RESEARCH ARTICLE

Deep learning via LSTM models for COVID-19 infection forecasting in India

Rohitash Chandra®[1]*, Ayush Jain[2]⊙, Divyanshu Singh Chauhan[3]⊙

1 Transitional Artificial Intelligence Research Group, School of Mathematics and Statistics, University of New South Wales, Sydney, Australia, **2** Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, Assam, India, **3** Department of Mechanical Engineering, Indian Institute of Technology Guwahati, Assam, India

⊙ These authors contributed equally to this work.
* rohitash.chandra@unsw.edu.au

Abstract

The COVID-19 pandemic continues to have major impact to health and medical infrastructure, economy, and agriculture. Prominent computational and mathematical models have been unreliable due to the complexity of the spread of infections. Moreover, lack of data collection and reporting makes modelling attempts difficult and unreliable. Hence, we need to re-look at the situation with reliable data sources and innovative forecasting models. Deep learning models such as recurrent neural networks are well suited for modelling spatiotemporal sequences. In this paper, we apply recurrent neural networks such as long short term memory (LSTM), bidirectional LSTM, and encoder-decoder LSTM models for multi-step (short-term) COVID-19 infection forecasting. We select Indian states with COVID-19 hotpots and capture the first (2020) and second (2021) wave of infections and provide two months ahead forecast. Our model predicts that the likelihood of another wave of infections in October and November 2021 is low; however, the authorities need to be vigilant given emerging variants of the virus. The accuracy of the predictions motivate the application of the method in other countries and regions. Nevertheless, the challenges in modelling remain due to the reliability of data and difficulties in capturing factors such as population density, logistics, and social aspects such as culture and lifestyle.

1 Introduction

The *coronavirus disease 2019* (COVID-19) is an infectious disease caused by *severe acute respiratory syndrome coronavirus 2* (SARS-CoV-2) [1–3] which became a global pandemic [4]. COVID-19 was first identified in December 2019 in Wuhan, Hubei, China with the first confirmed (index) case was traced back to 17th November 2019 [5]. The COVID-19 pandemic forced many countries to close their borders and enforced a partial or full lockdown which had a devastating impact on the world economy [6–8]. Other major impact has been on agriculture [9, 10] which is a major source of income for the population in rural areas, especially in the developing world. The sudden lockdown in some countries created a number of problems, especially for low income communities [11, 12], and migrant workers [13]. Currently (8th

October, 2021) [14], more than 238 million cases have been reported across the world which resulted in more than 4.8 million deaths, and about 215 million have recovered [15, 16]. A significant portion of the recovered suffer from "long-covid" [17, 18], a term which refer to prolonged health problems which can last for months to entire lifetime. In terms of vaccinations (8th October, 2021) [14], 46.3% of the world population has received at least one dose and 6.4 billion doses have been administered globally. Furthermore, about 2.4% of population in low-income countries have received at least one dose.

The case of India has been unique when it comes to management of COVID-19 pandemic [13]. The first COVID-19 case in India was reported on 30 January 2020. India had two major waves of infections, where the first wave was from April to October 2020 and the second wave was from February to June 2021 [19] India currently (8th October, 2021) [14, 15] has 33,935,309 confirmed cases with 450,408 deaths which makes the largest in Asia, the second highest in the world after the United States. The fatality rate of COVID-19 in India is among the lowest in the world as it ranks 124th with 322 deaths in a million people. In comparison, United States has 2,197 and the United Kingdom has 2,013 deaths in a million people. The second wave (2021) had a devastating effect on the Indian health system and during the time around its peak in daily cases, India had highest daily infection in the world [14, 15]. On the bright side, India also has one of the fastest recovery rates in the world with 236,610 active cases, and ranks 9th in the world although 2nd in total cases.

In terms of COVID-19 forecasting, prominent computational and statistical models have been unreliable due to the complexity of the spread of infections [20–22]. Given lack of data, it is challenging to develop a model while taking into consideration population density, effect of lock-downs, effect of viral mutations and variants such as the delta variant [23], logistics and travel, and qualitative social aspects such culture and lifestyle [24]. However, culture and lifestyle are examples of variable of interest that cannot be measured quantitatively. Due to qualitative nature of certain variables of interest such as lifestyle, and lack of data collection, modelling attempts have been mostly unreliable [25]. We need to re-look at the situation with latest data sources and most comprehensive forecasting models [26–28]. Moreover, a number of other limitations exists, such as noisy or unreliable data of active cases [29], changing mortality rate given different variants, and asymptotic carriers [30, 31]. There have been reports that the models lack a number of limitations and failed in several situations [25]. Despite these challenges, it has been shown that country based mitigation factors in terms of lockdown level and monitoring has a major impact on the rate of infection [32]. We note that limited work has been done using deep learning-based forecasting models for COVID-19 in India, although the country has one of world's largest population with highly populated and highly dense cities. There is a need to evaluate latest deep learning models for forecasting COVID-19 in India, that takes into account both the first (2020) and the second wave (2021) of infections.

Deep learning models such as recurrent neural networks (RNNs) are well suited for modelling spatiotemporal sequences [33–37] and modeling dynamical systems, when compared to simple neural networks [38–40]. The limitation in training RNNs for long-term dependencies in sequences where data sequences span hundreds or thousands of time-steps [41, 42] have been addressed by *long short-term memory* networks (LSTMs) [35]. LSTMs have been used for COVID-19 forecasting in China [43] with good performance results when compared to epidemic models. LSTMs have also been used for COVID-19 forecasting in Canada [44]. Other deep learning models such as convolutional neural networks (CNNs) have recently shown promising performance for time series forecasting [45, 46]. Hence, they would also be suited for capturing spatiotemporal relationship of COVID-19 transmission with neighbouring states in India.

In this paper, we employ LSTM models in order to forecast the spread of COVID-infections among selected states in India. We select Indian states with COVID-19 hotpots and capture the first and second wave of infections in order to later provide two months ahead forecast. We first employ univariate and multivariate time series forecasting approaches and compare their performance for short-term (4 days ahead) forecasting. We also present visualisation and analysis of the COVID-19 infections and provide open source software framework that can provide robust predictions as more data gets available. The software framework can also be applied to different countries and regions.

The rest of the paper is organised as follows. Section 2 presents a background and literature review of related work. Section 3 presents the proposed methodology with data analysis and Section 4 presents experiments and results. Section 5 provides a discussion with discussion of future work and Section 6 concludes the paper.

2 Related work

The pandemic has greatly affected and transformed work environment and lifestyle. COVID-19 lockdowns and restrictions of movement has given rise to e-learning [47–49] and telemedicine [50], and created opportunities in applications for geographical information systems [51]. The lockdown showed positive impact on the environment [52, 53], especially for highly populated and industrial nations with high air pollution rate [54]. Zambrano-Monserrate et. al highlighted the positive indirect effects revolve around the reduction air pollutants in China, France, Germany, Spain, and Italy [52]. However, the way medical pollutants and domestic waste were discarded during lockdowns has been an issue [52]. COVID-19 lockdowns and infection management raised concerns about prejudices against minorities and people of colour in developed countries such as the United States [55]. Furthermore, there has been a significant impact on mental health across the globe [56, 57].

It has been shown that in some countries, comprehensive identification and isolation policies have effectively suppressed the spread of COVID-19. Huang et. al [58] presented an evaluation of identification and isolation policies that effectively suppressed the spread of COVID-19 which further contributed to reduce casualties during the phase of a dramatic increase in diagnosed cases in Wuhan, China. The authors recommended that governments should swiftly execute the forceful public health interventions in the initial stage of the pandemic. However, such policies have not been that effective for other countries, such as the first wave of infections and associated lockdowns in India [26].

2.1 Modelling and forecasting COVID-19

A number of machine learning and statistical models have been used for modelling and forecasting COVID-19 in different parts of the world. Saba and Elsheikh presented simple autoregressive neural networks for forecasting the prevalence of COVID-19 outbreak in Egypt which showed relatively good performance when compared to officially reported cases [22]. Yousaf et. al used *auto-regressive integrated moving average* (ARIMA) model for forecasting COVID-19 for Pakistan [21]. The model predicted that the number of confirmed cases would increase by factor of 2.7 giving 95% prediction interval by the end of May 2020, to 5681—33079 cases. However, Pakistan reported around 70,000 cases [14] end of May 2020, and hence the model was poor in prediction. Velásquez and Lara used Gaussian process regression model for forecasting COVID-19 infection in the United States [20]. The authors show that COVID-19 would peak in United States around July 14th 2020, with about 132,074 deaths and 1,157,796 infected individuals at the peak stage. However, the actual cases by July 14th reached more

than 3.5 million with more than 139 thousand deaths [14, 15] which shows that the model was close in forecasting deaths but forecast of total cases was poor.

Chimmula and Zhand used LSTM neural networks for time series forecasting of COVID-19 transmission in Canada [44]. The authors predicted the possible ending point of the outbreak around June 2020 and compared transmission rate of Canada with Italy and the United States. Canada reached the daily new cases peak by 2nd May 2020 [14, 15], and since then, new cases has been drastically reducing. Therefore we can say that the approach by the authors was somewhat close in reporting the peak for COVID-19 in Canada. Chakraborty and Ghosh [27] used hybrid ARIMA and wavelet-based forecasting model for short-term (ten days ahead) forecasts of daily confirmed cases for Canada, France, India, South Korea, and the United Kingdom. The authors also applied an optimal regression tree algorithm to find essential causal variables that significantly affect the case fatality rates for different countries. Maleki et. al [28] used autoregressive time series models based on mixtures of normal distribution from confirmed and recovered COVID-19 cases worldwide.

Ren et al. [24] analysed spatiotemporal variations of the epidemics before utilizing the ecological niche models with nine socioeconomic variables for identifying the potential risk zones for megacities such as Beijing, Guangzhou and Shenzhen. The results demonstrated that the method was capable of being employed as an early forecasting tool for identifying the potential COVID-19 infection risk zones. Alzahrani et al. [59] used autoregressive and ARIMA models for COVID-19 in Saudi Arabia with data till 20th April 2020 and predicted 7668 daily new cases by 21st May 2020 given stringent precautionary control measures were not implemented. However, Saudi Arabia on 21st May 2020 reported 2532 actual cases [14, 15]; hence, the model has shown poor performance. Singh et al. [60] presented a hybrid of discrete wavelet decomposition and ARIMA models in application to one month forecast the casualties cases of COVID-19 in most affected countries back then which included France, Italy, Spain, United Kingdom and and United Sates. The study found that the hybrid model was better than standalone models. Dasilva et al. [26] employed machine learning methods such as Bayesian regression neural network, cubist regression, k-nearest neighbors, quantile random forest, and support vector regression with pre-processing based on variational mode decomposition for forecasting one, three, and six-days-ahead the cumulative COVID-19 cases in five Brazilian and American states up to April 28th, 2020. Yang et al. [43] presented an epidemiological model that incorporated the domestic migration data and the most recent COVID-19 epidemiological data to predict the epidemic progression. The model predicted peak by late February, showing gradual decline by end of April 2020. This was one of the few attempts in prediction of COVID-19 infection trend in China [14, 15]; however, the actual peak was observed in early February 2020 and the spread of infections ended by the middle of March 2020.

Next, we review key studies of COVID-19 forecasting with deep learning models in India. Anand et al. [61] focused on forecasting of COVID-19 cases in India using RNNs such as LSTM and gated-recurrent units (GRU) with the dataset from 30th January 2020 to 21st July 2020. Bhimala et al. [62] incorporated the weather conditions of different states to make improve forecasting of the COVID-19 cases in different states of India. The authors made assumption that different humidity levels in different states will lead to varying transmission of infection within the population. They demonstrated that LSTM model performed better in the medium and long range forecasting scale when integrated with the weather data. Shetty [63] presented real-time forecasting using a simple neural network for the COVID-19 cases in the state of Karnataka in India where parameter selection for the model was based on cuckoo search algorithm. The study reported that the mean-absolute percentage error (MAPE) was reduced from 20.73% to 7.03% and the proposed model was further tested on the Hungary

COVID-19 dataset and reported promising results. Tomar and Gupta [64] developed LSTM model for 30-day ahead prediction of COVID-19 positive cases in India where they also studied the effect of preventive measures on the spread of COVID-19. They showed that with preventive measures and lower transmission rate, the spread can be reduced significantly. Gupta et al. [65] forecasted COVID-19 cases of India using support vector machines, prophet, and linear regression models. Similarly, Bodapati et al. [66] forecasted the COVID-19 daily cases, deaths caused and recovered cases with the help of LSTM networks for the whole world. Chaurasia and Pal [67] used several forecasting models such as simple average, single exponential smoothing, Holt winter method, and ARIMA models for COVID-19 pandemic.

A number of machine learning methods used in conjunction with deep learning models for COVID-19 forecasting for the rest of the world. Battineni et al. [68] forecasted COVID-19 cases using a machine learning method known as prophet logistic growth model which estimated that by late September 2020, the outbreak can reach 7.56, 4.65, 3.01 and 1.22 million cases in the United States, Brazil, India and Russia, respectively. Nadler et al. [69] used a model embedded in a Bayesian framework coupled with a LSTM network to forecast cases of COVID-19 in developed and developing countries. Istaiteh et al. [70] compared the performance of ARIMA, LSTM, multilayer perceptron and convolutional neural network (CNN) models for prediction of COVID-19 cases all over the world. They reported that deep learning models outperformed ARIMA model, and furthermore CNN outperformed LSTM networks and multi-layer perceptron. Pinter et al. [71] used hybrid machine learning methods consisting of adaptive network-based fuzzy inference systems (ANFIS) and mutlilayer perceptron (simple neural network) for COVID-19 infections and mortality rate in Hungary.

