Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
A Deep Learning algorithm for modeling and forecasting of COVID-19 in five worst affected states of India

Junaid Farooq, Muhammad Abid Bazaz

PII: S1110-0168(20)30490-7
DOI: https://doi.org/10.1016/j.aej.2020.09.037
Reference: AEJ 1845

To appear in: Alexandria Engineering Journal

Received Date: 2 July 2020
Accepted Date: 23 September 2020

Please cite this article as: J. Farooq, M.A. Bazaz, A Deep Learning algorithm for modeling and forecasting of COVID-19 in five worst affected states of India, Alexandria Engineering Journal (2020), doi: https://doi.org/10.1016/j.aej.2020.09.037

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Production and hosting by Elsevier B.V. on behalf of Faculty of Engineering, Alexandria University.
A Deep Learning algorithm for modeling and forecasting of COVID-19 in five worst affected states of India

Junaid Farooq*, Muhammad Abid Bazaz

*Department of Electrical Engineering, National Institute of Technology Srinagar, India

Abstract

In this paper, deep learning is employed to propose an Artificial Neural Network (ANN) based online incremental learning technique for developing an adaptive and non-intrusive analytical model of Covid-19 pandemic to analyze the temporal dynamics of the disease spread. The model is able to intelligently adapt to new ground realities in real-time eliminating the need to retrain the model from scratch every time a new data set is received from the continuously evolving training data. The model is validated with the historical data and a forecast of the disease spread for 30-days is given in the five most affected states of India.

Keywords: Covid-19, Incremental Learning, ANN, Forecasting

1. Introduction

Covid-19 is a highly contagious epidemic disease caused by novel coronavirus (SARS-CoV-2) that originated in Wuhan, Hubei Province of China in late December 2019. World Health Organization (WHO) declared Covid-19 as a pandemic on 12th March 2020 [1]. Researchers and policy makers are working round the clock to find solutions and design strategies to control the pandemic and minimize its impact on human health and economy.

The transmission of SARS-CoV-2 in humans is mostly through respiratory droplets (sneezing, coughing and while talking) and through contaminated surfaces [2]. The most significant property of SARS-CoV-2 is that it can persist on a variety of surfaces from hours to 9 days at room temperature which makes its transmission more rapid [3]. This virus can cause Acute Respiratory Distress Syndrome (ARDS) or multiple organ dysfunction, which may lead to physiological deterioration and death of an infected individual [4].

Mathematical modeling of infectious diseases and epidemics has been employed as an important tool for analysis of disease characteristics and investigation of disease spread ever since the ground breaking work of Kermack and McKendrick in 1972 [5]. It plays a useful role in efficient decision making and optimal policy framing. Different models have been developed to analyse the transmission dynamics of many infectious diseases like malaria (Ronald Ross model)[6], cholera (Capasso and Pareri-Fontana model, 1979) [7], gonorrhea (Hethcote and Yorke model, 1984) [7], Ebola [8], H1N1 [9] etc. Different mathematical models have been developed for Covid-19 disease as well [10] [11].

In this work, deep learning is employed to propose an Artificial Neural Network (ANN) based online incremental learning technique to estimate parameters of a data stream guided analytical model of Covid-19 in order to aid in optimal policy formulation, efficient decision making, forecasting and simulation. Modeling and simulation of such problems poses an additional challenge of continuously evolving training data in which the model parameters change over time depending upon external factors. The main contribution of this work is that in a scenario of continuously evolving training data, unlike typical deep learning techniques, this model eliminates the need to retrain or rebuild the model from scratch every time a new training data set is received.

Covid-19 official data from 5 worst hit states of India has been taken as the case study. The first case of COVID-19 in India was reported on 30th January 2020 originating from Wuhan, China [12]. As on 27 June 2020, the total number of cases reported in India is 508,953 with 295,881 recoveries and 15,685 deaths [13]. Hence the number of active cases is 197,387. An overview of the spread of COVID-19 in different states of India is shown in Figure 1. The growth of Covid-19 in India is plotted in Figure 2. It is evident from these figures that the disease has spread all over the country and the worst hit states are Maharashtra, Delhi, Tamil Nadu, Gujarat and Uttar Pradesh. India comprises of a total of 28 states and 8 union territories all of which are affected by the Covid-19 pandemic.

