are shown in Figs. 5 and 6 respectively.

Mechanism of Rolling Update

In COVID-19 cases for achieving extended forecasts, a rolling update process is employed. It is used to upgrade the training example sequence depending on the present forecast values instantaneously for the purpose of training the model. It is actually done to enhance the rolling control of feedback evaluation in a bounded amount of time.

There are different parameters in the model. The input time step is taken as 3, which implies that new cases count of the next day is dependent on the cases of the last 3 days. The count of hidden neurons is 40. This is the input which is fed to hidden layer. The model used Adam optimizer, training of 500 epochs.

Fig. 5 Generator network structure

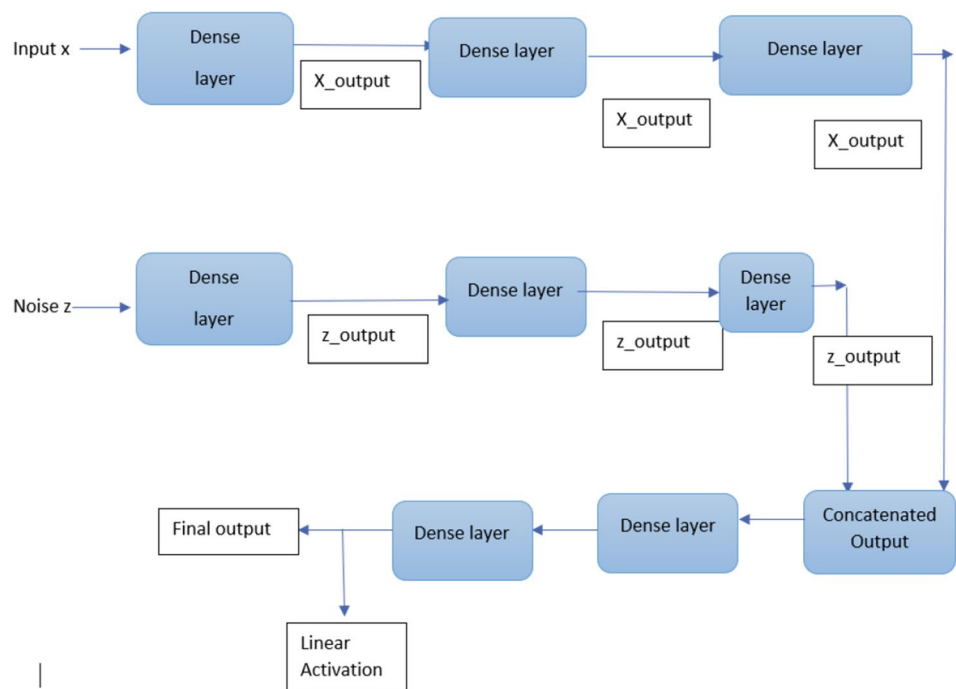


Fig. 6 Discriminator network structure

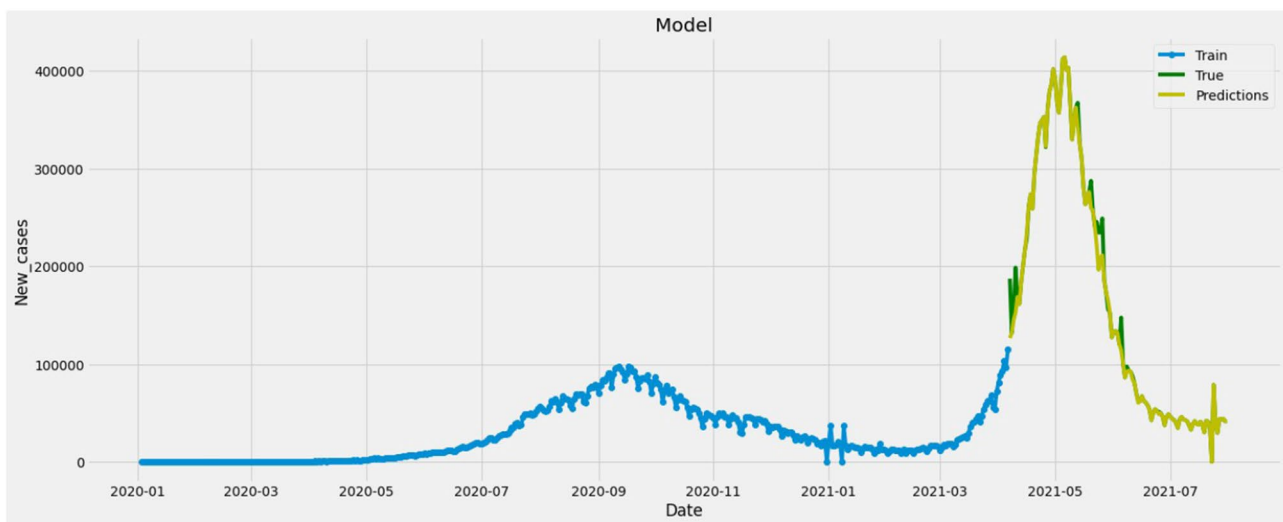
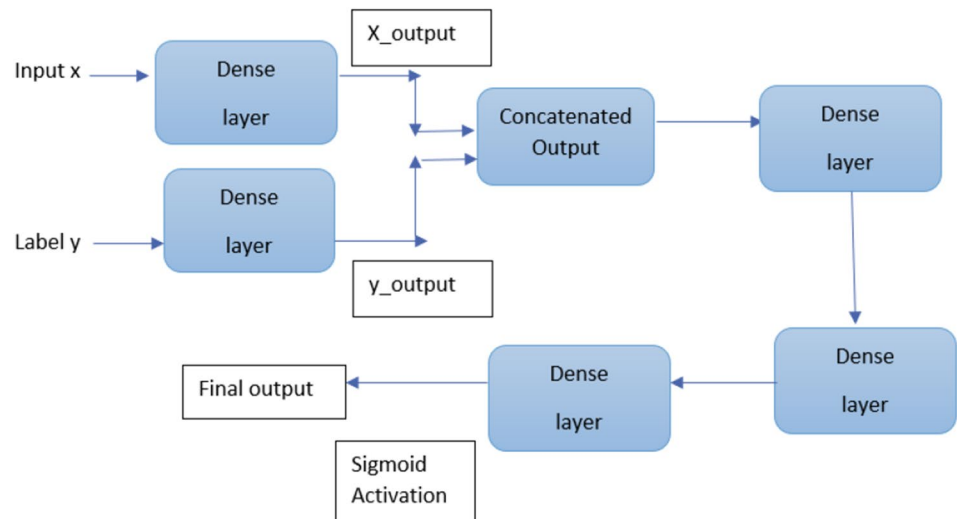