3 Methodology: Forecasting COVID-19 novel infections with deep learning models

We need to reconstruct the original time series into a state-space vector in order to train deep learning models for multi-step-ahead prediction. Taken's theorem expresses that the reconstruction can reproduce important features of the original time series [72]. Hence, an embedded phase space $Y(t) = [(x(t), x(t - T), . . ., x(t - (D - 1)T)]$ can be generated given an observed time series $x(t)$; where T is the time delay, D is the embedding dimension (window size) $t = 0$, 1, 2, . . ., $N - D - 1$, and N is the length of the original time series. Appropriate values for D and T need to selected to efficiently apply Taken's theorem for reconstruction [73]. Taken's proved that if the original attractor is of dimension d, then $D = 2d + 1$ would be sufficient [72].

3.1 LSTM network models

Recurrent neural networks (RNNs) have been prominent for modelling temporal sequences. RNNs feature a context layer to act as memory in order to project information from current state into future states, and eventually the output layer. Although number of different RNN architectures exist, the Elman RNN [33, 74] is one of the earliest which has been prominent for modelling temporal sequences and dynamical systems [39, 75, 76].

Training RNNs in the early days has been a challenging task. Backpropagation-through-time (BPTT) which is an extension of the backpropagation algorithm has been prominent in training RNNs [34]. BPTT features gradient-descent where the error is backpropagated for a deeper network architecture that features states defined by time. The RNN unfolded in time is similar to a multilayer perceptron that features multiple hidden layers. A major limitation of BPTT for simple RNNs has been the problem of learning long-term dependencies given vanishing and exploding gradients [41]. The LSTM network addressed this limitation with better capabilities in remembering the long-term dependencies using memory cells in the hidden
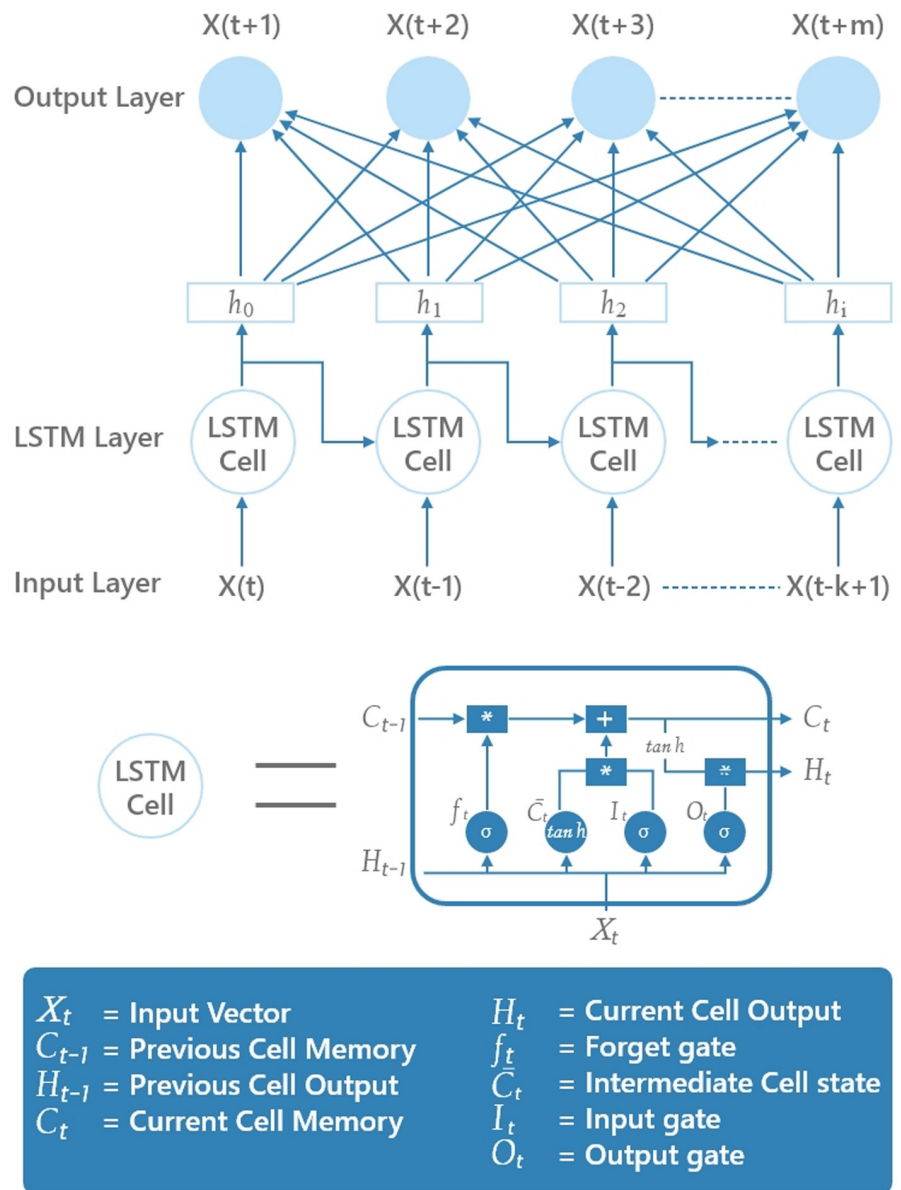
Fig 1. LSTM memory cell in the LSTM recurrent neural network.

layer [35]. The memory cells help in remembering the long-term dependencies in data as shown in Fig 1.

The LSTM network model calculates a hidden state output h_t by

$$
\begin{aligned}
i_t &= \sigma(x_t U^i + h_{t-1} W^i) \\
f_t &= \sigma(x_t U^f + h_{t-1} W^f) \\
o_t &= \sigma(x_t U^o + h_{t-1} W^o) \\
\tilde{C}_t &= \tanh(x_t U^c + h_{t-1} W^c) \\
C_t &= \sigma(f_t * C_{t-1} + i_t * \tilde{C}_t) \\
h_t &= \tanh(C_t) * o_t
\end{aligned}
\tag{1}
$$

where, i_t, f_t and o_t refer to the input, forget and output gates, at time t, respectively. c refers to the memory cell. x_t and h_t refer to the number of input features and number of hidden units, respectively. W and U are the weight matrices adjusted during learning along with b, which is the bias. Note that all the gates have the same dimensions d_h, which is given by the size of hidden state. $\tilde{C}_t$ is the intermediate cell state, and C_t is the current cell memory. The initial values at $t = 0$ are given by $C_0 = 0$ and $h_0 = 0$. Note that we denote (*) as element-wise multiplication.

3.2 Bi-directional LSTM networks

Conventional RNN and LSTM networks only make use of previous context state for determining future states. Bidirectional RNNs (BD-RNNs) [77] on the other hand, process information in both directions with two separate hidden layers which are then propagated forward to the same output layer. Hence, two independent RNNs are placed together to allow both backward and forward information about the sequence at every time step. The forward hidden sequence h_f, the backward hidden sequence h_b, and the output sequence y are computed by iterating the backward layer from $t = T$ to $t = 1$, and the forward layer from $t = 1$ to $t = T$.

Bi-directional LSTM networks (BD-LSTM) [78] can access longer-range context or state in both directions similar to BD-RNNs. BD-LSTM networks were originally proposed for word-embedding in natural language processing [78] tasks and have been used in several real-world sequence processing problems such as phoneme classification [78], continuous speech recognition [79], and speech synthesis [80].

BD-LSTM networks intake inputs in two ways; one from past to future, and another from future to past by running information backwards so that state information from the future is preserved. Given two hidden states combined in any point in time, the network can preserve information from both past and future as shown in Fig 2.

3.3 Encoder-decoder LSTM networks

The encoder-decoder LSTM network (ED-LSTM) [81] was introduced as a sequence to sequence model for mapping a fixed-length input to a fixed-length output. The length of the input and output may differ which makes them applicable in automatic language translation tasks (English to French for example) which can be extended to multi-step series prediction where both the input and outputs are of variable lengths. A latent vector representation is used to handle variable-length input and outputs by first encoding the input sequences, one at a time and then decoding it. We consider the input sequence $(x_1, \ldots, x_n)$ with corresponding output sequence $(y_1, \ldots, y_m)$, and estimate the conditional probability of the output sequence given an input sequence, i.e. $p(y_1, \ldots, y_m | x_1, \ldots, x_n)$. In the encoding phase, given an input sequence, the ED-LSTM network computes a sequence of hidden states. In the decoding phase, it defines a distribution over the output sequence given the input sequence as shown in Fig 3.

3.4 India: Situation report: 8th October, 2021

We provide a visual representation of the total number of COVID-19 infections for different states and union territories in India based on data till 8th October, 2021.

Tables 1–4 provide a rank of top ten Indian states with total cases 1st of every month. We see that largely populated states such as Maharashtra (population estimate of 123 million [82]) has been leading India in number of total cases through-out 2020. We note that state of Uttar Pradesh has estimate population of 238 million has managed better. Delhi has a relatively smaller population (estimated 19 million [82]), but high population density and hence been
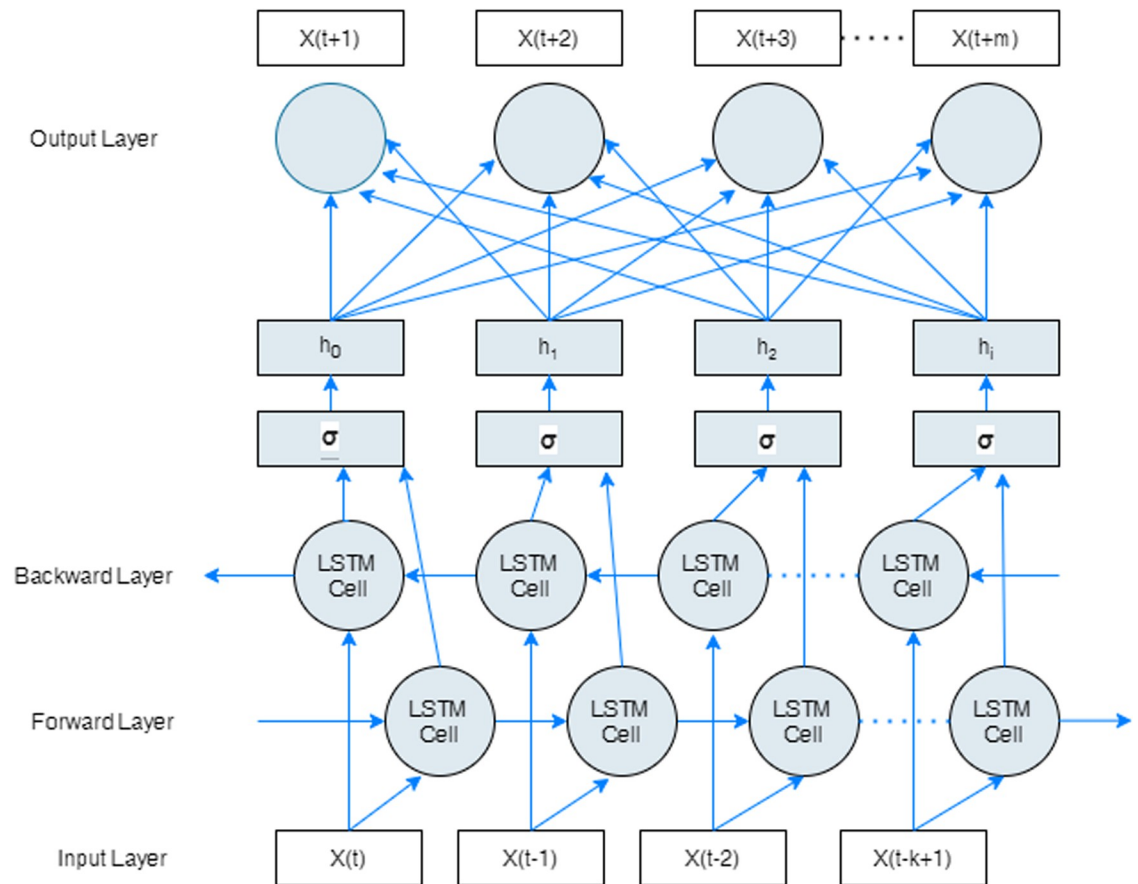
Fig 2. Bi-directional LSTM recurrent neural network.

https://doi.org/10.1371/journal.pone.0262708.g002

one of the leading states with COVID-19 infections (in top 6 throughout 2020). Tables 3 and 4 show that in 2021, Maharashtra continued leading; however, the second position was over-taken by Kerala from February which maintained the position since then. We note that from February to June 2021, India experienced the second-wave of infections from the delta-variant of the virus, with Maharastra and Kerala leading most of the time in terms of monthly infections. The first peak for novel cases in India was reached on September 16th 2020 with 97,894 daily and 93,199 weekly average novel cases [15]. The daily novel cases were steady for several months and then raised again from February 2021 for the second wave of infections. The peak of the second wave was reached around 7th May 2021, with 401,078 daily and weekly average of 389,672 novel infections [15]. The peak of deaths was reached around 21st May 2021 with 4194 deaths and 4188 weekly average.

