Visualization and Prediction of Trends of Covid-19 Pandemic During Early Outbreak in India Using DNN and SVR

Mahua Nandy Pal, Shuvankar Roy, Supriya Kundu, and Sasmita Subhadarsinee Choudhury

Abstract First known case of Covid-19 was found out in Wuhan, China in December, 2019. The virus itself is a novel virus, its harshness is unpredictable, its transmission ability is extremely powerful and its incubation period is comparatively larger. Covid-19 pandemic affected world health and socio-economy severely. So, it is required to know earlier whether the condition is continuing to get worse or how to scale up medical facilities like tracing, testing, treatment, quarantine etc. to fight against it. Early outbreak data for Novel Corona virus attack in India has been considered for this work. The trend of confirmed cases, recovery cases and deceased cases using deep neural network (DNN) and support vector regression (SVR) using Gaussian and exponential kernel functions are modeled. A comparative view of the prediction analysis is also considered.

Keywords DNN · Leaky ReLu · SVR · Exponential kernel · Gaussian kernel

1 Introduction

COVID-19 is essentially a severe acute respiratory syndrome virus with its origin in the Wuhan Province of China. At the end of 2019, the novel corona virus was identified as the origin of a number of pneumonia cases in Wuhan. Initially it was
considered as an epidemic throughout China but later a number of cases throughout
the world appeared. Thousands of people were suffering seriously from COVID-19,
and the curse began to grow exponentially in different places worldwide. In February
2020 World Health Organization (WHO) declared Corona Virus as a pandemic with
its spread in more than 168 countries in the world. The situation has been devastating
since then with increasing number of deaths in countries like Italy, Spain, USA
etc. with world class medical facilities. In India the number of infected cases has
crossed 3000 at the time of writing this article. The Virus do not have any proper
medicine or vaccination till date and in order to reduce the spread of COVID-19,
The Government of India had declared entire country under complete lockdown
excluding the essential services in order to combat this deadly Virus. The coronavirus
causes a range of symptoms such as pneumonia, fever, inhalation difficulty, and
lung contamination. These viruses are common in animals worldwide, but very few
cases had been known to affect humans before its outbreak in late 2019. WHO
renamed it as novel coronavirus that affected the lower respiratory tract of people
with pneumonia in Wuhan, China on 29th December of 2019. WHO mentioned the
disease as coronavirus disease (COVID-19). Infection preventive and control (IPC)
measures that may reduce the risk of exposure include the following: use of face
masks; covering coughs and sneezes with tissues that are then safely disposed of;
regular hand washing with soap or disinfection with hand sanitizer containing at least
60% alcohol (if soap and water are not available); avoidance of contact with infected
people and maintaining an appropriate distance as much as possible; and refraining
from touching eyes, nose, and mouth with unwashed hands.

Researchers are currently deficient of adequate data to predict the rapid growth
trend of this pandemic using machine learning tools. But predicting enables us
in understanding possible consequences which may affect enormously the socio-
economical growth. Visualizing and forecasting the pandemic behavior with AI
tools and data analysis enable us to prevent heavier health crisis and socio-economic
devastation. This may be considered as the motivation of this work.

As per the knowledge no works have been published yet which either represents
SVR analysis of Covid-19 data or DNN model trained on growth values of avail-
able data. The contribution of the work is that with minimal training samples quite
encouraging prediction results have been achieved. The DNN model we proposed
here is a simple model which takes less training time and is less prone to overfitting.

Following sections are arranged as follows. Section 2 represents literature survey,
Sect. 3 discusses deep neural network architecture and support vector regression,
Sect. 4 is dataset description, Sect. 5 is the representation of the early outbreak
scenario of India with respect to the World at that time. Sections 6 and 7 are the repre-
sentations and predictions using DNN and SVR models, Sect. 8 is error computation
and Sect. 9 is discussion about the work. Section 10 concludes the chapter.
2 Literature Survey