Fig. 7 India new cases

Results and Discussions

Results of COVID cases and deaths of five countries India, Italy, Spain, USA, Germany are visualised graphically and RMSE classification metric calculation. COVID-19 new cases and death cases predicted values of the five countries are visualised from Figs. 7, 8, 9, 10, 11, 12, 13, 14, 15, 16. The blue curve represents the training portion. The green curve represents true data from 7th April 2021 to 31st July, 2021. And the yellow curve shows the predicted values.

Covid-19 Forecasting of Daily New Cases

The COVID-19 true and predicted data of the countries India, USA, Spain, Italy, Germany are visualized graphically in the Figs. 8, 9, 10, 11, 12. The blue coloured curve depicts actual values and red-coloured curve depicts forecasted values. The predicted value and true values can be observed as overlapping with each other after April, 2021.

In Fig. 8, it can be seen that both curves are almost overlapping except sudden spikes in true value curve for weekend counts. In Fig. 9, most waves in the predicted curve are similar to the true values. In Fig. 10, only some spikes show dissimilarity when comparing new cases of Germany with our predicted values. In Fig. 11 also very little dissimilarity can be

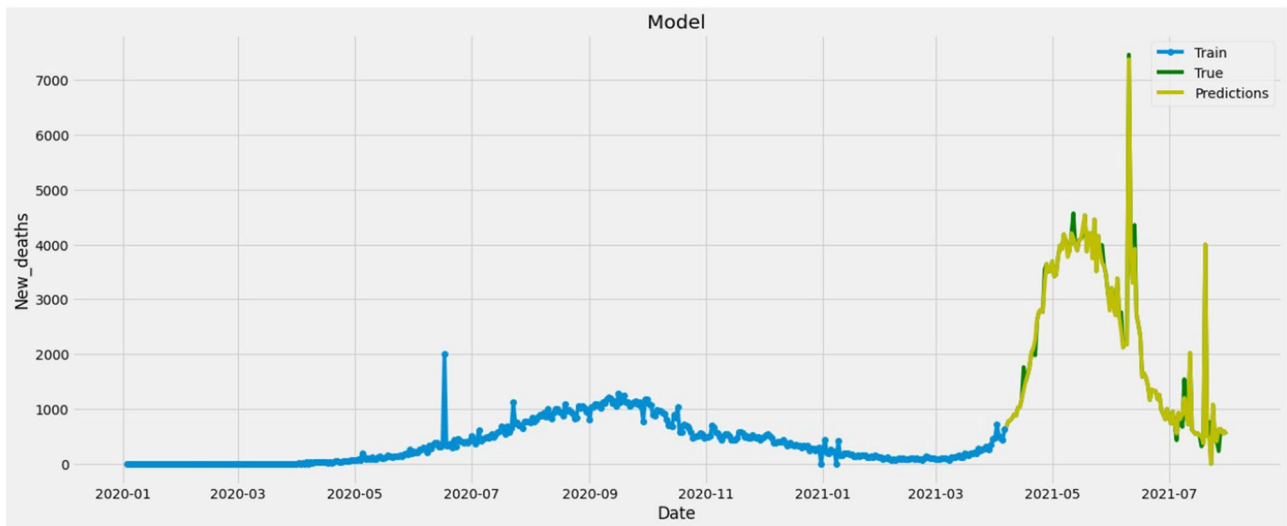


Fig. 8 India new deaths

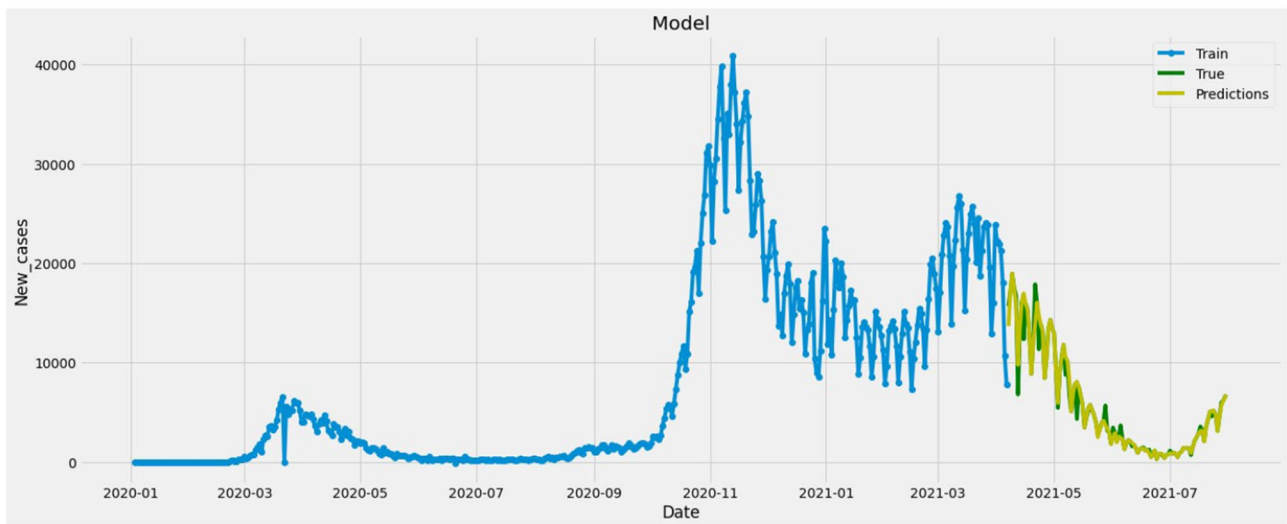


Fig. 9 Italy new cases

observed for new cases prediction in USA. In Fig. 12, similar results can be observed for Spain.

Covid-19 Forecasting of Daily New Death Cases

The COVID-19 true and predicted data of the countries India, USA, Spain, Italy, Germany are visualized graphically in the Figs. 7, 13, 14, 15, 16. The blue coloured curve depicts actual values and red-coloured curve depicts forecasted values.

The analysis of the graphs and classification metric implies that graphs and CGAN results are very close to the real cases of the five countries. The alpha value or significance level is considered to be 0.05 or 5%. When P value is smaller than alpha value, the results are statistically significant.

Alpha value is actually a threshold p value, that is actually decided by a team who conducted the test. It is decided before performing significance test (Z Test). To find the reported and forecasted value difference, Root Mean Square Error is to be found, by the equation:

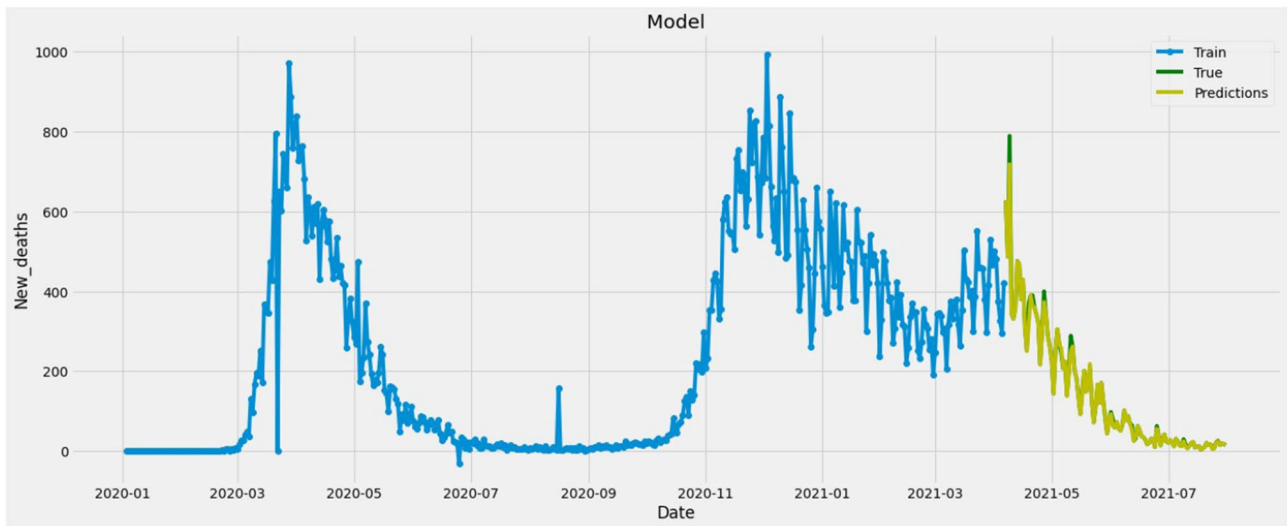


Fig. 10 Italy new deaths

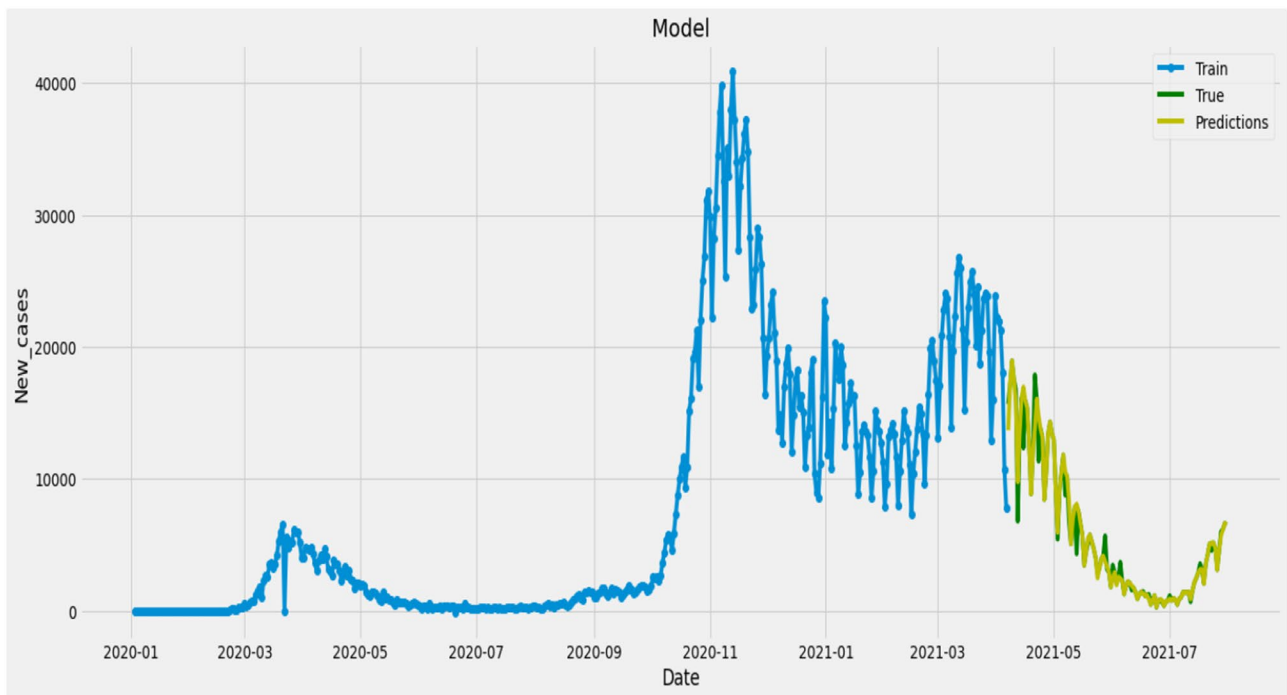


Fig. 11 Germany new cases

$$RMSE = \sqrt{\frac{\sum_1^n (y_i - y_j)^2}{n}} \tag{2}$$

Root Mean Squared Error values and *P* values are shown in Tables 2 and 3.

Mean Absolute Percentage Error (MAPE)

It evaluates how good a prediction model is. It calculates this accuracy in terms of percentage.

MAPE output is non-negative floating point. The best value is 0.0. But note the fact that bad predictions can lead to arbitrarily large MAPE values, especially if some y_{true} values are very close to zero. Note that we return a large