The government of India imposed a country wide complete lockdown on 24th March 2020 with strict restrictions on the movement of people while allowing only the essential services to operate under the supervision of administration and health
Covid-19 pandemic has emerged as a major challenge for the people, health workers and policy makers of the country.

The paper is organised as follows: Section 2 discusses an analytical epidemiological model for Covid-19 infectious disease. Section 3 deals with the adaptive deep learning technique of model parameter estimation and updation. In section 4, simulation results for 30-day ahead forecast of disease spread in the five worst affected states of India are presented.

2. Mathematical Modeling

2.1. Basic and Modified SIR Models

The pioneer work in development of mathematical models for infectious diseases was carried out by by [5] known as the susceptible-infectious-recovered (SIR) model. As one of the most classical models, it has been used by many researchers to study and analyse many infectious diseases like seasonal flu [16, 17], pandemic flu [18, 19], HIV/AIDS [20], SARS [21, 22] etc. These studies have shown that SIR models are reliable for analysis of the infectious disease spread and evaluation of the impact of prevention schemes in different scenarios. The basic SIR model is described by the following differential equations:

\[
\begin{align*}
\frac{dS}{dt} &= -\beta \frac{SI}{N} \\
\frac{dI}{dt} &= \beta \frac{SI}{N} - \gamma I \\
\frac{dR}{dt} &= \gamma I
\end{align*}
\]

where \( S, I \) and \( R \) are functions of time representing the number of susceptible, infected and recovered individuals in a population of size \( N \) at time \( t \). \( \beta \) is the rate of transmission and \( \gamma \) is the rate of recovery of infected individuals. It is assumed that those recovered develop immunity and do not catch the infection again in the time span of interest.

The basic SIR model can be modified in various ways to accommodate different scenarios. A modified SIR model known as SIRD (Susceptible-Infected-Recovered-Deceased) model is of our interest here and is based on the following assumptions:

(i) This model is fatal unlike a typical non-lethal SIR model which means that there is a positive probability of an infected person dying, \( P(\text{Death}) = \delta \) and \( \delta \geq 0 \).

(ii) A typical SIR model assumes that the Recovered group gains full immunity from reinfection. However, this model accommodates the possibility of a recovered person being reinfected with probability of reinfection, \( P(\text{Reinfection}) = \sigma \) and \( \sigma \geq 0 \).

(iii) The impact of new births and unrelated deaths in ignored and the total population remains constant, \( N = \text{constant} \forall t \).
The population is distributed randomly over the area. Therefore, there are four classes of individuals: Susceptible (S), Infected (I), Recovered (R) and Deceased (D) as described by the following equations and shown Figure 3:

\[
\frac{dS}{dt} = -\beta\frac{S \cdot I}{N} + \sigma R \tag{4}
\]
\[
\frac{dI}{dt} = \beta\frac{S \cdot I}{N} - \gamma I - \delta I \tag{5}
\]
\[
\frac{dR}{dt} = \gamma I - \sigma I \tag{6}
\]
\[
\frac{dD}{dt} = \delta I \tag{7}
\]

where:

\(\beta\) : Rate of Infection (Transmission of susceptible to infected)

\(\gamma\) : Rate of Recovery (Transmission of infected to recovered)

\(\delta\) : Rate of Death (Transmission of infected to deceased)

\(\sigma\) : Rate of Susceptibility (Transmission of recovered to susceptible)

It must be noted here that the process of mass action transmission is described by the non-linear term \(\beta SI/N\) where \(\beta\) = number of contacts per unit time by a person in group I required to transmit the disease to a person in group S, \(N - 1 \approx N\) = total number of possible contacts of a person, \(S/N\) = the fraction of possible contacts of a person that are from group S, \(I\) = the number of infected persons at time \(t\), therefore \(\beta SI/N\) = Number of people transmitted from group S to group I per unit of time.