A meeting of WHO Emergency Committee under the International Health Regulations (IHR) (2005) [1] about the epidemic of novel coronavirus 2019 in China, took place on 30th January, 20. According to them, China rapidly recognized the virus and shared its sequence; hence, other countries could diagnose it to save themselves. According to the mathematical model proposed in the research paper in The Lancet [2], the growth of epidemic spreading rate will decrease down if the transmission rate of the communicable infections diminished to 0.25. A composite Monte-Carlo model (CMCM) is projected in [3] which supports future predictions using non-deterministic data distributions from a deterministic model. Some factors are non-deterministic and contribute to high uncertainty such as gathering of people and some factors are deterministic such as historical data. They claimed that these characteristics are characterized most effectively in probabilistic distribution as non-deterministic variables to the MC model. Second, the sensitivity values obtained from the MC simulation is utilized as remedial feedback to the rules that are produced from a fuzzy rule induction (FRI) system. According to [4], with limited training samples, finding a forecasting model is a huge challenge in the field of machine learning. For this, three generally used methods have been used in the past, (1) augmenting the existing data, (2) using a panel selection to pick the most efficient model from several models, and (3) fine-tuning the parameters of an individual model for maximum achievable accuracy. A methodology that holds these data mining strategies is proposed in [4]. Reference [5] suggests AI-driven tools to forecast the nature of spread of COVID-19 outbreaks. AI driven tools are likely to have active learning-based cross-population train/test models that utilize multitudinal and multi-modal data. The recent trend is to explore the utilization of deep learning architecture in different fields of research. References [6, 7] is a representation of application of deep convolutional neural network. Reference [8] is a reference which represents a well-established CNN architecture capable of extracting features in biomedical applications. In general, deep learning techniques require large amount of training data. They are extremely computationally exhaustive and the training phase is very time consuming with expensive GPUs. Covid-19 patient response to treatment based on Convolutional Neural Networks and Whale Optimization has been discussed in [9]. Detection of affected images from lung CT scan images is presented in [10].

3 Preliminaries

3.1 Deep Neural Network (DNN)

Artificial neural networks having multiple number of layers is called deep neural network. A fully connected (FC) neural network layer is said to be a dense layer. It accepts the input from each of the previous units and produces outputs for all the
| Layer     | (Type)         | Output shape | Param# |
|-----------|----------------|--------------|--------|
| Input_1   | (Input layer)  | [(None), 1]  | 0      |
| Dense_11  | (Dense)        | (None, 160)  | 320    |
| LRelu_11  | (Leaky ReLu)   | (None, 160)  | 0      |
| Dense_12  | (Dense)        | (None, 320)  | 51,520 |
| LRelu_12  | (Leaky ReLu)   | (None, 320)  | 0      |
| Dense_13  | (Dense)        | (None, 160)  | 51,360 |
| LRelu_13  | (Leaky ReLu)   | (None, 160)  | 0      |
| Dense_14  | (Dense)        | (None, 80)   | 12880  |
| LRelu_14  | (Leaky ReLu)   | (None, 80)   | 0      |
| Dense_15  | (Dense)        | (None, 1)    | 81     |
| Output    | (Leaky ReLu)   | (None, 1)    | 0      |

Total params: 116,161
Trainable params: 116,161
Non trainable params: 0

output units. Activation function helps in learning the patterns of characteristic of data. In this model Leaky ReLu has been used as nonlinear activation function which is capable of learning the nonlinear trend of data. It is defined as,

\[ f(x) = \max(\alpha x, x) \]  

Here \( \alpha \) is a hyper parameter set to 0.3.

In this paper, a DNN architecture is proposed, which consists of only four layers of dense and Leaky ReLu. In comparison with different well-established deep architecture, this model is quite lightweight with less number of trainable parameters and takes less time to be trained. As a result, the proposed architecture is less prone to over-fitting also. There are 116,161 trainable parameters as a whole. Table 1 represents the model of DNN where layer wise detailed technical architecture of the network is given. Input and output shape mapping is visible properly from Fig. 1.

### 3.2 Support Vector Regression (SVR)

Support vector machine (SVM) is a popular machine learning tool for classification and regression. Specifying kernel functions allow facilitating higher dimension transformation of input data. In this work, two kernel functions have been explored to implement support vector regressor (SVR). Kernel functions shift data representation to a transformed coordinate system in the higher dimension feature space.

Gaussian kernel is represented by,
Fig. 1 DNN model
\[ k(x, y) = \exp \left( -\frac{||x - y||^2}{2\sigma^2} \right) \]  

(2)

and exponential is represented by,

\[ k(x, y) = \exp \left( -\frac{||x - y||}{2\sigma^2} \right) \]  

(3)

The parameter \( \sigma \) should be optimized in an application-oriented way.

3.3 Dataset Description

The dataset [11] is contributed and updated on daily basis by Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE). This dataset is supported by ESRI Living Atlas Team and the Johns Hopkins University Applied Physics Lab (JHU APL). Johns Hopkins University is currently using a GitHub repository to store its data. It is in 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository.