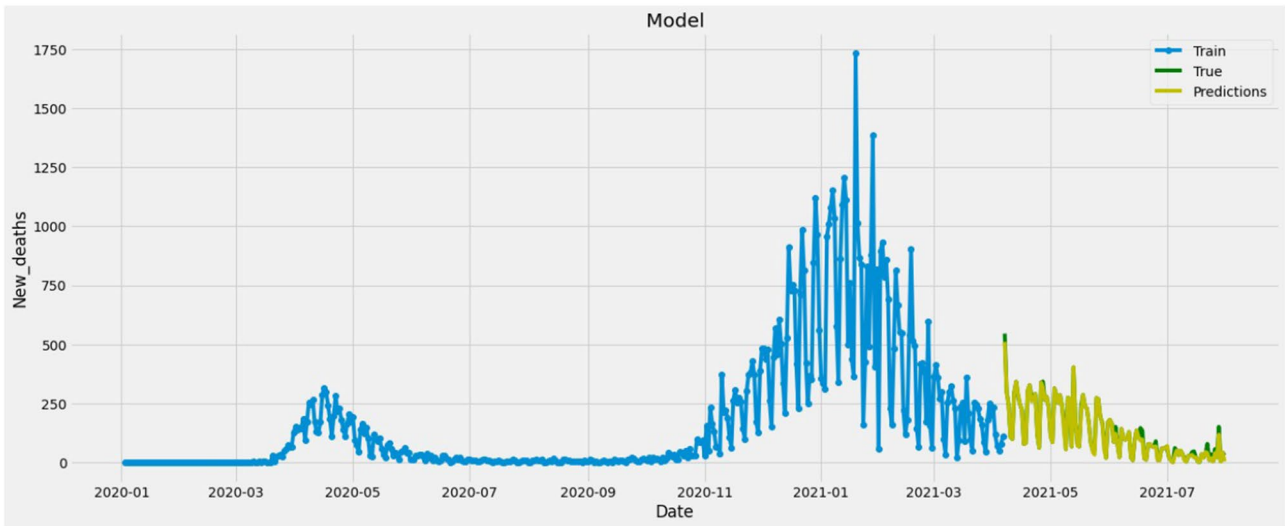


Fig. 12 Germany new deaths

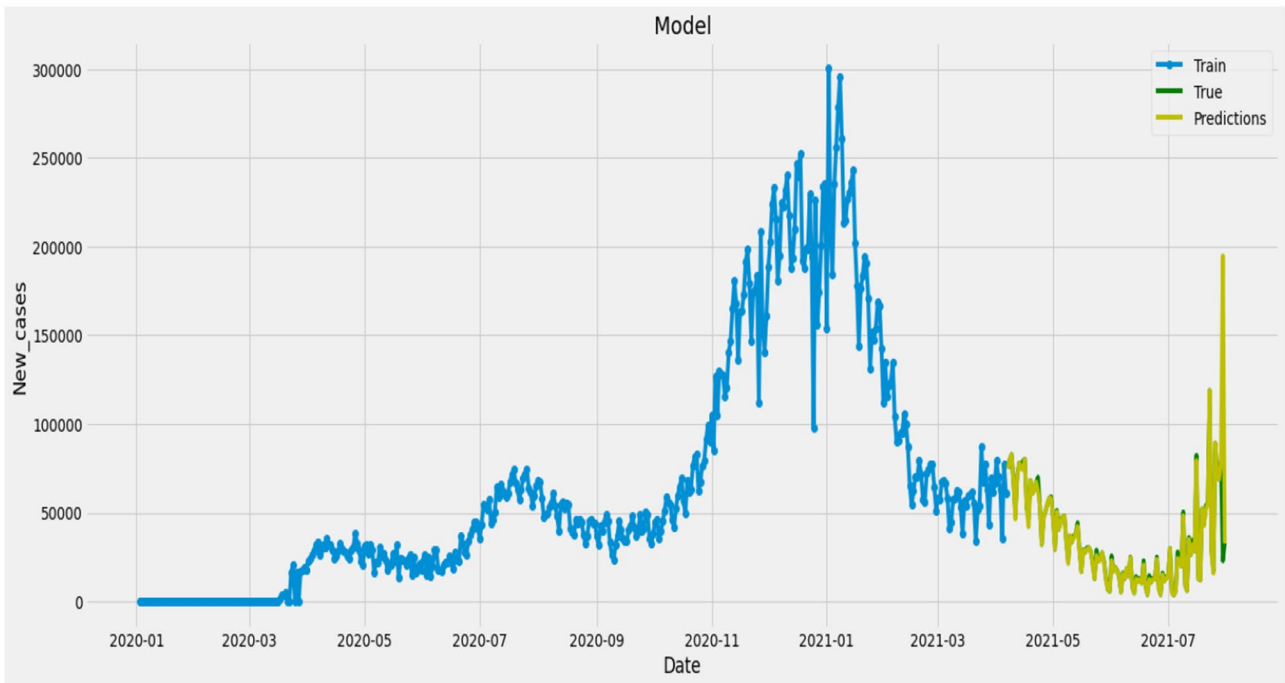


Fig. 13 USA new cases

value instead of inf when y_{true} is zero. MAPE is calculated by the equation

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \bar{y}}{y_i} \right| \tag{3}$$

Mean absolute percentage error (MAPE) is shown in Table 4 It gives minimum errors in our predicted data. It

shows maximum error for India with 2.38 value and minimum error for Germany with 0.54 value.

R² Score

R-squared (R^2) or coefficient of determination is a statistical measure that represents the proportion of the variance for a dependent variable that is explained by an