Figs 4 and 5 present the total number of novel weekly cases for different groups of Indian states and union territories, which covers both the first and second wave of infections. We notice that the number of cases significantly increased after May 2020 which marks the first wave and then declined. Fig 4 (Panel a) focusing on major affected states show that Maharastra led the first and second wave of infections followed by Karnataka. In Fig 4 Panel (b), considering the Eastern states, we find that new cases in West Bengal drastically increased for the first wave of infections and it took Bihar longer to reach the peak when compared to Odisha and the others. In the second wave, West Bengal led the other states by a large margin. In the
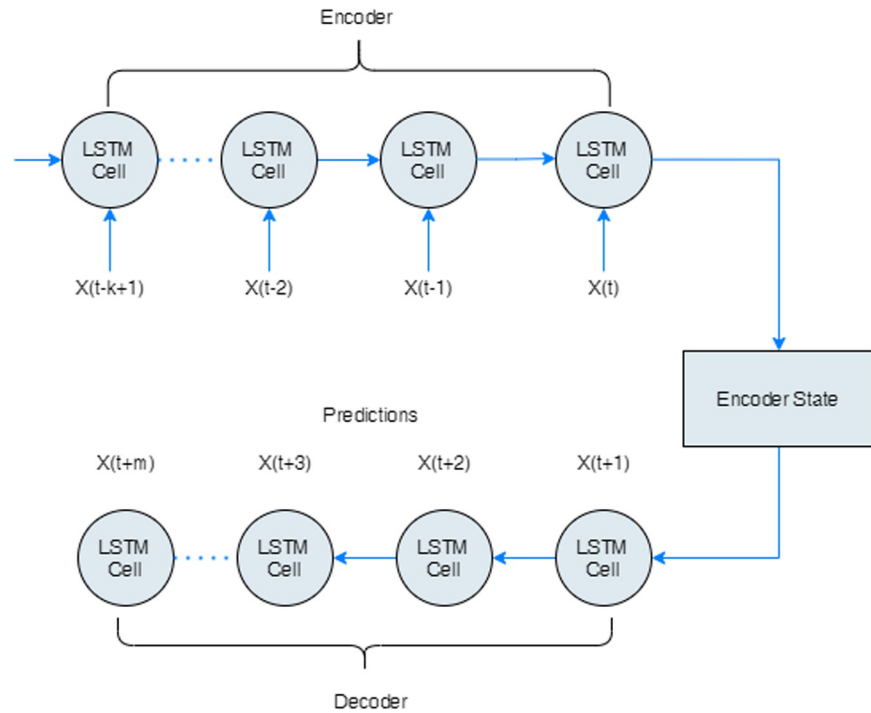
Fig 3. Encoder-decoder LSTM recurrent neural network. Note that the outputs $(y_1, \ldots, y_m)$ with corresponding input $(x_1, \ldots, x_n)$ are not explicitly shown.

https://doi.org/10.1371/journal.pone.0262708.g003

Table 1. Rank of states by number of novel total cases taken at the 1st of every month for April to August 2020 [15]. The number of novel cases are shown in brackets.

Rank	April	May	June	July	August
1	Maharashtra	Maharashtra	Maharashtra	Maharashtra	Maharashtra
	(302)	(10498)	(67655)	(174761)	(422118)
2	Kerala	Gujarat	Tamil Nadu	Tamil Nadu	Tamil Nadu
	(241)	(4395)	(22333)	(90167)	(245859)
3	Tamil Nadu	Delhi	Delhi	Delhi	Andhra Pradesh
	(234)	(3515)	(19844)	(87360)	(140933)
4	Delhi	Madhya Pradesh	Gujarat	Gujarat	Delhi
	(152)	(2719)	(16779)	(32557)	(135598)
5	Uttar Pradesh	Rajasthan	Rajasthan	Uttar Pradesh	Karnataka
	(103)	(2584)	(8831)	(23492)	(124115)
6	Karnataka	Tamil Nadu	Madhya Pradesh	West Bengal	Uttar Pradesh
	(101)	(2323)	(8089)	(18559)	(85461)
7	Telengana	Uttar Pradesh	Uttar Pradesh	Rajasthan	West Bengal
	(96)	(2281)	(7823)	(18014)	(70188)
8	Rajasthan	Andhra Pradesh	West Bengal	Telengana	Telengana
	(93)	(1463)	(5501)	(16339)	(62703)
9	Andhra Pradesh	Telengana	Bihar	Karnataka	Gujarat
	(83)	(1039)	(3815)	(15242)	(61438)
10	Gujarat	West Bengal	Andhra Pradesh	Andhra Pradesh	Bihar
	(82)	(795)	(3679)	(14595)	(51233)

https://doi.org/10.1371/journal.pone.0262708.t001

Table 2. Rank of states by number of novel total cases taken at the 1st of every month for September to December 2020 [15]. The number of novel cases are shown in brackets.

Rank	September	October	November	December
1	Maharashtra	Maharashtra	Maharashtra	Maharashtra
	(792541)	(1384446)	(1678406)	(1823896)
2	Andhra Pradesh	Andhra Pradesh	Karnataka	Karnataka
	(434771)	(693484)	(823412)	(884897)
3	Tamil Nadu	Karnataka	Andhra Pradesh	Andhra Pradesh
	(428041)	(601767)	(823348)	(868064)
4	Karnataka	Tamil Nadu	Tamil Nadu	Tamil Nadu
	(342423)	(597602)	(724522)	(781915)
5	Uttar Pradesh	Uttar Pradesh	Uttar Pradesh	Kerala
	(230414)	(399082)	(481863)	(602982)
6	Delhi	Delhi	Kerala	Delhi
	(174748)	(279715)	(43310)	(570374)
7	West Bengal	West Bengal	Delhi	Uttar Pradesh
	(162778)	(257049)	(386706)	(543888)
8	Bihar	Odisha	West Bengal	West Bengal
	(136457)	(219119)	(373664)	(483484)
9	Telengana	Kerala	Odisha	Odisha
	(127697)	(196106)	(290116)	(318725)
10	Assam	Telengana	Telengana	Telengana
	(109040)	(193600)	(240048)	(270318)

https://doi.org/10.1371/journal.pone.0262708.t002

Table 3. Rank of states by number of novel total cases taken at the 1st of every month for January to May 2021 [15]. The number of novel cases are shown in brackets.

Rank	January	February	March	April	May
1	Maharashtra	Maharashtra	Maharashtra	Maharashtra	Maharashtra
	(2026399)	(2155070)	(2812980)	(4602472)	(5746892)
2	Karnataka	Kerala	Kerala	Kerala	Karnataka
	(939387)	(1059403)	(1124584)	(1571183)	(2604431)
3	Kerala	Karnataka	Karnataka	Karnataka	Kerala
	(929178)	(951251)	(997004)	(1523142)	(2526579)
4	Andhra Pradesh	Andhra Pradesh	Andhra Pradesh	Uttar Pradesh	Tamil Nadu
	(887836)	(889916)	(901989)	(1252324)	(2096516)
5	Tamil Nadu	Tamil Nadu	Tamil Nadu	Tamil Nadu	Andhra Pradesh
	(838340)	(851542)	(886673)	(1166756)	(1693085)
6	Delhi	Delhi	Delhi	Delhi	Uttar Pradesh
	(635096)	(639289)	(662430)	(1149333)	(1691488)
7	Uttar Pradesh	Uttar Pradesh	Uttar Pradesh	Andhra Pradesh	Delhi
	(600299)	(603527)	(617194)	(1101690)	(1426240)
8	West Bengal	West Bengal	West Bengal	West Bengal	West Bengal
	(569998)	(575118)	(586915)	(828366)	(1376377)
9	Odisha	Odisha	Chhattisgarh	Chhattisgarh	Chhattisgarh
	(335072)	(337191)	(349187)	(728700)	(971463)
10	Rajasthan	Rajasthan	Odisha	Rajasthan	Rajasthan
	(317491)	(320336)	(340917)	(598001)	(939958)

https://doi.org/10.1371/journal.pone.0262708.t003

Table 4. Rank of states by number of novel total cases taken at the 1st of every month for June to September 2021 [15]. The number of novel cases are shown in brackets.

Rank	June	July	August	September
1	Maharashtra	Maharashtra	Maharashtra	Maharashtra
	(6061404)	(6303715)	(6464876)	(6541119)
2	Kerala	Kerala	Kerala	Kerala
	(2924165)	(3390761)	(4057233)	(4613937)
3	Karnataka	Karnataka	Karnataka	Karnataka
	(2843810)	(2905124)	(2949445)	(2972620)
4	Tamil Nadu	Tamil Nadu	Tamil Nadu	Tamil Nadu
	(2479696)	(2559597)	(2614872)	(2655572)
5	Andhra Pradesh	Andhra Pradesh	Andhra Pradesh	Andhra Pradesh
	(1889513)	(1966175)	(2014116)	(2045657)
6	Uttar Pradesh	Uttar Pradesh	Uttar Pradesh	Uttar Pradesh
	(1706107)	(1708441)	(1709335)	(1709761)
7	West Bengal	West Bengal	West Bengal	West Bengal
	(1499783)	(1528019)	(1548604)	(1565645)
8	Delhi	Delhi	Delhi	Delhi
	(1434188)	(1436265)	(1437764)	(1438685)
9	Chhattisgarh	Chhattisgarh	Odisha	Odisha
	(994480)	(1002008)	(1007750)	(1023735)
10	Rajasthan	Odisha	Chhattisgarh	Chhattisgarh
	(952422)	(977268)	(1004451)	(1005229)

https://doi.org/10.1371/journal.pone.0262708.t004

Northern states, shown in Fig 5 (Panel a), we find that Uttar Pradesh leading the first and second wave of infections which is not surprising since it is the most populous state of India. In the case of the relatively lowly populated states (small states) shown in Fig 5 (Panel b), we find that Goa and Tripura lead the first wave of infections and later in the second wave, Goa overtakes the rest of the states, significantly.

Fig 6 presents daily active cases and cumulative (total) deaths for key Indian states for 2021 [15]. We notice that the different states, such as Maharastra and Tamil Nadu have few to several weeks of lag in reaching the peak of novel daily cases. In terms of deaths, we do not see a sharp increase after July 2021 in most of states. Note that we chose not to show daily deaths in the same graphs since the scales between active cases and deaths are quite different.

4 Results

In this section, we present results of prediction of COVID-19 daily cases in India using prominent LSTM neural network models that includes, BD-LSTM and ED-LSTM with architectural details given earlier (Section 3).

4.1 Experimental design

Our experiments consider the evaluation of the respective models for univariate and multivariate prediction tasks. The data has been accessed from Indian Institute of Statistical Science—Bangalore [83], which was originally sourced from Ministry of Health and Family Welfare, Government of India website [84]. The dataset is based on daily novel cases which is normalised taking the maximum number of daily cases over the entire data into account. We start our analysis from 15th April 2020 and use rolling mean of 3 days to smoothen the original data.
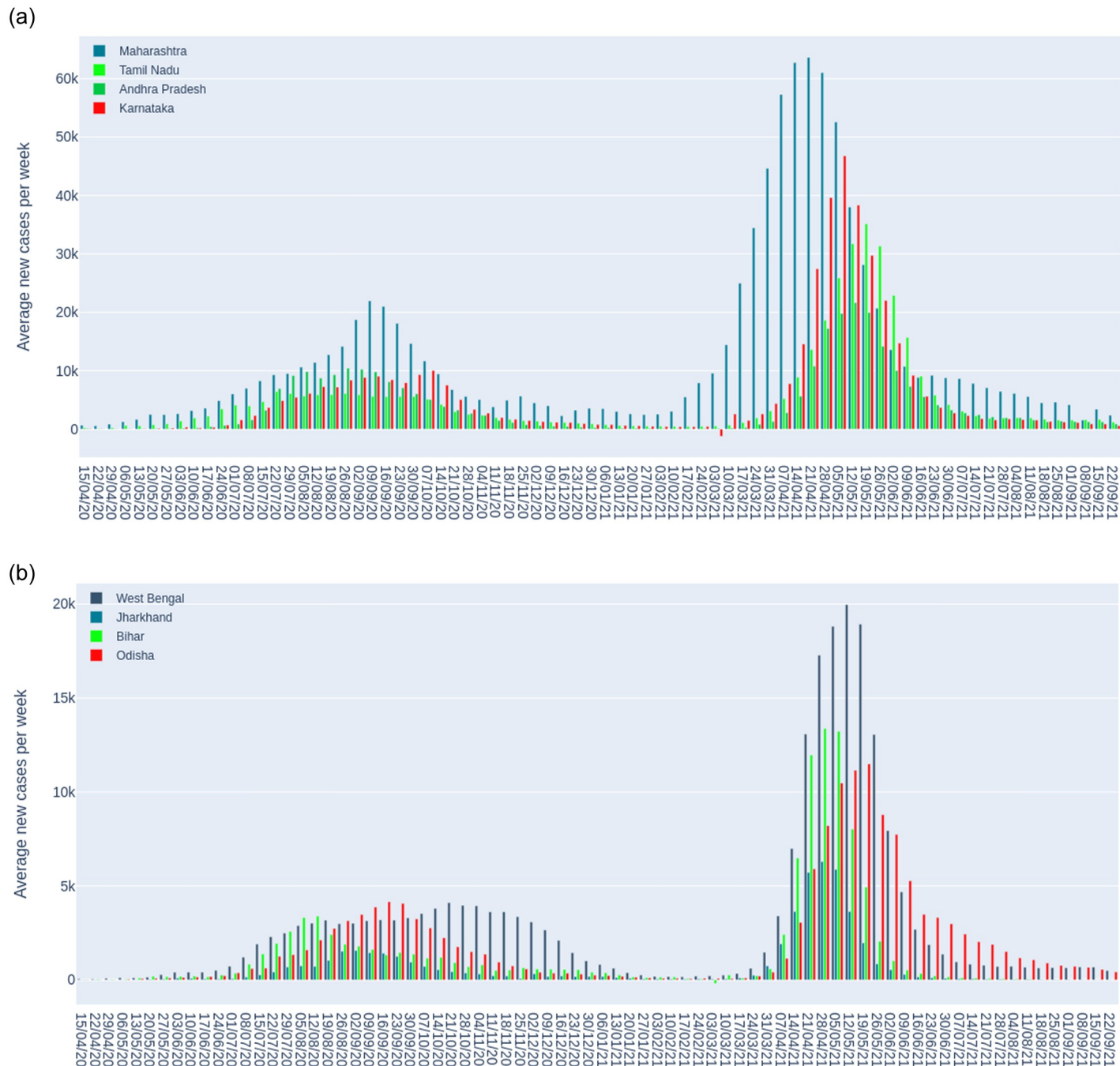
(a)

(b)

Fig 4. Weekly average of new cases for groups of Indian states and union territories [15].