4 Implementation Requirements

Python 3.6.8 is the implementation software requirement. Google’s deep learning library TensorFlow has been used to implement the work. Code has been executed in Google online Colab cloud environment (Tesla K80 GPU, 12 GB VRAM) with 4* Intel(R) Xeon(R) CPU @ 2.20 GHz, 13.51 GB RAM, 358.27 GB HDD.

5 Covid-19 Pandemic—India with Respect to World

Following are the representations of Covid-19 confirmed cases, deaths and recovery during early stage in India. The data visualization has been considered for the duration of 30th January, 2020 to 30th March, 2020. Figures 2, 3 and 4 are the data representation of India with respect to the world during the same duration and Figs. 5, 6 and 7 are the data representation of India with respect to the most affected countries at that stipulated duration.
Fig. 2 Confirmed case data visualization during an early outbreak stage of Covid-19 in India with respect to the then world data

Fig. 3 Recovery data visualization during an early outbreak stage of Covid-19 in India with respect to the then world data

6 Representation and Prediction with DNN Model

The period of early outbreak of covid-19 in India during 30th January to 30th March, 2020 has been considered. DNN was trained with consecutive growth values instead of using the time series data itself. The trained DNN predicts next 15 days confirmed cases, recovery cases and death cases data. Figures 8, 9 and 10 represent DNN training data plots whereas Figs. 11, 12 and 13 represent DNN prediction data plots. Prediction
Fig. 4 Death data visualization during an early outbreak stage of Covid-19 in India with respect to the then world data

Fig. 5 Confirmed case data visualization during an early outbreak stage of Covid-19 in India with respect to the then most affected countries of the world data

error representation has been done graphically by plotting absolute prediction error as well as by computing the average absolute prediction error percentage (Table 2).
7 Representation and Prediction with SVR Model

Work experimentation has been done using SVR with Gaussian and exponential kernels. Figures 14, 15 and 16 represent mapping of data up to 30th March, 2020. Figures 17, 18 and 19 are prediction representations of next 15 days respectively (Tables 3, 4 and 5).
Errors incurred by previously discussed three variations of prediction models have been compared. Absolute prediction error (APE) is defined as,

\[ APE = |Val_{pred} - Val_{obs}| \] (4)
These comparative error plots are helping in proper visualization of the day wise error trend. These three plots are shown in Fig. 20.

To ascertain the observation from day wise APE plots, average of absolute prediction error percentages (AAPEP) have been computed which is defined as,

\[
AAPEP = \frac{\sum_{i=1}^{n} \left( \frac{|\text{Val}_{\text{Pred}} - \text{Val}_{\text{Obs}}|}{\text{Val}_{\text{Obs}}} \right) \times 100}{n}
\]  (5)
where \( n \) is no. of actual prediction observation.

In Fig. 21, AAPEP values obtained from three models for three different types of data are presented.
### Table 2: Actual and predicted cases of Covid-19 for next 15 days by DNN

| Date       | Actual confirmed cases | DNN prediction-confirmed cases | Actual deceased cases | DNN prediction-deceased cases | Actual recovery cases | DNN prediction-recovery cases |
|------------|------------------------|--------------------------------|-----------------------|--------------------------------|-----------------------|--------------------------------|
| 31st Mar, 20 | 1397                   | 1403                           | 35                    | 108                            | 123                   | 62                             |
| 1st April, 20 | 1998                   | 1563                           | 58                    | 121                            | 148                   | 94                             |
| 2nd April, 20 | 2543                   | 1732                           | 72                    | 135                            | 191                   | 127                            |
| 3rd April, 20 | 2567                   | 1909                           | 72                    | 150                            | 192                   | 162                            |
| 4th April, 20 | 3082                   | 2095                           | 86                    | 165                            | 229                   | 200                            |
| 5th April, 20 | 3588                   | 2289                           | 99                    | 182                            | 229                   | 241                            |
| 6th April, 20 | 4778                   | 2491                           | 136                   | 198                            | 375                   | 284                            |
| 7th April, 20 | 5311                   | 2702                           | 150                   | 216                            | 421                   | 330                            |
| 8th April, 20 | 5916                   | 2921                           | 178                   | 234                            | 506                   | 378                            |
| 9th April, 20 | 6725                   | 3148                           | 226                   | 253                            | 620                   | 430                            |
| 10th April, 20 | 7598                   | 3383                           | 246                   | 273                            | 774                   | 484                            |
| 11th April, 20 | 8446                   | 3626                           | 288                   | 293                            | 969                   | 542                            |
| 12th April, 20 | 9205                   | 3877                           | 331                   | 314                            | 1080                  | 604                            |
| 13th April, 20 | 10453                  | 4137                           | 358                   | 336                            | 1181                  | 668                            |
| 14th April, 20 | 11487                  | 4405                           | 393                   | 359                            | 1359                  | 737                            |
9 Discussion