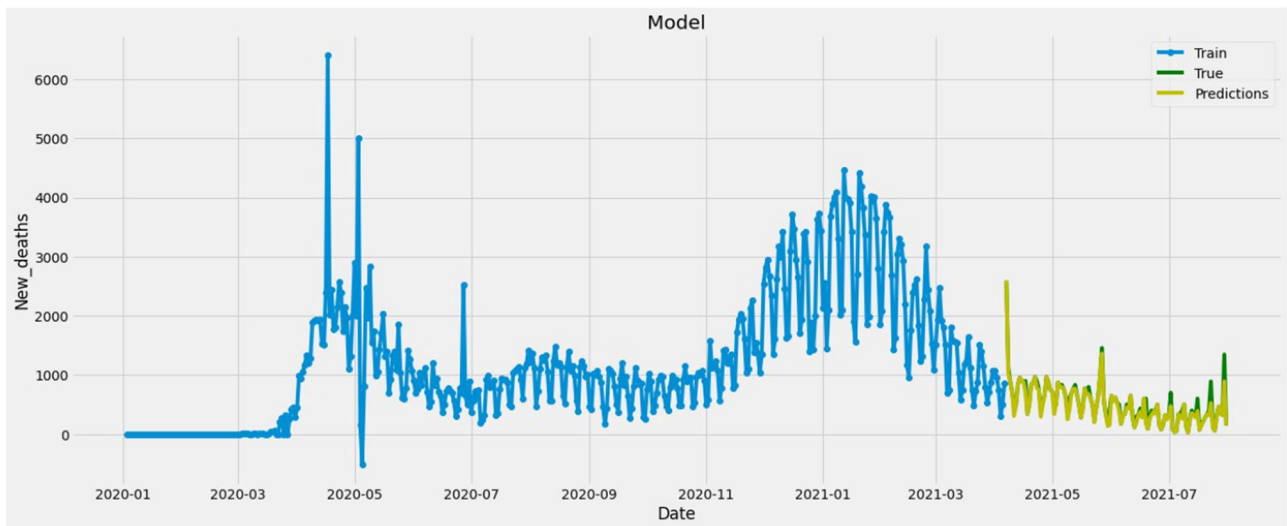


Fig. 14 USA new deaths

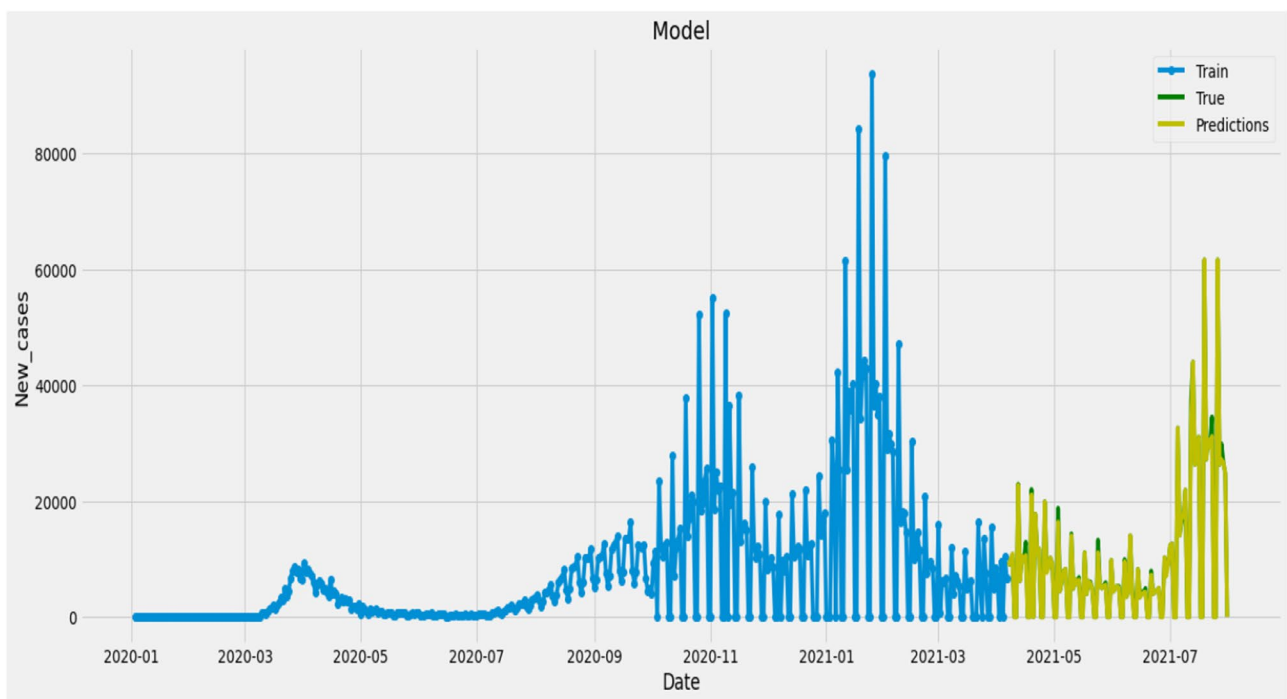


Fig. 15 Spain new cases

independent variable or variables in a regression model. It is also called the coefficient of determination.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \tag{4}$$

Coefficient of determination is shown in Table 5. All the values in Table 5 are within 0.75–0.98 range, which shows significance of our predicted values.