We reconstruct the univariate and multivariate time series into a state-space vector using Taken's theorem [72] with selected values for embedding dimension window ($D = 6$) and time-lag ($T = 2$) for multi-step ahead (MSA) prediction. We consider four prediction horizons; i.e. $MSA = 4$, where each step is a prediction horizon.

The Adam optimizer is used for training the respective LSTM models. Tables 5 and 6 describe the topology for the respective LSTM models for univariate and multivariate cases,
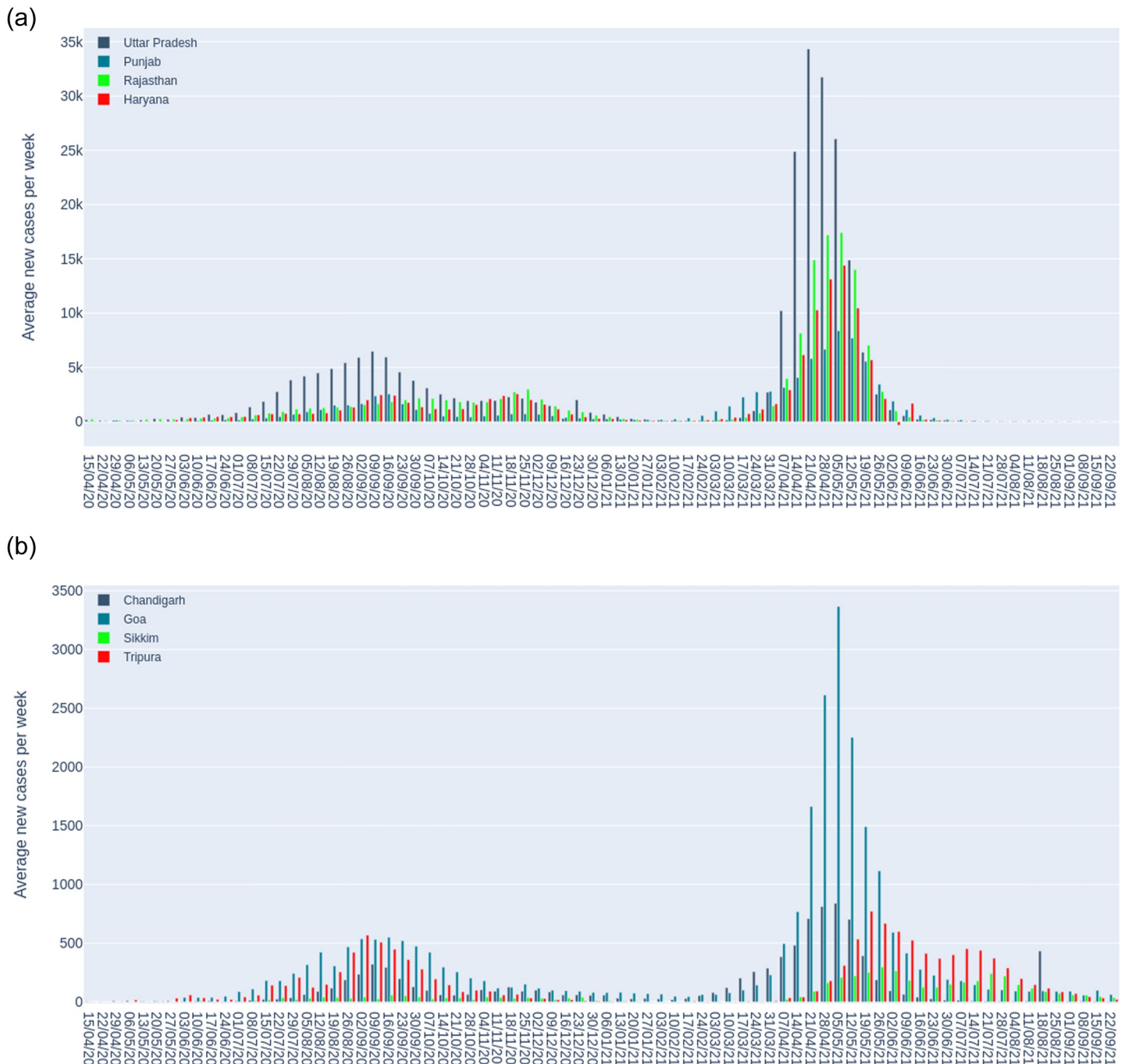
(a)

(b)

Fig 5. Weekly average of new cases for groups of Indian states and union territories [15].

https://doi.org/10.1371/journal.pone.0262708.g005

respectively. In case of multivariate model, the input contains four features which represents the adjacent states in relation to the state taken into account; i.e. the case of Maharashtra (Maharashtra, Gujarat, Madhya Pradesh, Uttar Pradesh) and the case of Delhi (Delhi, Rajasthan, Uttar Pradesh, Haryana). In multivariate model of India, we take all the states as input features to the multivariate model. We note that similar to univariate model, the multivariate
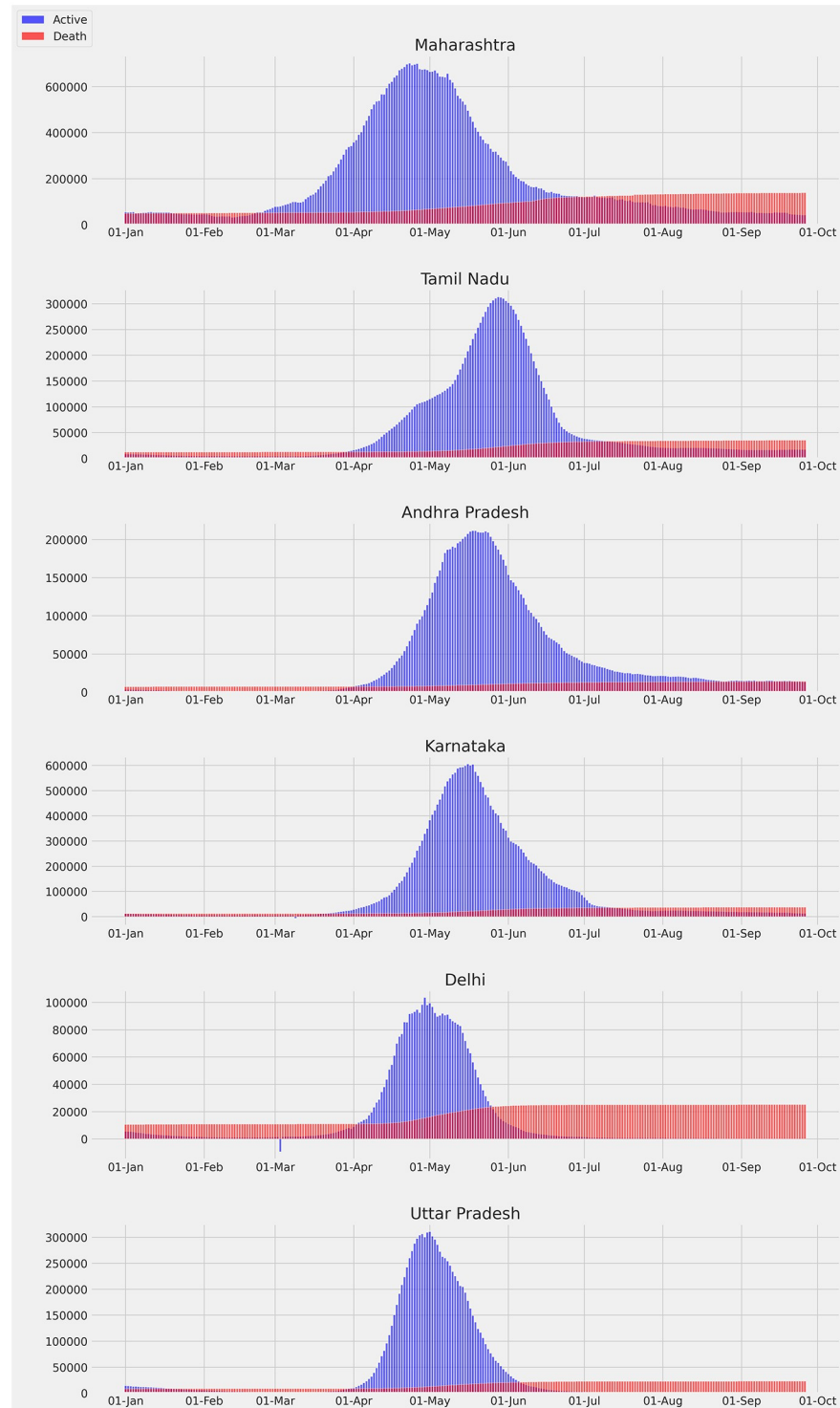
Fig 6. Daily active cases and cumulative death in key states of India for 2021 [15].

https://doi.org/10.1371/journal.pone.0262708.g006

Table 5. Respective LSTM model topologies for the univariate case.

Method	Input	Hidden layer 1	Hidden layer 2	Output
LSTM	(6,1)	32	32	(1,4)
BD-LSTM	(6,1)	32	16	(1,4)
ED-LSTM	(6,1)	32	-	(1,4)

model considers a selected embedding dimension window ($D = 6$) for multi-step-ahead prediction ($MSA = 4$). Tables 5 and 6 provides the details for the respective LSTM model typologies in terms of the hidden neurons and layers.

We review the performance of the respective methods in terms of scalability and robustness which refers to the ability of maintaining consistent prediction performance as the prediction horizon increases. We use the root mean squared error (RMSE) in Eq 2 as the main performance measure for prediction accuracy

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{2}$$

where $y_i, \hat{y}_i$ are the observed data, predicted data, respectively. N is the length of the observed data. We use RMSE for each prediction horizon and for each problem, we report the mean error for the respective prediction horizons.