3. ANN based Adaptive Incremental Learning (ANNAIL) of Model Parameters

The next task is to learn the model parameters which can be quite challenging in an epidemic scenario like Covid-19 as the model parameters are supposed to change with time. This section proposes an Artificial Neural Network (ANN) based Adaptive Incremental Learning technique (ANNAIL) for online learning of the SIRVD Model parameters with the following assumptions:

(i) The rate of infection \(\beta\) as a function of time \((t)\) is the major challenge for parameter learning. It is affected by external factors like degree of social distancing, lockdown etc. In case of a lockdown \(\beta\) decreases exponentially. Therefore, in order to take into account both the lockdown and no lockdown scenarios, \(\beta\) has been modelled as:

\[
\beta = \begin{cases} 
\beta_0, & \text{Outside Lockdown} \\
\beta_0 e^{-\frac{t}{t_l}} + \beta_1, & \text{During Lockdown} 
\end{cases} \tag{8}
\]

where \(t_l\) is the time when the lockdown begins. Therefore, the learning algorithm has to learn 3 parameters \((\beta_0, \beta_1, t_p)\) to find \(\beta\).

(ii) The rate of reinfection \(\sigma\) has been assumed to be zero for Covid-19 Disease. This is based on the fact that the human body produces a non-specific innate response to a viral infection initially using neutrophils, macrophages, and dendritic cells. However, this is followed by a more specific adaptive response in the form of development of proteins called immunoglobulins which act as antibodies specifically binding to the virus. This is coupled with the formation of T-cells which generate cellular immunity by identifying and eliminating the cells that are infected with the virus. Generally, sufficient presence of such antibodies in collaboration with cellular immunity prevents reinfection after recovery. Although every recovered patient may not develop complete immunity, but that is the case for the most of them [23]. In case of Covid-19, although research is till going on to reach a conclusive opinion, promising studies suggest that nearly all the recovered patients develop such antiviral immunoglobulin-G (IgG) antibodies and are immune to reinfection [24]. This lays the basis for mass Serological testing for Covid-19 being practised by many governments across the world in which the blood samples of people are tested for the presence of these antibodies indicating present or past Covid-19 infection. Further, the data on reinfection, even if rare, is not available.

(iii) Rate of Recovery \(\gamma\) and Rate of Death \(\delta\) are affected by factors like change in health care facilities, possible overcrowding of hospitals, development of new drugs to manage or treat the disease etc.

For a typical neural network or any other technique of model parameter estimation, the training data is required first to train the model before applying it on future scenarios. However, in case of an epidemic like Covid-19, the training data is continuously evolving with time and the model needs to be trained and
executed at the same time as the model parameters may change over time based on different external factors like government policies, social distancing etc which can be known only from newly arriving data sets. To overcome this challenge, a deep learning based technique is proposed here for the model to learn these parameters from new data sets in an adaptive manner while continuously updating the old models without the need to build the model from scratch every time a new training data set is received.

### 3.1. Adaptive Incremental Learning

Deep learning and other machine learning techniques stand out in solving problems of data based model parameter estimation due to their state-of-the-art results. However, they face the problem of catastrophic forgetting which reduces their performance as new training data becomes available with time. This is because the typical neural networks require the entire dataset to update the model each time a new training data set becomes available with time. To address such issues, different incremental learning algorithms have been suggested in the literature [25, 26, 27].

Incremental Learning refers to an online learning technique of continuous model adaptation under a scenarios of continuously evolving training data. Therefore, storage or access to the previously observed data is not required each time a new data set is received, as in case of an epidemic like Covid-19. In order to adapt the model parameters in light of new data, it is not needed to use all the previously accumulated data for developing the model from scratch. Rather, the Learning Network modifies the previous hypothesis to adapt to the new data chunk.