The recent Covid-19 pandemic behaviour and its dynamic nature with rapid and wide probability of change in trend are tried to visualize with respect to India in this work. It is very important to forecast as it is associated with threatening of life and social economy. Support vector regression using two different kernel functions and DNN model are used for representation of the growth of available data. Prediction of future data and prediction error is visualized, too, using those models. Average absolute
Fig. 16  SVR death case training data plot

Fig. 17  SVR confirmed case test data plot

prediction error in confirmed case, recovery case and death case prediction for DNN and two models of SVR are as shown in Table 6. DNN is most effective in predicting confirmed cases whereas, Gaussian kernel effectively represents forecasting of death and recovery cases. As per overall observation, we can say that, DNN is the most efficient option among three in forecasting confirmed cases and decreased cases but the model is comparatively less efficient compared to others while predicting recovery cases. Train data availability during the early outburst period in India was not sufficient for learning the trend exactly for a DNN. But it is tried to capture the
trend by executing training of DNN on growth data, which is least available in case of death data during the period considered. This is the reason why the model failed to learn the death trend properly and the error percentage is high in death case. This may be considered as the limitation of the work.

Fig. 18  SVR recovery test data plot

Fig. 19  SVR death test data plot
Table 3  Actual and predicted confirmed cases of Covid-19 by SVR using Gaussian and exponential kernel functions

| Date             | Actual confirmed case | Gaussian kernel prediction | Exponential kernel prediction |
|------------------|-----------------------|---------------------------|------------------------------|
| 31st March, 20   | 1397                  | 1529                      | 1454                         |
| 1st April, 20    | 1998                  | 1816                      | 1646                         |
| 2nd April, 20    | 2543                  | 2668                      | 1829                         |
| 3rd April, 20    | 2567                  | 2259                      | 2003                         |
| 4th April, 20    | 3082                  | 2389                      | 2169                         |
| 5th April, 20    | 3588                  | 2473                      | 2327                         |
| 6th April, 20    | 4778                  | 2305                      | 2478                         |
| 7th April, 20    | 5311                  | 2573                      | 2622                         |
| 8th April, 20    | 5916                  | 2610                      | 2759                         |
| 9th April, 20    | 6725                  | 2645                      | 2891                         |
| 10th April, 20   | 7598                  | 2680                      | 3017                         |
| 11th April, 20   | 8446                  | 2715                      | 3138                         |
| 12th April, 20   | 9205                  | 2750                      | 3254                         |
| 13th April, 20   | 10453                 | 2785                      | 3366                         |
| 14th April, 20   | 11487                 | 2820                      | 3473                         |

Table 4  Actual and predicted deceased cases of Covid-19 by SVR using Gaussian and exponential kernel functions

| Date             | Actual deceased cases | Gaussian kernel prediction | Exponential kernel prediction |
|------------------|-----------------------|---------------------------|------------------------------|
| 31st March, 20   | 35                    | 39                        | 37                           |
| 1st April, 20    | 58                    | 48                        | 42                           |
| 2nd April, 20    | 72                    | 59                        | 47                           |
| 3rd April, 20    | 72                    | 72                        | 52                           |
| 4th April, 20    | 86                    | 86                        | 57                           |
| 5th April, 20    | 99                    | 101                       | 62                           |
| 6th April, 20    | 136                   | 116                       | 67                           |
| 7th April, 20    | 150                   | 131                       | 72                           |
| 8th April, 20    | 178                   | 145                       | 77                           |
| 9th April, 20    | 226                   | 158                       | 82                           |
| 10th April, 20   | 246                   | 169                       | 86                           |
| 11th April, 20   | 288                   | 179                       | 90                           |
| 12th April, 20   | 331                   | 187                       | 94                           |
| 13th April, 20   | 358                   | 194                       | 98                           |
| 14th April, 20   | 393                   | 200                       | 102                          |
Table 5  Actual and predicted recovery cases of Covid-19 by SVR using Gaussian and exponential kernel functions