We present the mean and 95% confidence interval for 30 experiment runs with different initialisation of model parameter space (weights and biases) in all the experiments. We use a dropout rate of 0.2 for the respective models in all the experiments.
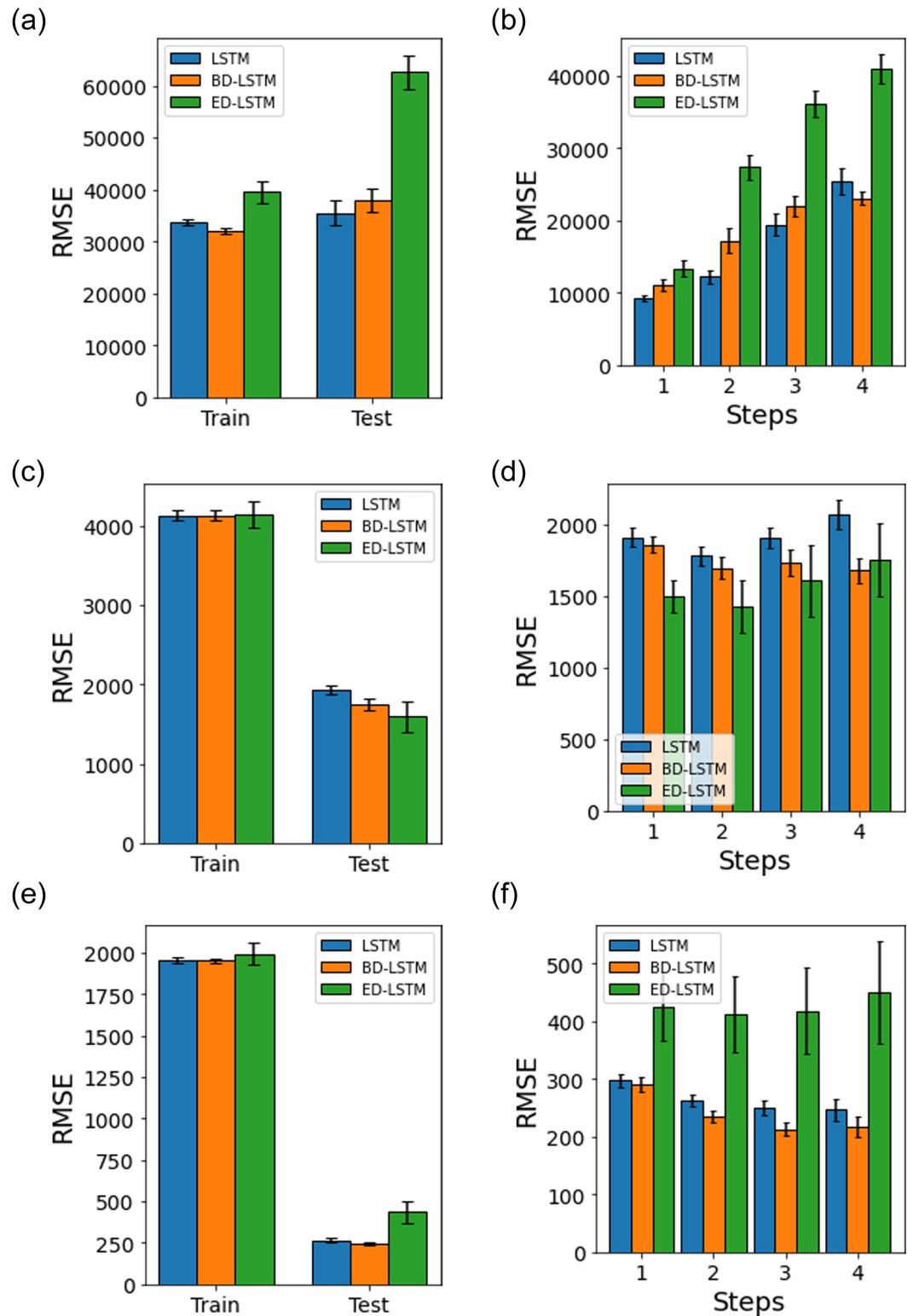
4.2 Prediction performance

We first evaluate the optimal strategy for creating training and testing datasets. We use static-split of training samples from 15th April 2020 to 15th May 2021. Our test set features data from 16th May 2021 to 27th September 2021; hence, the training data covers half of the second wave of the cases. In random-split, we create the train and test sets by randomly shuffling the dataset with the same size of the dataset as done for the static-split. We show results for entire case of India, and two leading states of COVID-19 infections, i.e. Maharashtra and Delhi. We investigate the effect of the univariate and multivariate approaches on the three models, (LSTM, BD-LSTM, ED-LSTM). Finally, using the best model, we provide a two month outlook for novel daily cases with a recursive approach, i.e. by feeding back the predictions into the trained models.

Fig 7 shows univariate LSTM, BD-LSTM and ED-LSTM models with a static-split of train/test dataset. Fig 8 shows univariate random splitting of train/test datasets using the same models. We observe that the prediction for the India dataset has a unique trend where the model is improving with increase in the prediction horizon (steps) when compared to Maharashtra and Delhi cases (Panels d and f). The corresponding cases in random-split given in Fig 8 show a

Table 6. Respective LSTM model topologies for the multivariate case.

Method	Input	Hidden-layer	Output
LSTM	(6,4)	32	(1,4)
BD-LSTM	(6,4)	32	(1,4)
ED-LSTM	(6,4)	32	(1,4)

Fig 7. Univariate LSTM, BD-LSTM and ED-LSTM model performance accuracy (RMSE) with a static-split for the train and test datasets, and test prediction horizons (steps). The error bars represent the mean and 95% confidence interval for 30 experiment runs.
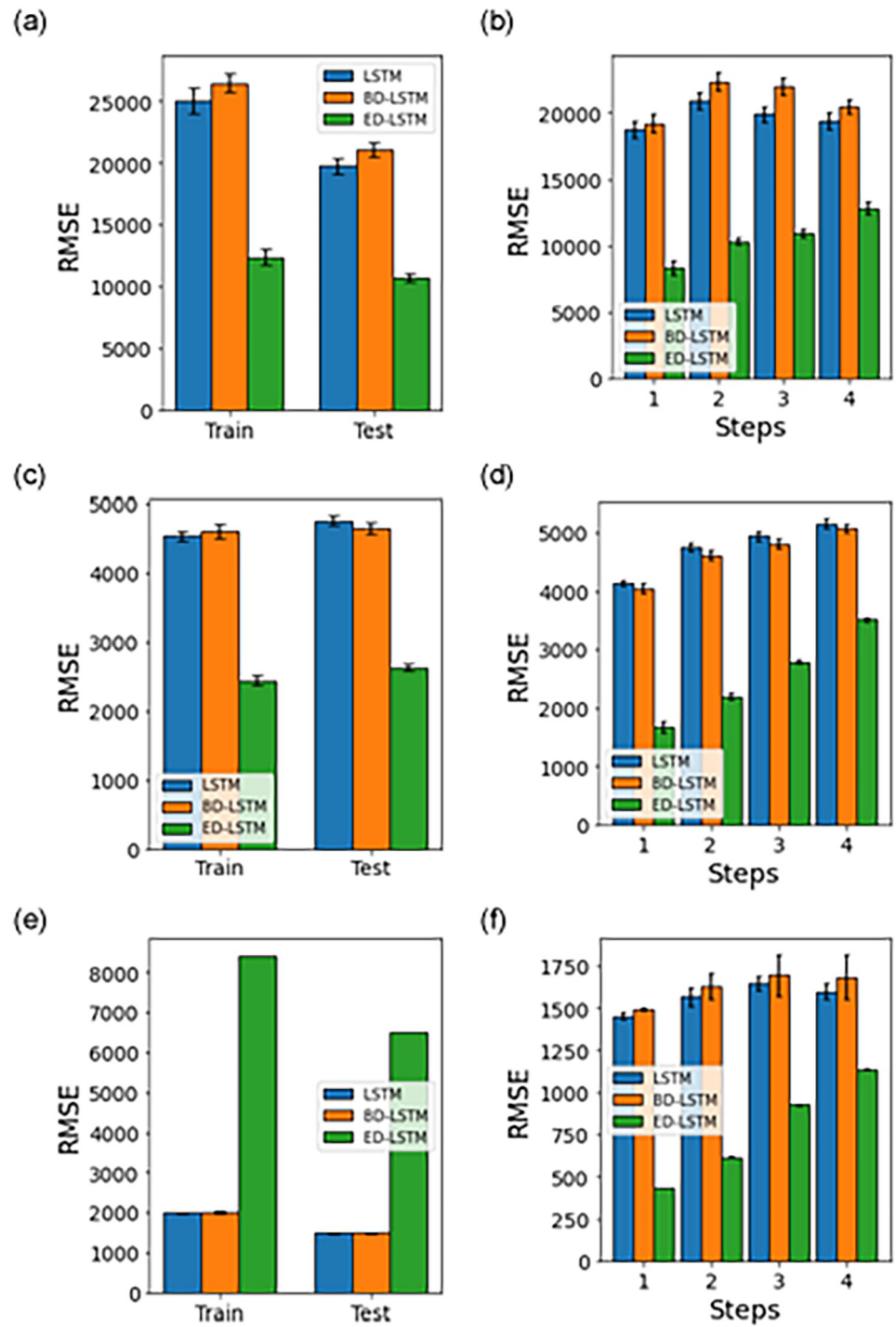
https://doi.org/10.1371/journal.pone.0262708.g007

Fig 8. Univariate LSTM, BD-LSTM and ED-LSTM model performance accuracy (RMSE) with a random-split for the train and test datasets, and the test prediction horizons (steps). The error bars represent the mean and 95% confidence interval for 30 experiment runs.

https://doi.org/10.1371/journal.pone.0262708.g008

different trend and better accuracy (with lower RMSE) where ED-LSTM provides the best test accuracy. In general, we find that random-split with ED-LSTM provides the best performance accuracy for the given univariate models.

Fig 9 shows results for the multivariate approach, where we use the same methods used in the univariate approach (LSTM, ED-LSTM, BD-LSTM). We find that ED-LSTM model gives best performance for the test datasets for all the respective datasets. Fig 10 shows results for the case of random shuffling of train/test dataset using the respective methods. We notice that ED-LSTM at times, provides slightly worse performance when compared to BD-LSTM. Moreover, the performance is not as good when compared to static-split (Fig 9) and in general, we find that ED-LSTM with static-split provides the best performance accuracy.

Tables 7 and 8 provide a summary of results in terms of the test dataset performance accuracy (RMSE) by the respective models for random-split and static-split, which have been given in Figs 7–10. In the univariate models, ED-LSTM provides the best performance accuracy across most of the three different datasets while LSTM performs best only for a single case (Fig 7, Panel a). In multivariate models, BD-LSTM and ED-LSTM provide the best performance accuracy for most cases while LSTM performs best only for a single case (Fig 10, Panel a).

Next, we select two univariate recursive models using random-split for the three datasets to provide a two months outlook for COVID-19 daily infections. In this approach, we use the predictions using the test dataset and extend it further for two months (October and November 2021), recursively. Fig 11 presents results for univariate LSTM and BD-LSTM models. The uncertainty (95% confidence interval shaded in green) and mean prediction is shown in solid black line for 30 experiment runs. We notice that there is a trend of general decline in cases and we also find that the LSTM models well capture the spike and fall in cases every few days.

5 Discussion

The COVID-19 pandemic in India was hit by two major peaks with one in May-October 2020 and the other more deadly in April-June 2021. Surprisingly, the first Indian peak in new cases was reached around the time when the government began lifting nationwide lockdown and focused more on state-level and hot-spot based lock downs [85, 86]; however, there were strict restrictions, such as maintaining social distance and use of face-masks [87]. The second wave struck due to multiple factors and highly-infectious variant-of-concern, also known as SARS-CoV-2 delta variant [23]. A lack of preparation by the authorities in setting up temporary hospitals, shortage of resources such as oxygen and poor management of lockdowns led to major rise of the cases.

There are number of challenges in COVID-19 forecasting due to the nature of the infections, reporting of cases, and effect of lock downs. Nevertheless, despite the challenges and given limited dataset, we have been successful in developing LSTM models for forecasting trend of daily new cases. Our long-term forecasts for two months (October and November 2021) show a steady decline in new cases in India in the respective states. We find that Delhi's two monthly forecasts provide (Fig 11) more uncertainty when compared to Maharashtra and India datasets. We also notice that there is similar level of uncertainty by LSTM and BD-LSTM models for India and Delhi datasets. The major reason the uncertainties are different when you compare Delhi and Maharashtra with rest of India is due to the difference in the trend of cases. In Delhi [15], multiple peaks were observed since the first wave of infections in 2020, whereas in the India dataset, there were only two major peaks. In Maharashtra, a minor peak was observed in November 2021 (Fig 4, Panel a) after the first wave of infections. Hence, the tends captured in the training dataset were relatively different from the states (Maharashtra and Delhi), when compared to India dataset. This suggests that the predictions for
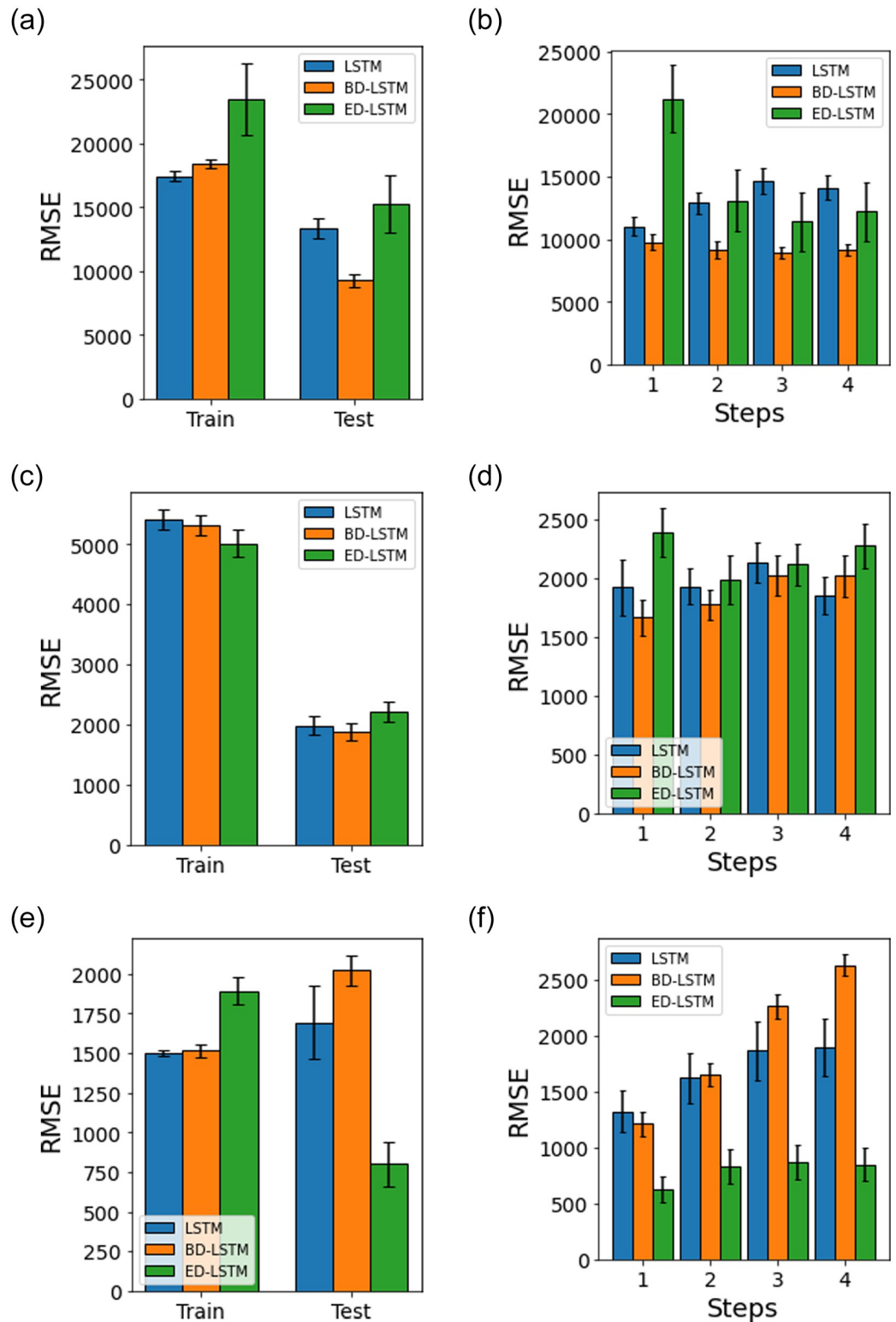
Fig 9. Multivariate model performance accuracy (RMSE) using LSTM, BD-LSTM and ED-LSTM with a static-split for the train and test datasets, and test prediction horizon (steps). The error bars represent the mean and 95% confidence interval for 30 experiment runs.