### 3.2. Problem Formulation

In this paper, hypothesis generation via an Artificial Neural Network (ANN) is proposed. Let $D_{j-1}$ be the data set received between time $t_{j-1}$ and $t_j$, and $h_{j-1}$ be the hypothesis generated on this data set. The hypothesis $h_j$ for a new data set $D_j$ received between time $t_j$ and $t_{j+1}$ is a function of $D_j$ and $h_{j-1}$ only as under:

$$h_j = f(h_{j-1}, D_j) \quad (9)$$

The experience gained from this step is stored and integrated to support in future adaptation process. Thus the objective here is to integrate the previously learned knowledge into the new raw data set to adapt the model parameters accordingly; and to accumulate this experience over time to increase the model efficiency, accuracy and flexibility.

### 3.3. Proposed Framework

The proposed framework for the above problem is shown in Figure 4. The ANN is based on a non-linear activation function for successful regression analysis. The hidden layers are represented by the function $f_{NN}$. With the continuous data stream, the weight distribution functions are generated to describe the learning capability of the ANN where the decision boundary is adjusted to focus especially on the hard to learn data examples. The algorithm for this framework is given in Algorithm 1.

This algorithm is run in a top-down and horizontal signal flow, as shown in Figure 4. The adaptive nature of this algorithm is due to the mapping function based on ANN which estimates the initial distribution function $\Phi_{t-1}$ for $D_t$ while providing a quantitative approach to indicate the learning power of the new data set based on previously trained model. $\Phi_{t-1}$ is applied to the new data set to find pseudo-error, $\gamma$. Thus a hard to learn example will have higher $\gamma$ in step (iii) of the Learning procedure, and will in turn receive higher weight in step (iv). This ensures the adaptive nature of the algorithm.
Algorithm 1: The ANN-AIL framework

(1) Inputs:
   Previous Knowledge at time \( (t-1) \);
   - \( D_{t-1} \) with \( m \) instances: \( \{ x_i, y_i \} \), \( i=1,2,...,m \), where \( x_i \) represents the input vector and \( y_i \) represents the output.
   - Associated mapping function output: \( \Phi_{t-1} \).
   New Data set at time \( t \)
   \( D_t \) with \( m \) instances: \( \{ x_i, y_i \} \), \( i=1,2,...,m \).

(2) Learning Procedure
   (i) Find initial distribution function for \( D_t \)
   \[ \Phi_{t-1} = f_{NN}(D_{t-1}, D_t, \Phi_{t-1}) \]
   where \( f_{NN} \) is the ANN defined mapping function.
   (ii) Apply hypothesis \( h_{t-1} \) to \( D_t \), and find the pseudo-error
   \[ \varepsilon_{t-1} = \sum_{j \in h_{t-1}(x), y_j} \Phi_{t-1}(j) \]
   (iii) Define
   \[ \Upsilon_{t-1} = \frac{\varepsilon_{t-1}}{1 - \varepsilon_{t-1}} \]
   (iv) Update the ANN’s mapping function for \( D_t \):
   \[ \Phi_t(j) = \frac{\Phi_{t-1}(j)}{N_t} \times \begin{cases} \Upsilon_{t-1} & \text{if } h_{t-1}(x_j) = y_j \\ 1 & \text{otherwise} \end{cases} \]
   where \( N_t \) is a normalization constant.
   (v) Develop new hypothesis \( h_t \) from \( D_t \) and \( \Phi_t \), using
   \[ h_t = f_{NN}(h_{t-1}, D_t) \]
   (vi) Repeat the above procedure for \( D_{t+1} \).

(3) Output
   The final hypothesis is:
   \[ h_{\text{final}}(x) = \arg \max_{y \in Y} \sum_{f \in h_{t-1}(x) \rightarrow y} \log \left( \frac{1}{1 + e^{-f}} \right) \]
   where \( T \) is the set of incrementally developed hypotheses in the learning life.

3.4 Mapping Function Design

Mapping function connects the past experiences to the new data in an adaptive fashion. There can be many ways to design the mapping function. However, we implemented the non-linear regression by an ANN based approximation of mapping function owing to its flexibility. Any such Neural Network based function approximation technique can be used. As in illustration we take the Multilayer Perceptron (MLP) in this paper. This is shown in Figure 5.