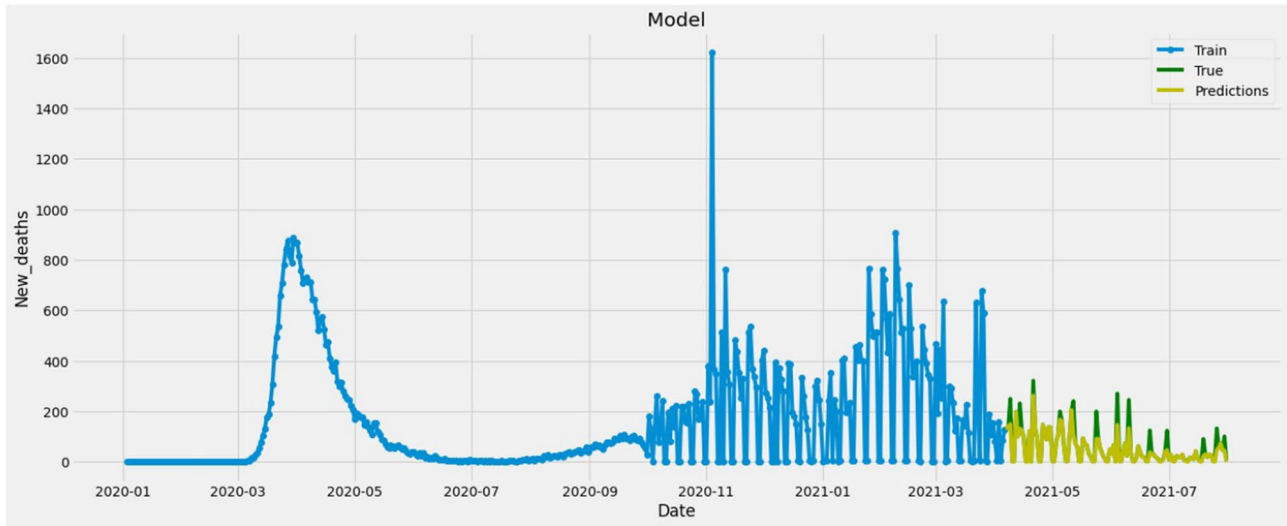


Fig. 16 Spain new deaths

Table 2 RMSE values

| Country | New cases | Cumulative cases | New deaths | Cumulative deaths |
|---------|-----------|------------------|------------|-------------------|
| India | 557.4 | 448.49 | 163.35 | 503.44 |
| Germany | 665.65 | 231.3 | 412.6 | 242.06 |
| USA | 709.35 | 542.8 | 587.54 | 142.43 |
| Italy | 734.7 | 681.7 | 308.5 | 221.4 |
| Spain | 603.56 | 126.3 | 253.8 | 203.4 |