https://doi.org/10.1371/journal.pone.0262708.g009
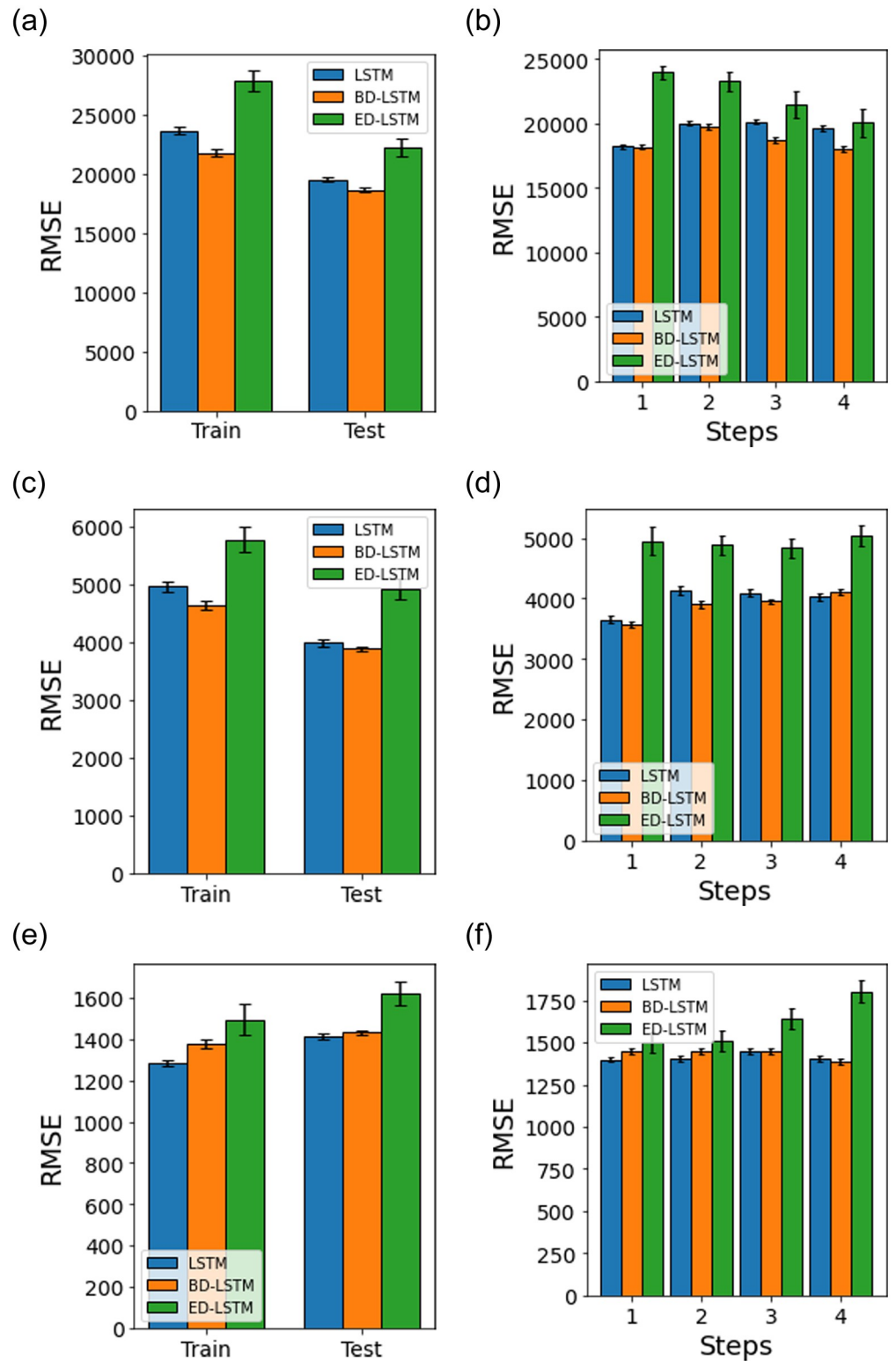
Fig 10. Multivariate model performance accuracy (RMSE) using LSTM, BD-LSTM and ED-LSTM with a random-split for the train and test datasets, and test prediction horizons (steps). The error bars represent the mean and 95% confidence interval for 30 experiment runs.

https://doi.org/10.1371/journal.pone.0262708.g010

Table 7. Univariate model performance accuracy on test dataset (RMSE mean and standard deviation for 30 experimental runs across 4 prediction horizons).

Model	India				Delhi				Maharashtra			
	Random Split		Static Split		Random Split		Static Split		Random Split		Static Split	
	RMSE	Std. Dev.	RMSE	Std. Dev.	RMSE	Std. Dev.	RMSE	Std. Dev.	RMSE	Std. Dev.	RMSE	Std. Dev.
LSTM	19734	1703	35540	6728	1568	34	266	30	4762	195	1925	160
BD-LSTM	21058	1746	37948	6301	1628	75	242	22	4653	214	1748	195
ED-LSTM	10732	1038	62595	8932	825	2	435	185	2365	146	1594	543

https://doi.org/10.1371/journal.pone.0262708.t007

Maharashtra and Delhi are less certain since they had multiple peaks and outbreaks in the past. The second wave of infections in 2021 began in Maharashtra [88, 89] which has been the state that held one of the most number of COVID-19 cases and novel daily infections (Tables 1–4) [15].

The model uncertainty increases due to the limitations in the dataset and models. Our framework has been limited in capturing social-cultural aspects, population density, and level of lockdowns due to missing information and data. Moreover, inter-state travels and the chaotic nature of spread of COVID-19 infections makes it increasingly harder to provide reliable long-term forecasts. In order to improve forecasting results, the models need to incorporate more features in the data. The model needs to capture features such as travel behaviour, level of lockdowns, compliance in masks and other restrictions, social and cultural lifestyle, local area population density, work and income thresholds, state-wise vaccination rate, and accessibility to information.

We take into account the population density as five Indian cities make the top 50 mostly densely populated cities in the world [90], where Mumbai ranks 5th and Delhi the 40th. The impact of COVID-19 on Indian gross domestic product (GDP) is significant, but not as bad when compared to some of the developed western nations [91, 92]. One of the most crucial aspect of management of spread of COVID-19 infections is the role the government played in timely closing their international borders and enforcing lock-downs to various degrees. We need to note that different countries have different geographical and population dynamics, such as population density and culture. It is not a good idea to compare cities given difference in population density although the overall population may be similar. Overall, it is also important to look at cultural factors such as rituals [93], and role of nuclear and extended families [94]. In countries such as India, there is large portion of inter-state migrant workers [95] and also a large portion of the population is in rural areas [96] that also have extended families. These factors made further challenges in containing the spread of COVID-19 infections and are hard to be captured by computational and mathematical models.

In future work, it is important to incorporate robust uncertainty quantification in collection and sampling data model training and model parameters; hence, Bayesian deep learning framework for COVID-19 forecasts would be needed [97–100]. Moreover, ensemble-based

Table 8. Multivariate model prediction accuracy on the test dataset (RMSE mean and standard deviation for 30 experimental runs across 4 prediction horizons).

Model	India				Delhi				Maharashtra			
	Random Split		Static Split		Random Split		Static Split		Random Split		Static Split	
	RMSE	Std. Dev.	RMSE	Std. Dev.	RMSE	Std. Dev.	RMSE	Std. Dev.	RMSE	Std. Dev.	RMSE	Std. Dev.
LSTM	19524	496	13325	2145	1413	38	1693	650	3981	162	1983	414
BD-LSTM	18677	521	9271	1438	1449	43	2021	259	3888	108	1886	389
ED-LSTM	22274	2174	15250	6356	1621	163	801	394	4925	503	2211	478

https://doi.org/10.1371/journal.pone.0262708.t008
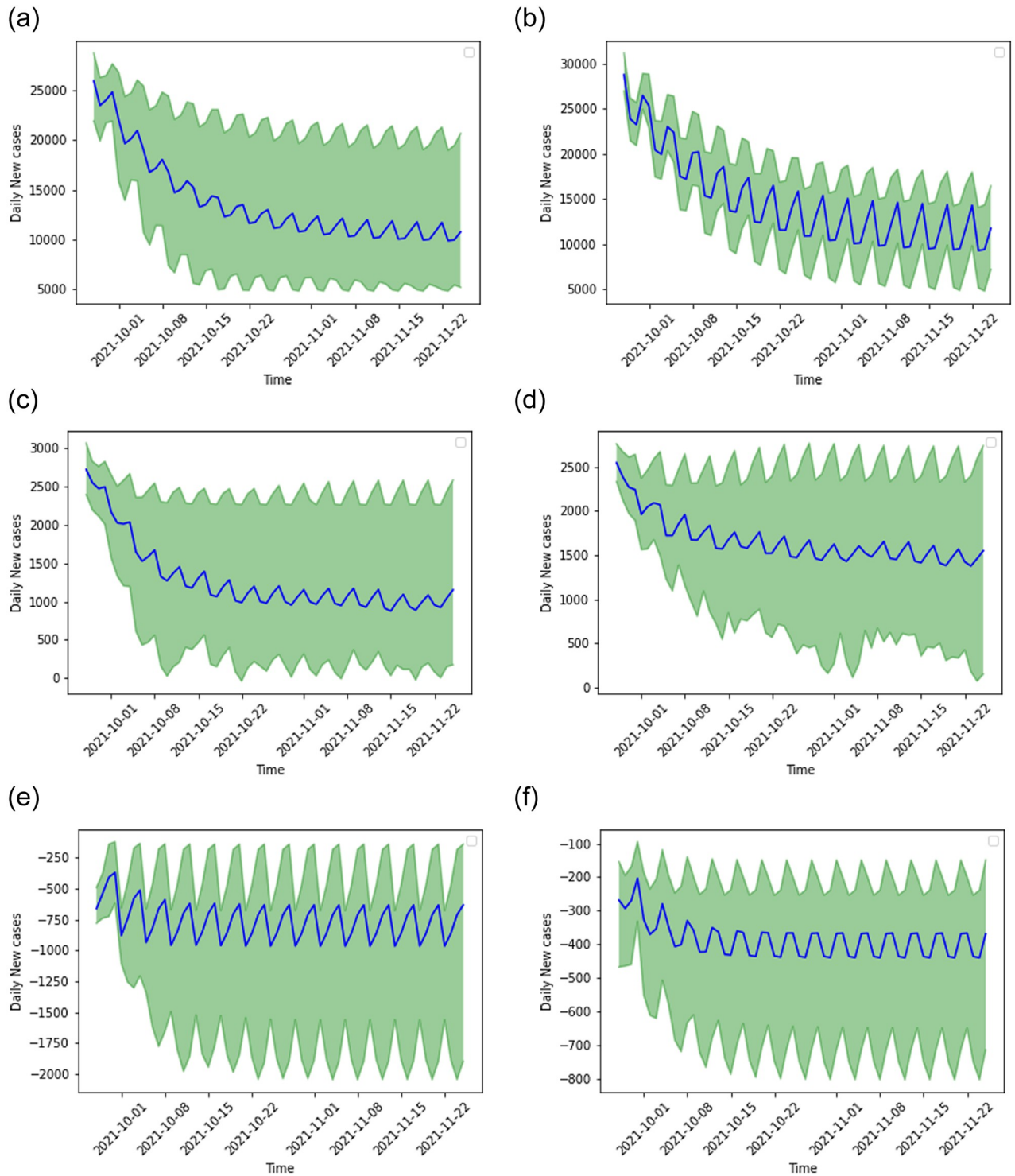
Fig 11. Recursive univariate LSTM and BD-LSTM models predictions for next 60 days (October and November 2021). The uncertainty (95% confidence interval shaded in green) and mean prediction is shown in solid black line.

https://doi.org/10.1371/journal.pone.0262708.g011

learning methods can be used to combine the different types of LSTM models used in this study. We could also develop similar models for death rate and other trends related to COVID-19. Moreover, deep learning models could be used to jointly model the rise and fall of cases and the effect it has on the economy.