The input is an n-dimensional vector (for example, Number of infections, deaths, recoveries in an epidemic) of example \( t \). The distribution function is currently estimated as \( J_{t-1} \), \( W \) represents the weights of a layer. Backpropagation is used to tune the weights \( W \) of different layers, where error function is defined as

\[ e(k) = J_{t-1}(k-1) - \Phi_{t-1}(k-1) \]  
\[ E(k) = \frac{1}{2} e^2(k) \]

where \( k \) is the training epoch of the backpropagation. The neural network gives the following output:

\[ J(k) = \frac{1 - e^{-w}}{1 + e^{-w}} \]  
\[ \nu(k) = \sum_{f=1}^{N_h} w_f^{(2)}(k) g_f(k) \]  
\[ g_f(k) = \frac{1 - e^{-h_f(k)}}{1 + e^{-h_f(k)}} \]  
\[ h_f(k) = \sum_{i=1}^{n} w_f^{(1)}(k) x_{i,f}(k) \]
Outputs of vaccination (Inputs: Number of New infections, deaths and recoveries; rate the actual data. The inputs and outputs in this algorithm were: the results given by the model during testing are very close to 6. and the remaining 20 % was used for testing as shown in Figure 3. Model Validation of the MLP. 

Backpropagation: Weight adjustment for the hidden to output layer is described as:

$$\Delta w_f^{(2)} = \alpha(k) \left( -\frac{\partial E(k)}{\partial w_f^{(2)}(k)} \right)$$ (16)

$$\frac{\partial E(k)}{\partial w_f^{(2)}(k)} = \frac{\partial E(k)}{\partial J(k)} \frac{\partial J(k)}{\partial \nu(k)} \frac{\partial \nu(k)}{\partial w_f^{(2)}(k)}$$

$$= e(k) \frac{1}{2} (1 - (J(k))^2) g_f(k)$$ (17)

Similarly, weight adjustments for the input to hidden layer is described as:

$$\Delta w_{f,j}^{(1)} = \alpha(k) \left( -\frac{\partial E(k)}{\partial w_{f,j}^{(1)}(k)} \right)$$ (18)

$$\frac{\partial E(k)}{\partial w_{f,j}^{(1)}(k)} = \frac{\partial E(k)}{\partial J(k)} \frac{\partial J(k)}{\partial g_f(k)} \frac{\partial g_f(k)}{\partial h_f(k)} \frac{\partial h_f(k)}{\partial w_{f,j}^{(1)}(k)}$$

$$= e(k) \frac{1}{2} (1 - (J(k))^2) w_f^{(2)}(k) \cdot \frac{1}{2} (1 - g_f^2(k)). x_{j}(k)$$ (19)

where $\alpha(k)$ describes the learning rate. Estimation of initial distribution function $f_0$ for $D$: required only the feedforward path of the MLP.

3.5. Model Validation

Here, the model is validated for Covid-19 in the whole country of India, where the first 80% of data was used for training and the remaining 20% was used for testing as shown in Figure 6.

It is clearly evident from the plots shown in the figure that the results given by the model during testing are very close to the actual data. The inputs and outputs in this algorithm were: Inputs: Number of New infections, deaths and recoveries; rate of vaccination (α). Outputs: $\beta_0, \beta_1, \tau_0, \delta, \gamma$

These outputs are fed to the analytical SIRD model at every time instant when a new set of input data is received to simulate and forecast different scenarios. Model Validation for each of the 5 worst hit states is separately presented in the Results and Discussion section.

4. Results and Discussion

For forecast error calculation, percentage mean absolute error (MAE) has been defined as:

$$MAE\% = \frac{1}{n} \sum_{j=1}^{n} \left| \frac{y_j - \hat{y}_j}{y_j} \right| \times 100$$

where $y$ represents the observed values and $\hat{y}$ is the predicted value for $n$ number of observations. Figures 7-11 show the forecast of Covid-19 growth in the worst hit states of India for next 30 days. These results are summed up in Table 1.