Table 3 P values

| Country | New cases | Cumulative cases | New deaths | Cumulative deaths |
|---------|---------------------|---------------------|----------------------|---------------------|
| India | 0.02 | 0.007 | 0.008e ⁻² | 2.15e ⁻¹ |
| Germany | 0.003 | 2.43e ⁻¹ | 0.0016 | 3.3e ⁻² |
| USA | 3.6e ⁻² | 2.6e ⁻¹ | 2.87e ⁻¹ | 3.05e ⁻² |
| Italy | 2.18e ⁻² | 1.28e ⁻¹ | 3.41e ⁻¹ | 2.11e ⁻² |
| Spain | 0.014 | 2.6e ⁻² | 3.2e ⁻¹ | 2.5e ⁻² |

Table 5 R² score, the coefficient of determination

| Country | New cases | Cumulative cases | New deaths | Cumulative deaths |
|---------|-----------|------------------|------------|-------------------|
| India | 0.87 | 0.77 | 0.89 | 0.84 |
| Germany | 0.768 | 0.8 | 0.9 | 0.96 |
| USA | 0.75 | 0.86 | 0.78 | 0.98 |
| Italy | 0.83 | 0.75 | 0.93 | 0.88 |
| Spain | 0.89 | 0.94 | 0.84 | 0.95 |

Table 4 Mean absolute percentage error

| Country | New cases | Cumulative cases | New deaths | Cumulative deaths |
|---------|-----------|------------------|------------|-------------------|
| India | 2.38 | 2.49 | 1.26 | 0.66 |
| Germany | 0.54 | 2.40 | 0.84 | 3.42 |
| USA | 0.84 | 2.64 | 0.40 | 3.52 |
| Italy | 0.59 | 2.18 | 3.33 | 2.38 |
| Spain | 1.94 | 3.87 | 3.04 | 1.91 |

Table 6 Explained variance score

| Country | New cases | Cumulative cases | New deaths | Cumulative deaths |
|---------|-----------|------------------|------------|-------------------|
| India | 0.54 | 0.60 | 0.61 | 0.62 |
| Germany | 0.68 | 0.84 | 0.56 | 0.64 |
| USA | 0.51 | 0.90 | 0.76 | 0.50 |
| Italy | 0.73 | 0.51 | 0.633 | 0.60 |
| Spain | 0.63 | 0.85 | 0.71 | 0.85 |

Table 7 Performance measures from existing regression algorithms for New Cases in India

| Regressor | Expected variance | R^2 | MSE | RMSE |
|---------------------|-------------------|---------|------------|---------|
| Logistic regression | – | – | 41,690.592 | 204.182 |
| Lasso | 0.0 | – 0.255 | 0.124 | 0.352 |
| Ridge regression | 0.977 | 0.976 | 0.002 | 0.048 |
| ElasticNet | 0.193 | – 0.009 | 0.099 | 0.316 |
| Proposed | 0.950 | 0.943 | 0.006 | 0.075 |

Table 8 True and depicted daily new cases comparison

| Country | Date | Reported data | Predicted data |
|---------|-----------|---------------|----------------|
| India | 06.6.2021 | 100,636 | 100,321 |
| | 07.6.2021 | 86,498 | 86,212 |
| | 08.6.2021 | 92,596 | 92,321 |
| | 09.6.2021 | 93,463 | 93,112 |
| | 10.6.2021 | 92,291 | 92,345 |
| Germany | 06.6.2021 | 1964 | 1912 |
| | 07.6.2021 | 1444 | 1421 |
| | 08.6.2021 | 2253 | 2289 |
| | 09.6.2021 | 3275 | 3264 |
| | 10.6.2021 | 2747 | 2732 |
| USA | 06.6.2021 | 5395 | 5356 |
| | 07.6.2021 | 15,496 | 15,400 |
| | 08.6.2021 | 13,013 | 13,000 |
| | 09.6.2021 | 18,647 | 18,670 |
| | 10.6.2021 | 14,545 | 14,500 |
| Italy | 06.6.2021 | 2275 | 2289 |
| | 07.6.2021 | 1270 | 1265 |
| | 08.6.2021 | 1894 | 1800 |
| | 09.6.2021 | 2198 | 2450 |
| | 10.6.2021 | 2078 | 2043 |
| Spain | 06.6.2021 | 3 | 10 |
| | 07.6.2021 | 9542 | 9456 |
| | 08.6.2021 | 3504 | 3605 |
| | 09.6.2021 | 4427 | 4489 |
| | 10.6.2021 | 14,004 | 14,390 |

Explained Variance Score

It is used to measure the discrepancy between a model and actual data. In other words, it is the part of the model's total variance that is explained by factors that are actually present and is not due to error variance. The great score is 1.0 and lesser values are bad (Table 6).

Explained variance is shown in Fig. 6.

Table 7 shows comparative results for predictions of proposed method with Logistic Regression, Lasso, Ridge

Regression and ElasticNet methods for daily new cases data of India. From the obtained results, it is evident that only Ridge Regression is providing comparative better results than our proposed method. Lasso and ElasticNet methods are showing comparative results also. But Logistic Regression is failing to provide proper R^2 and Variance scores. So from Table 7, it is evident that our proposed method is providing comparative results for this COVID-19 prediction problem. We have applied scaling on data to make it compatible along all regression algorithms. That is why it is providing comparative results with other algorithms.

Table 8 shows predicted and reported data of the new COVID cases of India, Italy, Germany, Spain, USA from 6th June 2021 to 10th June, 2021. The values show that reported and predicted data are very close to each other.