6 Conclusions

We presented a framework for employing LSTM-based models for COVID-19 daily novel infection forecasting for India. Our research incorporated some of the latest and most prominent forecasting tools via deep learning, and highlighted the challenges given limited data and the nature of the spread of infections.

Our results show the challenges of forecasting given limited data which is highly biased given that we have two major peaks when considering the pandemic in India. We found that the India and Maharashtra datasets had similar trend in novel cases and model performance. We evaluated univariate and multivariate LSTM-based models with different ways of creating training and test data. The LSTM model variants showed certain strengths and limitations in different scenarios that made it difficult to choose a single model. Generally, we found that the univariate random-split ED-LSTM model provides the best test performance in comparison to rest of the models. Therefore, the data from the adjacent states did not have much effect in the multivariate model since it could not outperform the univariate model. The two months ahead forecast showed a general decline in new cases; however, the authorities need to be vigilant.

Author Contributions

Conceptualization: Rohitash Chandra.

Investigation: Rohitash Chandra, Ayush Jain.

Methodology: Rohitash Chandra, Ayush Jain, Divyanshu Singh Chauhan.

Software: Ayush Jain.

Supervision: Rohitash Chandra.

Validation: Divyanshu Singh Chauhan.

Visualization: Ayush Jain.

Writing – original draft: Rohitash Chandra.

Writing – review & editing: Rohitash Chandra, Ayush Jain, Divyanshu Singh Chauhan.

References

1. Gorbalenya A. E., Baker S. C., Baric R. S., de Groot R. J., Drosten C., Gulyaeva A. A., et al. "The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2," *Nature Microbiology*, vol. 5, no. 4, p. 536, 2020. https://doi.org/10.1038/s41564-020-0695-z

2. Monteil V., Kwon H., Prado P., Hagelkrüys A., Wimmer R. A., Stahl M., et al. "Inhibition of SARS-CoV-2 infections in engineered human tissues using clinical-grade soluble human ACE2," *Cell*, vol. 181, no. 4, pp. 905–913, 2020. https://doi.org/10.1016/j.cell.2020.04.004 PMID: 32333836

3. WHO, "Coronavirus disease 2019 (COVID-19): situation report, 73," pp. 1–13, 2020. [Online]. Available: https://apps.who.int/iris/handle/10665/331686

4. Cucinotta D. and Vanelli M., "WHO declares COVID-19 a pandemic." *Acta bio-medica: Atenei Parmensis*, vol. 91, no. 1, pp. 157–160, 2020. https://doi.org/10.23750/abm.v91i1.9397 PMID: 32191675

5. Andersen K. G., Rambaut A., Lipkin W. I., Holmes E. C., and Garry R. F., "The proximal origin of SARS-CoV-2," *Nature medicine*, vol. 26, no. 4, pp. 450–452, 2020. https://doi.org/10.1038/s41591-020-0820-9 PMID: 32284615

6. A. Atkeson, "What will be the economic impact of COVID-19 in the US? Rough estimates of disease scenarios," National Bureau of Economic Research, Tech. Rep., 2020.

7. N. Fernandes, "Economic effects of coronavirus outbreak (COVID-19) on the world economy," *Available at SSRN 3557504*, 2020.

8. Maliszewska M., Mattoo A., and Dominique V. D. M., "The potential impact of COVID-19 on GDP and trade: A preliminary assessment," *The World Bank*, vol. 9211, pp. 1–26, 2020.

9. C. E. Hart, D. J. Hayes, K. L. Jacobs, L. L. Schulz, and J. M. Crespi, "The impact of COVID-19 on Iowa's corn, soybean, ethanol, pork, and beef sectors," *Center for Agricultural and Rural Development, Iowa State University. CARD Policy Brief*, 2020.

10. Siche R., "What is the impact of COVID-19 disease on agriculture?" *Scientia Agropecuaria*, vol. 11, no. 1, pp. 3–6, 2020. https://doi.org/10.17268/sci.agropecu.2020.01.00

11. Liem A., Wang C., Wariyanti Y., Latkin C. A., and Hall B. J., "The neglected health of international migrant workers in the COVID-19 epidemic," *The Lancet Psychiatry*, vol. 7, no. 4, p. e20, 2020. https://doi.org/10.1016/S2215-0366(20)30076-6 PMID: 32085842

12. Kluge H. H. P., Jakab Z., Bartovic J., D'Anna V., and Severoni S., "Refugee and migrant health in the COVID-19 response," *The Lancet*, vol. 395, no. 10232, pp. 1237–1239, 2020. https://doi.org/10.1016/S0140-6736(20)30791-1 PMID: 32243777

13. Lancet T., "India under COVID-19 lockdown," *Lancet*, vol. 395, no. 10233, p. 1315, 2020.

14. H. Ritchie, E. Mathieu, L. Rodes-Guirao, C. Appel, C. Giattino, E. Ortiz-Ospina, et al, "Coronavirus pandemic (COVID-19)," *Our World in Data*, 2020, https://ourworldindata.org/coronavirus.

15. Dong E., Du H., and Gardner L., "An interactive web-based dashboard to track COVID-19 in real time," *The Lancet infectious diseases*, vol. 20, no. 5, pp. 533–534, 2020. https://doi.org/10.1016/S1473-3099(20)30120-1 PMID: 32087114

16. Johns Hopkins Coronavirus Resource Center (CRC), "COVID-19 Dashboard, Center for Systems Science and Engineering (CSSE), Johns Hopkins University." [Online]. Available: https://coronavirus.jhu.edu/map.html

17. Callard F. and Perego E., "How and why patients made long covid," *Social Science & Medicine*, vol. 268, p. 113426, 2021. https://doi.org/10.1016/j.socscimed.2020.113426 PMID: 33199035

18. Mahase E., "Covid-19: What do we know about "long covid"?" *BMJ*, vol. 370, 2020. PMID: 32665317

19. Jain V. K., Iyengar K. P., and Vaishya R., "Differences between first wave and second wave of COVID-19 in India," *Diabetes & Metabolic Syndrome*, 2021. https://doi.org/10.1016/j.dsx.2021.05.009 PMID: 33992554

20. Arias Velásquez R. M. and Mejía Lara J. V., "Forecast and evaluation of COVID-19 spreading in USA with reduced-space gaussian process regression," *Chaos, Solitons & Fractals*, vol. 136, p. 109924, 2020. https://doi.org/10.1016/j.chaos.2020.109924 PMID: 32501372

21. Yousaf M., Zahir S., Riaz M., Hussain S. M., and Shah K., "Statistical analysis of forecasting COVID-19 for upcoming month in Pakistan," *Chaos, Solitons & Fractals*, vol. 138, p. 109926, 2020. https://doi.org/10.1016/j.chaos.2020.109926

22. Saba A. I. and Elsheikh A. H., "Forecasting the prevalence of COVID-19 outbreak in egypt using non-linear autoregressive artificial neural networks," *Process Safety and Environmental Protection*, vol. 141, pp. 1–8, 2020. https://doi.org/10.1016/j.psep.2020.05.029 PMID: 32501368

23. Del Rio C., Malani P. N., and Omer S. B., "Confronting the delta variant of SARS-CoV-2, summer 2021," *JAMA*, vol. 326, no. 11, pp. 1001–1002, 2021. https://doi.org/10.1001/jama.2021.14811

24. Ren H., Zhao L., Zhang A., Song L., Liao Y., Lu W., et al. "Early forecasting of the potential risk zones of COVID-19 in China's megacities," *Science of The Total Environment*, vol. 729, p. 138995, 2020. https://doi.org/10.1016/j.scitotenv.2020.138995 PMID: 32353723

25. Chin V., Samia N. I., Marchant R., Rosen O., Ioannidis J. P., Tanner M. A., et al. "A case study in model failure? COVID-19 daily deaths and ICU bed utilisation predictions in New York state," *European Journal of Epidemiology*, vol. 35, no. 8, pp. 733–742, 2020. https://doi.org/10.1007/s10654-020-00669-6 PMID: 32780189

26. da Silva R. G., Ribeiro M. H. D. M., Mariani V. C., and dos Santos Coelho L., "Forecasting Brazilian and American COVID-19 cases based on artificial intelligence coupled with climatic exogenous variables," *Chaos, Solitons & Fractals*, vol. 139, p. 110027, 2020. https://doi.org/10.1016/j.chaos.2020.110027 PMID: 32834591

27. Chakraborty T. and Ghosh I., "Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis," *Chaos, Solitons & Fractals*, vol. 135, p. 109850, 2020. https://doi.org/10.1016/j.chaos.2020.109850 PMID: 32355424

28. Maleki M., Mahmoudi M. R., Wraith D., and Pho K.-H., "Time series modelling to forecast the confirmed and recovered cases of COVID-19," *Travel Medicine and Infectious Disease*, p. 101742, 2020. https://doi.org/10.1016/j.tmaid.2020.101742

29. Fauci A. S., Lane H. C., and Redfield R. R., "Covid-19—navigating the uncharted," *New England Journal of Medicine*, vol. 382, no. 13, pp. 1268–1269, 2020. https://doi.org/10.1056/NEJMe2002387 PMID: 32109011

30. Ye F., Xu S., Rong Z., Xu R., Liu X., Deng P., et al. "Delivery of infection from asymptomatic carriers of COVID-19 in a familial cluster," *International Journal of Infectious Diseases*, vol. 94, pp. 133–138, 2020. https://doi.org/10.1016/j.ijid.2020.03.042

31. Bai Y., Yao L., Wei T., Tian F., Jin D.-Y., Chen L., et al. "Presumed asymptomatic carrier transmission of COVID-19," *JAMA*, vol. 323, no. 14, pp. 1406–1407, 2020. https://doi.org/10.1001/jama.2020.2565 PMID: 32083643

32. Anderson R. M., Heesterbeek H., Klinkenberg D., and Hollingsworth T. D., "How will country-based mitigation measures influence the course of the COVID-19 epidemic?" *The Lancet*, vol. 395, no. 10228, pp. 931–934, 2020. https://doi.org/10.1016/S0140-6736(20)30567-5 PMID: 32164834

33. Elman J. L. and Zipser D., "Learning the hidden structure of speech," *The Journal of the Acoustical Society of America*, vol. 83, no. 4, pp. 1615–1626, 1988. https://doi.org/10.1121/1.395916

34. Werbos P. J., "Backpropagation through time: what it does and how to do it," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990. https://doi.org/10.1109/5.58337

35. Hochreiter S. and Schmidhuber J., "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997. https://doi.org/10.1162/neco.1997.9.8.1735 PMID: 9377276

36. Schmidhuber J., "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015. https://doi.org/10.1016/j.neunet.2014.09.003 PMID: 25462637

37. Connor J. T., Martin R. D., and Atlas L. E., "Recurrent neural networks and robust time series prediction," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 240–254, 1994. https://doi.org/10.1109/72.279188 PMID: 18267794

38. Omlin C. W., Thornber K. K., and Giles C. L., "Fuzzy finite state automata can be deterministically encoded into recurrent neural networks," *IEEE Trans. Fuzzy Syst.*, vol. 6, pp. 76–89, 1998. https://doi.org/10.1109/91.660809

39. C. W. Omlin and C. L. Giles, "Training second-order recurrent neural networks using hints," in *Proceedings of the Ninth International Conference on Machine Learning*. Morgan Kaufmann, 1992, pp. 363–368.

40. Giles C. L., Omlin C., and Thornber K. K., "Equivalence in knowledge representation: Automata, recurrent neural networks, and dynamical fuzzy systems," *Proceedings of the IEEE*, vol. 87, no. 9, pp. 1623–1640, 1999. https://doi.org/10.1109/5.784244

41. Hochreiter S., "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty*, *Fuzziness and Knowledge-Based Systems*, vol. 6, no. 02, pp. 107–116, 1998. https://doi.org/10.1142/S0218488598000094

42. Bengio Y., Simard P., Frasconi P. et al., "Learning long-term dependencies with gradient descent is difficult," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 157–166, 1994. https://doi.org/10.1109/72.279181 PMID: 18267787

43. Yang Z., Zeng Z., Wang K., Wong S.-S., Liang W., Zanin M., et al. "Modified seir and ai prediction of the epidemics trend of COVID-19 in china under public health interventions," *Journal of Thoracic Disease*, vol. 12, no. 3, p. 165, 2020. https://doi.org/10.21037/jtd.2020.02.64 PMID: 32274081

44. Chimmula V. K. R. and Zhang L., "Time series forecasting of COVID-19 transmission in canada using lstm networks," *Chaos*, *Solitons & Fractals*, vol. 135, p. 109864, 2020. https://doi.org/10.1016/j.chaos.2020.109864 PMID: 32390691

45. S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," in *Advances in neural information processing systems*, 2015, pp. 802–810.