Maharashtra, with the highest number of recorded cases, accounts for 30% of the total country’s caseload. Mumbai, the capital of Maharashtra and dubbed as the financial capital of India, has witnessed the steepest rise in the growth of Covid-19 and currently forms the Covid-19 epicentre of India. This is particularly due to high population density, tourist influx and dependence on public transport. Dharavi, a locality in Mumbai, is of particular concern as it is considered to be one of Asia’s largest slums with an area of just over 2.1 square kilometres and a population of about 1,000,000. Therefore, the rules of social distancing are difficult to observe there. Further, Mumbai has shortage of ICU beds and dedicated COVID-19 hospitals as well. As shown in Figure 7, Maharashtra is expected to see a continuous upsurge of cases and deaths in coming weeks. MAE describing the forecast error for Maharashtra for past 30 days is as low as 1.86% which can be attributed to the extensive testing and reporting.

Delhi, the capital of India, stands next to Maharashtra in facing the brunt of Covid-19 pandemic. Our model records an MAE of 19.79% for Delhi over past 30 days. The apparent reasons for this relatively large error is low testing rate and higher mobility of infected people. The government of India allowed migrant workers and people from other states stranded in Delhi to move to their home states after 2nd phase of the lockdown which disturbed the model assumption that there is zero inflow and outflow of population. Further, missed reports of deaths were recently added to the official data by the Ministry of Health which increased the forecast error. As shown in Figure 8, Delhi will continue seeing moderate rate of increase in the number of cases in coming weeks, partly due to now updated quarantine facilities and health care.

The scenario for Tamil Nadu as shown in Figure 9 is similar to that of Delhi except that the forecast error is lower owing to better testing and reporting in the state. Further, the mortality rate in Tamil Nadu is low owing to its relatively better health infrastructure. Chennai, the capital of Tamil Nadu, is known...
Figure 7: 30-day Forecast for Maharashtra: Red line shows the start of prediction window

(a) Prediction comparison with observed data for past 30 days
(b) Total number of cases
(c) Active Infections
(d) Deaths

Figure 8: 30-day Forecast for Delhi: Red line shows the start of prediction window

(a) Prediction comparison with observed data for past 30 days
(b) Total number of cases
(c) Active Infections
(d) Deaths
Figure 9: 30-day Forecast for Tamil Nadu: Red line shows the start of prediction window.

(a) Prediction comparison with observed data for past 30 days
(b) Total number of cases
(c) Active Infections
(d) Deaths

Figure 10: 30-day Forecast for Gujarat: Red line shows the start of prediction window.

(a) Prediction comparison with observed data for past 30 days
(b) Total number of cases
(c) Active Infections
(d) Deaths
Figure 11: 30-day Forecast for Uttar Pradesh: Red line shows the start of prediction window

Table 1: Covid-19 scenario in the worst hit states of India (2020)

| State       | First Case Recorded | Total cases on 27 June | Total Deaths on 27 June | Mortality rate on 27 June | Cases Forecast upto 27 July | Deaths Forecast upto 27 July | Mortality Forecast on 27 July |
|-------------|---------------------|------------------------|-------------------------|---------------------------|-----------------------------|-----------------------------|-----------------------------|
| Maharashtra | 10 Mar              | 147,741                | 6931                    | 4.70%                     | 300,300                     | 19,078                      | 6.35%                       |
| Delhi       | 03 Mar              | 73,780                 | 2429                    | 3.30%                     | 175,720                     | 5851                        | 3.32%                       |
| Tamil Nadu  | 18 Mar              | 70,977                 | 911                     | 1.28%                     | 229,220                     | 2872                        | 1.68%                       |
| Gujarat     | 20 Mar              | 29,520                 | 1753                    | 5.93%                     | 48,183                      | 2617                        | 5.43%                       |
| Uttar Pradesh | 04 Mar            | 20,193                 | 611                     | 3.03%                     | 48,754                      | 1545                        | 3.17%                       |
for medical tourism in India. However, Tamil Nadu state is expected to see a steep rise in the number of cases and deaths for next one month and hence the administration should frame the policies accordingly.