| Date         | Actual recovery cases | Gaussian kernel prediction | Exponential kernel prediction |
|--------------|-----------------------|---------------------------|-----------------------------|
| 31st March, 20 | 123                   | 107                       | 101                         |
| 1st April, 20  | 148                   | 120                       | 107                         |
| 2nd April, 20  | 191                   | 134                       | 113                         |
| 3rd April, 20  | 192                   | 148                       | 119                         |
| 4th April, 20  | 229                   | 162                       | 125                         |
| 5th April, 20  | 229                   | 176                       | 131                         |
| 6th April, 20  | 375                   | 190                       | 137                         |
| 7th April, 20  | 421                   | 203                       | 143                         |
| 8th April, 20  | 506                   | 216                       | 149                         |
| 9th April, 20  | 620                   | 228                       | 155                         |
| 10th April, 20 | 774                   | 239                       | 161                         |
| 11th April, 20 | 969                   | 249                       | 167                         |
| 12th April, 20 | 1080                  | 258                       | 173                         |
| 13th April, 20 | 1181                  | 266                       | 179                         |
| 14th April, 20 | 1359                  | 273                       | 185                         |

Fig. 20  Absolute prediction error (APE) plots for DNN, support vector regression with Gaussian and exponential kernels
Fig. 21 Average absolute prediction error percentage (AAPEP) plots for DNN, support vector regression with Gaussian and exponential kernels

|                      | Average absolute prediction error |
|----------------------|----------------------------------|
|                      | Date          | Confirmed cases | Recovery cases | Death cases |
| DNN                  | 42.7506       | 45.9664         | 99.3057        |
| SVR (Gaussian)       | 44.6309       | 7.1762          | 7.1762         |
| SVR (exponential)    | 44.7340       | 16.7569         | 16.7569        |

### 10 Conclusion and Future Work

Considering general population, at this moment there is no vaccination available for preventing COVID-19. The best prevention to subside this pandemic is to avoid being exposed to the virus. Forecasting its behaviour is necessary to analyze whether it is continuing or getting to be worse day by day or the chain of pandemic outbreak should soon be collapsed. It will be helpful for timely decision making and remedial action.

According to the above analysis DNN prediction performance minimizes error in confirmed cases predictions whereas SVR with Gaussian kernel exhibit minimal error in recovery and death cases prediction. DNN would have performed better if more training samples might be available.

The future scope of this work is to properly analyse the training set representation to the model so that the model learning becomes efficient with limited data available and can frame the trend prediction more powerfully.
References

1. World Health Organization (WHO): Statement on the Second Meeting of the International Health Regulations (2005) Emergency Committee Regarding the Outbreak of Novel Coronavirus (2019-nCoV). 30 January 2020. Geneva, Switzerland (2020)
2. Wu, J.T., Leung, K., Leung, G.M.: Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study, The Lancet (January 31)1–9 (2020). https://doi.org/10.1016/S0140-6736(20)30260-9
3. Fong, S.J., Li, G., Dey, N., Crespo, R.G., Viedma, E.H.: Composite Monte Carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction. Appl. Soft Comput. 106282 (2020)
4. Fong, S.J., Li, G., Dey, N., Crespo, R.G., Viedma, E.H.: Finding an accurate early forecasting model from small dataset: a case of 2019-nCoV novel coronavirus outbreak. Int. J. Interact. Multim. Artif. Intell. 6, 1 (2020)
5. Santosh, K.C.: AI-driven tools for coronavirus outbreak: need of active learning and cross-population train/test models on multitudinal/multimodal data. J. Med. Syst. 44, 93 (2020)
6. Fu, H., Xu, Y., Wong, D.W.K.: Retinal vessel segmentation via deep learning network and fully connected conditional random fields. In: 2016 IEEE Symposium on Biomedical Imaging
7. Liskowski, P., Krawice, K.: Segmenting retinal blood vessels with deep neural networks. IEEE Trans. Med. Imaging 35(11), 2369–2380 (2016)
8. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation (2015). arXiv:1505.04597 [cs.CV]
9. ELGhamrawy, S.M., Hassanien, A.E.: Diagnosis and prediction model for COVID19 patients response to treatment based on convolutional neural networks and whale optimization algorithm using CT image. medRxiv 2020.04.16.20063990 (2020). https://doi.org/10.1101/2020.04.16.20063990
10. Rajinikanth, V., Dey, N., Joseph Raj, A.N., Hassanien, A.E., Santosh, K.C., Sri Madhava Raja, N.: Harmony-search and Otsu based system for coronavirus disease (COVID-19) detection using lung CT scan images (2020). arXiv preprint arXiv:2004.03431
11. Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by Johns Hopkins CSSE. https://github.com/CSSEGISandData/COVID-19