Discussions and Suggestions

Development of Vaccine and Its Side Effects

Presently, there are 115 candidate vaccines for [17] COVID-19. In total, there are 33 candidate vaccines in stage 3 clinical trials. So far, 15 vaccines have been authorized across several countries. After a thorough evaluation of phase 3 clinical trial data major side effects after getting the Sputnik V vaccine were flu-like illness, headache, fatigue, and reactions at the injection site. It is based on surveys from 12,296 participants who had received two doses of either the vaccine or the placebo. There is quite a lot of turmoil regarding the vaccine, with some scientists unsure of the data and remaining are scared of side effects that the shot employs [18]. However, 64 countries have been using the vaccine, in addition to Russia.

Transmission of COVID-19 Through Air

COVID-19 is a disease that spreads through air and is a danger to the world, according to various researchers. Aerosol transmission is another mode of transmission of COVID-19. Not going outside house, frequent cleaning of hands and maintaining social distance are compulsory measures for avoiding airborne transferral of COVID-19 according to WHO with various other international authorities. Wearing face masks is extremely essential to stop inhaling aerosol. Social gathering must be avoided. Aerosol transmission is not universally accepted till now but some researchers found evidences of it [19]. So, precautionary measures must be taken to be prevented from the dangerous effects of the pandemic.

Conclusion

COVID-19 has [20] brought the world to a standstill due to its rampant spread. World has to cope with the pandemic as it is not known how long the pandemic will be there. We used deep learning model to predict new cases and new death cases for India, USA, Spain, Italy and Germany of COVID-19 in this paper. Collecting the data of COVID-19 really hard because of its limited availability for forecasting time series data. Conditional Generative adversarial network is used to model the predictions. The Root mean squared error is evaluated here. The predictions of this paper's CGAN model which provides current situation's results is close to actual COVID-19 data. This paper is the first evaluation of India, USA, Spain, Italy, Germany COVID-19 new cases and deaths. Our study will be essential for the five countries to take proper steps before being dominated by COVID-19 pandemic. We can inspect the losses incurred by the economies of the countries at the end of COVID-19 and take precautionary steps to overcome it, which can help the countries to come back into normal conditions. The COVID-19 new cases and death cases of several other countries can be predicted later on.

Declarations

Conflict of Interest The authors declare that they have no familiar monetary motives that could have portrayed to impact the task presented in this paper.

References

- Kathryn HJ. Will covid-19 generate global preparedness? *The Lancet*. 2020;395(10229):1013–4.
- Li Q, Guan X, Peng W, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau EHY, Wong JY, et al. Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New Engl J Med*. 2020. <https://doi.org/10.1056/NEJMoa2001316>.
- Shao N, Cheng J, Chen W. The reproductive number r_0 of covid-19 based on estimate of a statistical time delay dynamical system. *MedRxiv*. 2020.
- Kraemer MUG, Yang C-H, Gutierrez B, Chieh-Hsi W, Klein B, Pigott DM, Plessis LD, Faria NR, Li R, Hanage WP, et al. The effect of human mobility and control measures on the covid-19 epidemic in china. *Science*. 2020;368(6490):493–7.
- Ozaslan M, Muhamad Safdar I, Kilic H, Khailany RA. Practical measures to prevent covid-19: a mini-review. *J Biol Sci*. 2020;20:100–2.
- Huang C, Wang Y, Li X, Ren L, Zhao J, Yi Hu, Zhang Li, Fan G, Jiuyang Xu, Xiaoying Gu, et al. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The lancet*. 2020;395(10223):497–506.
- Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *The lancet*. 2020;395(10223):470–3.
- Arora P, Kumar H, Ketan Panigrahi B. Prediction and analysis of covid-19 positive cases using deep learning models: a descriptive case study of India. *Chaos Solit Fract*. 2020;139:110017.
- Abdulmajeed K, Adeleke M, Popoola L. Online forecasting of covid-19 cases in nigeria using limited data. *Data Brief*. 2020;30:105683.
- Massonnaud C, Roux J, Crépey P. Covid-19: forecasting short term hospital needs in france. *medrxiv*. 2020.
- Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-based analysis, modelling and forecasting of the covid-19 outbreak. *PLoS One*. 2020;15(3): e0230405.
- Le QV, Smola AJ, Canu S. Heteroscedastic gaussian process regression. In: *Proceedings of the 22nd international conference on Machine learning*, pages 489–496. 2005.
- Mirza M, Osindero S. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*. 2014.
- Mao X, Li Q, Xie H, Lau RYK, Wang Z, Paul Smolley S. Least squares generative adversarial networks. In: *Proceedings of the IEEE international conference on computer vision*, pages 2794–2802. 2017.
- Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. In: *International conference on machine learning*, pages 214–223. PMLR. 2017.
- Clevert D-A, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*. 2015.
- Forni G, Mantovani A. Covid-19 vaccines: where we stand and challenges ahead. *Cell Death Differ*. 2021;28(2):626–39.
- Male V. Are covid-19 vaccines safe in pregnancy? *Nat Rev Immunol*. 2021. <https://doi.org/10.1038/s41577-021-00525-y>.
- Morawska L, Milton DK. It is time to address airborne transmission of coronavirus disease 2019, (covid-19). *Clin Infect Dis*. 2020. <https://doi.org/10.1093/cid/ciaa939>.
- Schwamm LH, Erskine A, Licurse A. A digital embrace to blunt the curve of covid19 pandemic. *NPJ Digit Med*. 2020;3(1):1–3.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.