The case of Gujarat as shown in Figure 10 is alarming as the mortality rate is 5.93% which is almost double the national average. This can be ascribed to the delay in testing and relatively low quality of public healthcare. The total number of cases is expected to rise up to 48 thousand within next one month. Although the number of cases in Uttar Pradesh, as shown in Figure 11, is lower considering the fact that it is the largest populated state in India with a population of around 204 million, the Covid-19 cases in Uttar Pradesh are expected to grow more than double within next one month.

5. Conclusion

This work studies the spread of Covid-19 disease in India’s five worst hit states. An analytical epidemiological model was developed for Covid-19 pandemic where model parameters are continuously updated to intelligently adapt to new data sets using an ANN based adaptive online Incremental Learning technique. In a scenario of continuously evolving training data, unlike typical Deep Learning techniques, the model eliminates the need to retrain or rebuild the model from scratch every time a new training data set is received. India was taken as a case study. However, this model can be applied to any population in the world and would be a useful tool for policy makers, health officials and researchers in improving decision making efficiency, policy formulation, monitoring and forecasting of an epidemic. The model used here is non-intrusive, adaptive, intelligent, real-time and online in nature, therefore it is immune to loss of accuracy, reliability or computational performance arising due to limitations like run-time duration, size of training data, computational complexity, change in transmission dynamics due to mutations in virus or bacteria, change in prevention mechanisms or government policies. Even if the epidemic continues for decades in the whole world, the model will keep working efficiently on daily basis without any decay in performance or RTE (run-time environment). This work is a demonstration of the applicability and usefulness of the incremental learning technique, as an adaptive deep learning strategy, in data science problems where training data is not available beforehand but evolves continuously with time. Modeling and Simulation work was carried in Matlab.

The simulation results suggest an alarming rise in the number of cases in coming weeks with flattening of infection curves still far from sight. This should prompt the healthcare officials and policy makers to take appropriate decisions to combat the rise of disease spread. It is pertinent to mention here that the estimates provided herein are as good as the data provided by the official sources. Under-testing and under-reporting of disease can subject the forecasts to unavoidable errors. However, the information provided by this paper can provide significant insights into the pattern of disease spread and help in making a realistic assessment of the situation necessary for optimal policy framing.

Funding

The doctoral research funding from Ministry of Human Resource Development, Government of India, in favor of the first author is duly acknowledged.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

[1] WHO, https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-mission-briefing-on-covid-19—12-march-2020 (2020).
[2] WHO, https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations (2020).
[3] N. van Doremalen, T. Bushmaker, D. H. Morris, M. G. Holbrook, A. Gamble, B. N. Williamson, A. Tamin, J. L. Harcourt, N. J. Thornburg, S. I. Gerber, J. O. Lloyd-Smith, E. de Wit, V. J. Munster, Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1, The New England journal of medicine, 382 (16) (2020) 15641567, https://doi.org/10.1056/NEJMrc2004973.
[4] P. G. Gibson, L. Qin, S. Puah, COVID-19 ARDS: clinical features and differences to usual pre-COVID ARDS, The Medical Journal of Australia Preprint, https://www.mja.com.au/journal/2020/covid-19-ards-clinical-features-and-differences-usual-pre-covid-ards.
[5] W. O. Kemark, A. G. McKendrick, Contributions to the mathematical theory of epidemics, Proceedings of the Royal Society A 115 (5) (1927) 700–721.
[6] F. Brauer, J. Wu, P. V. D. Driessche, Mathematical Epidemiology, Springer-Verlag, Berlin Heidelberg, 2008.
[7] J. D. Murray, Mathematical Biology: 1. An Introduction, Springer Science & Business Media, 2002.
[8] A. C. Osemwinyen, A. Diakhaby, Mathematical Modelling of the Transmission Dynamics of Ebola Virus, Applied and Computational Mathematics 4 (4) (2015) 313–320, doi: 10.11648/j.acm.20150404.19.
[9] B. J. Cohurn, B. G. Wagner, S. Blower, Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1), BMC medicine 7 (30), doi: 10.1186/1741-7015-7-30.
[10] D. Rafiq, A. Batool, M. A. Bazaz, Three months of covid-19: A systematic review and meta-analysis, Reviews in Medical Virology (2020) e2113 doi:10.1002/rmv.2113.
[11] D. Rafiq, S. A. Suhail, M. A. Bazaz, Evaluation and prediction of covid-19 in India: A case study of worst hit states, Chaos, Solitons & Fractals 139 (2020) 110014. doi:https://doi.org/10.1016/j.chaos.2020.110014.
[12] PIB, Pib, https://pib.gov.in/pressreleaseframepage.aspx?appid=1601095 (2020).
[13] Govt., Covid-19.in, https://www.mygov.in/covid-19/?cbs=1. (2020).
[14] Wikipedia, Covid-19 pandemic lockdown in india, https://en.wikipedia.org/wiki/COVID-19_pandemic_lockdown_in_India (2020).
[15] Brookings, India’s health structure, https://www.brookings.edu/blog/up-front/2020/03/24/is-indias-health-infrastructure-equipped-to-handle-an-epidemic/ (2020).
[16] C. B. Bridges, M. J. Kuehnert, C. B. Hall, Transmission of influenza: implications for control in health care settings, Clinical Infectious Diseases 37 (8) (2003) 1094–1101.
[17] S. Cauchemez, F. Carrat, C. Viboud, A. J. Valleron, P. Y. Boelle, A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data, Statistics in Medicine 23 (22) (2004) 3469–3487.
F. Carrat, J. Luong, H. Lao, A. V. Sall, C. Lajaunie, H. Wackernagel, A small-world-like model for comparing interventions aimed at preventing and controlling influenza pandemics, BMC Medicine 4 (3) (2006) 26–40.