46. Wang H.-z., Li G.-q., Wang G.-b., Peng J.-c., Jiang H., and Liu Y.-t., "Deep learning based ensemble approach for probabilistic wind power forecasting," *Applied energy*, vol. 188, pp. 56–70, 2017. https://doi.org/10.1016/j.apenergy.2016.11.111

47. Owusu-Fordjour C., Koomson C., and Hanson D., "The impact of COVID-19 on learning-the perspective of the Ghanaian student," *European Journal of Education Studies*, vol. 7, no. 3, pp. 88–101, 2020.

48. Ting D. S. W., Carin L., Dzau V., and Wong T. Y., "Digital technology and COVID-19," *Nature medicine*, vol. 26, no. 4, pp. 459–461, 2020. https://doi.org/10.1038/s41591-020-0824-5 PMID: 32284618

49. Biavardi N. G., "Being an Italian medical student during the COVID-19 outbreak," *International Journal of Medical Students*, vol. 8, no. 1, pp. 49–50, 2020. https://doi.org/10.5195/ijms.2020.489

50. Leite H., Hodgkinson I. R., and Gruber T., "New development:'Healing at a distance'—telemedicine and COVID-19," *Public Money & Management*, pp. 1–3, 2020.

51. Zhou C., Su F., Pei T., Zhang A., Du Y., Luo B., et al. "COVID-19: Challenges to GIS with big data," *Geography and Sustainability*, vol. 1, no. 1, pp. 77–87, 2020. https://doi.org/10.1016/j.geosus.2020.03.005

52. Zambrano-Monserrate M. A., Ruano M. A., and Sanchez-Alcalde L., "Indirect effects of COVID-19 on the environment," *Science of the Total Environment*, vol. 728, p. 138813, 2020. https://doi.org/10.1016/j.scitotenv.2020.138813 PMID: 32334159

53. Muhammad S., Long X., and Salman M., "COVID-19 pandemic and environmental pollution: a blessing in disguise?" *Science of The Total Environment*, vol. 728, p. 138820, 2020. https://doi.org/10.1016/j.scitotenv.2020.138820 PMID: 32334164

54. Kerimray A., Baimatova N., Ibragimova O. P., Bukenov B., Kenessov B., Plotitsyn P., et al. "Assessing air quality changes in large cities during COVID-19 lockdowns: The impacts of traffic-free urban conditions in almaty, kazakhstan," *Science of The Total Environment*, vol. 730, p. 139179, 2020. https://doi.org/10.1016/j.scitotenv.2020.139179 PMID: 32387822

55. Millett G. A., Jones A. T., Benkeser D., Baral S., Mercer L., Beyrer C., et al. "Assessing differential impacts of COVID-19 on Black communities," *Annals of Epidemiology*, 2020. https://doi.org/10.1016/j.annepidem.2020.05.003 PMID: 32419766

56. Rajkumar R. P., "COVID-19 and mental health: A review of the existing literature," *Asian journal of psychiatry*, vol. 52, p. 102066, 2020. https://doi.org/10.1016/j.ajp.2020.102066 PMID: 32302935

57. Gao J., Zheng P., Jia Y., Chen H., Mao Y., Chen S., et al. "Mental health problems and social media exposure during COVID-19 outbreak," *PLOS One*, vol. 15, no. 4, p. e0231924, 2020. https://doi.org/10.1371/journal.pone.0231924 PMID: 32298385

58. Huang Y., Wu Y., and Zhang W., "Comprehensive identification and isolation policies have effectively suppressed the spread of COVID-19," *Chaos*, *Solitons & Fractals*, vol. 139, p. 110041, 2020. https://doi.org/10.1016/j.chaos.2020.110041 PMID: 32834599

59. Alzahrani S. I., Aljamaan I. A., and Al-Fakih E. A., "Forecasting the spread of the COVID-19 pandemic in saudi arabia using arima prediction model under current public health interventions," *Journal of Infection and Public Health*, vol. 13, no. 7, pp. 914–919, 2020. https://doi.org/10.1016/j.jiph.2020.06.001 PMID: 32546438

60. Singh S., Parmar K. S., Kumar J., and Makkhan S. J. S., "Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (ARIMA) models in application to one month forecast the casualties cases of COVID-19," *Chaos*, *Solitons & Fractals*, vol. 135, p. 109866, 2020. https://doi.org/10.1016/j.chaos.2020.109866 PMID: 32395038

61. Anand A., Lamba Y., and Roy A., "Forecasting COVID-19 transmission in India using deep learning models," *Letters in Applied NanoBioScience*, vol. 10, no. 2, pp. 2044–2055.

62. K. R. Bhimala, G. K. PATRA, R. Mopuri, and S. R. Mutheneni, "A deep learning approach for prediction of SARS-CoV-2 cases using the weather factors in India," *Authorea Preprints*, 2020. [Online]. Available: https://doi.org/10.22541/au.160275979.91541585/v1

63. Shetty R. P. and Pai P. S., "Forecasting of COVID 19 cases in Karnataka state using artificial neural network (ANN)," *Journal of The Institution of Engineers (India): Series B*, pp. 1–11, 2021.

64. Tomar A. and Gupta N., "Prediction for the spread of COVID-19 in India and effectiveness of preventive measures," *Science of The Total Environment*, vol. 728, p. 138762, 2020. https://doi.org/10.1016/j.scitotenv.2020.138762 PMID: 32334157

65. Gupta A. K., Singh V., Mathur P., and Travieso-Gonzalez C. M., "Prediction of COVID-19 pandemic measuring criteria using support vector machine, prophet and linear regression models in Indian scenario," *Journal of Interdisciplinary Mathematics*, pp. 1–20, 2020.

66. S. Bodapati, H. Bandarupally, and M. Trupthi, "COVID-19 time series forecasting of daily cases, deaths caused and recovered cases using long short term memory networks," in *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*. IEEE, 2020, pp. 525–530.

67. Chaurasia V. and Pal S., "Application of machine learning time series analysis for prediction COVID-19 pandemic," *Research on Biomedical Engineering*, pp. 1–13, 2020.

68. Battineni G., Chintalapudi N., and Amenta F., "Forecasting of COVID-19 epidemic size in four high hitting nations (USA, Brazil, India and Russia) by Fb-Prophet machine learning model," *Applied Computing and Informatics*, pp. 1–10, 2020.

69. P. Nadler, R. Arcucci, and Y. Guo, "A neural sir model for global forecasting," in *Machine Learning for Health*. PMLR, 2020, pp. 254–266.

70. O. Istaiteh, T. Owais, N. Al-Madi, and S. Abu-Soud, "Machine learning approaches for covid-19 fore-casting," in *2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*. IEEE, 2020, pp. 50–57.

71. Pinter G., Felde I., Mosavi A., Ghamisi P., and Gloaguen R., "COVID-19 pandemic prediction for hun-gary; a hybrid machine learning approach," *Mathematics*, vol. 8, no. 6, 2020. [Online]. Available: https://doi.org/10.3390/math8060890

72. F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, War-wick 1980*, ser. Lecture Notes in Mathematics, 1981, pp. 366–381.

73. Frazier C. and Kockelman K., "Chaos theory and transportation systems: Instructive example," *Trans-portation Research Record: Journal of the Transportation Research Board*, vol. 20, pp. 9–17, 2004. https://doi.org/10.3141/1897-02

74. Elman J. L., "Finding structure in time," *Cognitive Science*, vol. 14, pp. 179–211, 1990. https://doi.org/10.1207/s15516709cog1402_1

75. Omlin C. W. and Giles C. L., "Constructing deterministic finite-state automata in recurrent neural net-works," *J. ACM*, vol. 43, no. 6, pp. 937–972, 1996. https://doi.org/10.1145/235809.235811

76. Chandra R., "Competition and collaboration in cooperative coevolution of Elman recurrent neural net-works for time-series prediction," *Neural Networks and Learning Systems, IEEE Transactions on*, vol. 26, pp. 3123–3136, 2015. https://doi.org/10.1109/TNNLS.2015.2404823 PMID: 25769175

77. Schuster M. and Paliwal K., "Bidirectional recurrent neural networks," *Signal Processing, IEEE Trans-actions on*, vol. 45, pp. 2673–2681, 12 1997. https://doi.org/10.1109/78.650093

78. Graves A. and Schmidhuber J., "Framewise phoneme classification with bidirectional lstm and other neural network architectures," *Neural networks: the official journal of the International Neural Network Society*, vol. 18, pp. 602–10, 07 2005. https://doi.org/10.1016/j.neunet.2005.06.042 PMID: 16112549

79. Y. Fan, Y. Qian, F.-L. Xie, and F. K. Soong, "Tts synthesis with bidirectional lstm based recurrent neu-ral networks," in *INTERSPEECH*, 2014.

80. A. Graves, N. Jaitly, and A.-r. Mohamed, "Hybrid speech recognition with deep bidirectional lstm," in *2013 IEEE workshop on automatic speech recognition and understanding*. IEEE, 2013, pp. 273–278.

81. I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds., 2014, pp. 3104–3112.

82. "Unique identification authority of India," May 2020, [Online; accessed 16-July-2020]. [Online]. Avail-able: https://uidai.gov.in/images/state-wise-aadhaar-saturation.pdf

83. S. Athreya, N. Gadhiwala, and A. Mishra, "COVID-19 India-Timeline an understanding across states and union territories." 2020, ongoing Study at http://www.isibang.ac.in/∼athreya/incovid19.

84. "COVID-19 India data," 2020, Ministry of Health and Welfare, Government of India: https://www.mohfw.gov.in/.

85. Debnath R. and Bardhan R., "India nudges to contain COVID-19 pandemic: A reactive public policy analysis using machine-learning based topic modelling," *PLOS One*, vol. 15, no. 9, pp. 1–25, 09 2020. https://doi.org/10.1371/journal.pone.0238972 PMID: 32915899

86. Rai B., Shukla A., and Dwivedi L. K., "Dynamics of COVID-19 in India: A review of different phases of lockdown," *Population Medicine*, vol. 2, no. July, 2020. https://doi.org/10.18332/popmed/125064

87. Rab S., Javaid M., Haleem A., and Vaishya R., "Face masks are new normal after COVID-19 pan-demic," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 6, pp. 1617–1619, 2020. https://doi.org/10.1016/j.dsx.2020.08.021 PMID: 32889402

88. Joshi M., Kumar M., Srivastava V., Kumar D., Rathore D., Pandit R., et al. "First detection of SARS-CoV-2 Delta variant (B. 1.617. 2) in the wastewater of (Ahmedabad), India," *medRxiv*, 2021. [Online]. Available: https://doi.org/10.1101/2021.07.07.21260142

89. Yang W. and Shaman J., "COVID-19 pandemic dynamics in India and impact of the SARS-CoV-2 Delta (B. 1.617. 2) variant," *medRxiv*, 2021. [Online]. Available: https://doi.org/10.1101/2021.06.21.21259268 PMID: 34845460

90. Wikipedia contributors, "List of cities proper by population density, Wikipedia," 2020, [Online; accessed 16-July-2020]. [Online]. Available: https://en.wikipedia.org/wiki/List_of_cities_proper_by_population_density

91. S. M. Dev, R. Sengupta et al., "COVID-19: impact on the indian economy," *Indira Gandhi Institute of Development Research, Mumbai Working Papers*, April, 2020. [Online]. Available: https://ideas.repec.org/p/ind/igiwpp/2020-013.html

92. "The World Economic Outlook April 2020: The Great Lockdown, International Monetary Fund," April 2020, [Online; accessed 16-July-2020]. [Online]. Available: https://www.imf.org/en/Publications/WEO/Issues/2020/04/14/weo-april-2020

93. Imber-Black E., "Rituals in the time of COVID-19: imagination, responsiveness, and the human spirit," *Family process*, vol. 59, no. 3, pp. 912–921, 2020. https://doi.org/10.1111/famp.12581 PMID: 32663322

94. Lebow J. L., "Family in the age of COVID-19," *Family process*, vol. 59, no. 2, p. 309–312, June 2020. [Online]. Available: https://europepmc.org/articles/PMC7273068 PMID: 32412686

95. Dandekar A. and Ghai R., "Migration and reverse migration in the age of covid-19," *Economic and Political Weekly*, vol. 55, no. 19, pp. 28–31, 2020.

96. Kumar A., Nayar K. R., and Koya S. F., "COVID-19: challenges and its consequences for rural health care in india," *Public Health in Practice*, p. 100009, 2020. https://doi.org/10.1016/j.puhip.2020.100009 PMID: 34171041

97. R. M. Neal, "Bayesian learning via stochastic dynamics," in *Advances in Neural Information Processing Systems 5*. Morgan-Kaufmann, 1993, pp. 475–482.

98. Pall J., Chandra R., Azam D., Salles T., Webster J. M., Scalzo R., et al. "Bayesreef: A bayesian inference framework for modelling reef growth in response to environmental change and biological dynamics," *Environmental Modelling & Software*, p. 104610, 2020. https://doi.org/10.1016/j.envsoft.2019.104610

99. Chandra R., Jain K., Deo R. V., and Cripps S., "Langevin-gradient parallel tempering for Bayesian neural learning," *Neurocomputing*, vol. 359, pp. 315–326, 2019. https://doi.org/10.1016/j.neucom.2019.05.082

100. Chandra R. and Kapoor A., "Bayesian neural multi-source transfer learning," *Neurocomputing*, vol. 378, pp. 54–64, 2020. https://doi.org/10.1016/j.neucom.2019.10.042