C. E. Mills, J. M. Robins, M. Lipsitch, Transmissibility of 1918 pandemic influenza, Nature 432 (7019) (2004) 904–906.

K. O. Okosun, O. D. Makinde, I. Takaidza, Analysis of recruitment and industrial human resources management for optimal productivity in the presence of the HIV/AIDS epidemic, Journal of Biological Physics 39 (1) (2013) 99–121.

M. Lipsitch, T. Cohen, B. Cooper, J. M. Robins, S. Ma, L. James, M. Murray, Transmission dynamics and control of severe acute respiratory syndrome, Science 300 (5627) (2003) 1966–1970.

S. Riley, C. Fraser, C. A. Donnelly, A. C. Ghani, L. J. Abu-Raddad, A. J. Hedley, R. M. Anderson, Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health intervention, Science 300 (5627) (2003) 1961–1966.

WHO, immunity passports in the context of covid-19, https://www.who.int/news-room/commentaries/detail/immunity-passports-in-the-context-of-covid-19 (2020).

Q. Long, B. Liu, H. Deng, et al, Antibody responses to SARS-CoV-2 in patients with COVID-19, Nature MedicineHttps://doi.org/10.1038/s41591-020-0897-1.

H. Bouchachia, B. Gabrys, Z. Sahel, Overview of Some Incremental Learning Algorithms, in: Proceedings of the 2007 IEEE International Conference on Fuzzy Systems, 2007, pp. 1–6, doi: 10.1109/FUZZY.2007.4295640.

F. M. Castro, M. J. Marn-Jimnez, N. Guil, C. Schmid, K. Alahari, End-to-End Incremental Learning, Vol. 11206, Springer, Cham, 2018, computer Vision: ECCV 2018. Lecture Notes in Computer Science.

A. Gepperth, B. Hammer, Incremental learning algorithms and applications, European Symposium on Artificial Neural Networks (ESANN).
Declaration of Interest Statement

**Manuscript title:** A Deep Learning algorithm for modeling and forecasting of COVID-19 in five worst affected states of India.

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers’ bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

**Author names:**

Junaid Farooq  
(Corresponding Author)  
Electrical Engineering Department  
National Institute of Technology Srinagar, India - 190006  
Email: junaid_phd017@nitsri.net

Muhammad Abid Bazaz  
(Co-author)  
Electrical Engineering Department  
National Institute of Technology Srinagar, India - 190006  
Email: abid@nitsri.net