

ξ boost: An AI-Based Data Analytics Scheme for COVID-19 Prediction and Economy Boosting

Darshan Vekaria, Aparna Kumari¹, Member, IEEE, Sudeep Tanwar², Member, IEEE,
and Neeraj Kumar³, Senior Member, IEEE

Abstract—The coronavirus (COVID-19) outbreak has a significant impact on people’s lives, occupations, businesses, and economies globally. The world economic market is experiencing a big shift and the share market has observed crashes day-by-day. Even, the Indian economy has witnessed a slowdown in the current pandemic, and recovery of it is quite difficult. The restrictions and restrain strategies (e.g., lockdown and social distancing) introduced by the government leave many professions and facilities in a dormant state, catalyzing economy downfall. It necessitates to improve economy along with control strategies of COVID-19, which is a challenging task. To handle the above-mentioned issues, this article proposes a novel economy-boosting scheme, i.e., ξ boost, which is a fusion of artificial intelligence (AI) and big data analytics (BDA) integrated with the Internet-of-Things (IoT)-based data communication. Here, a bidirectional long short-term memory (LSTM) model is anticipated for early prediction of total positive cases as well as the economy. Then, it calculates an optimal subsegment of days, in which trade and commerce related restrictions could be reduced to control a sharp decline in the economy. Next, a spark-based pre and post unlock (PPU) analytics is carried out on the rise of COVID-19 cases to validate the intensity of testing in the country and deciding economy-boosting activities. Then, the ξ boost scheme is evaluated based on various factors such as prediction accuracy and others while comparing to existing approaches. It facilitates healthy and profitable smart cities by the means to control pandemic with subsequent economy rise.

Index Terms—Artificial intelligence (AI), bidirectional long short-term memory (LSTM), big data analytics (BDA), coronavirus (COVID-19), economy prediction, Internet of Things (IoT), spark.

I. INTRODUCTION

THE CORONAVIRUS (COVID-19) global pandemic has significantly impacted the health and lifestyle of people. Governments are enforced to implement lockdowns; social distancing and work from home policies. Aiming to save

Manuscript received September 28, 2020; revised December 5, 2020; accepted December 23, 2020. Date of publication December 25, 2020; date of current version October 22, 2021. (Corresponding author: Neeraj Kumar.)

Darshan Vekaria, Aparna Kumari, and Sudeep Tanwar are with the Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad 382481, India (e-mail: 16bit042@nirmauni.ac.in; 18ftphde22@nirmauni.ac.in; sudeep.tanwar@nirmauni.ac.in).

Neeraj Kumar is with the Thapar Institute of Engineering and Technology, Patiala 147004, India, also with the Department of Informatics, King Abdulaziz University, Jeddah 21589, Saudi Arabia, also with the Department of Computer Science and Information Engineering, Asia University, Taichung 41354, Taiwan, and also with School of Computing, University of Petroleum and Energy Studies, Dehradun 248007, India (e-mail: neeraj.kumar@thapar.edu).

Digital Object Identifier 10.1109/JIOT.2020.3047539

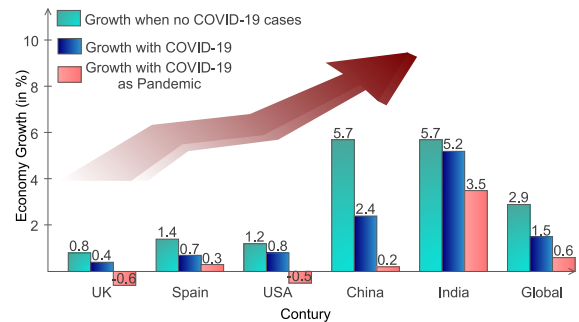


Fig. 1. Economy growth in COVID-19 outbreak.

the lives of the citizens, several steps such as travel restrictions, social distancing, quarantines, scrapping of large scale events have been incorporated since the beginning of 2020 [1]. Initially originated from the Wuhan city of China, the catastrophic effects of this outbreak has hit the economy of several developed and developing countries like the USA, India, and more. Fig. 1 shows the economic growth in different countries by considering the COVID-19 rise on different scales [2].

The world-wide economy is predicted to diminish by approximately 3% in the year 2020, which could be considered as one of the most impulsive downfalls since the 1930 [3]. The rise of infection has severely affected various micro, small, and medium scaled enterprises such as tourism & hospitality, aviation, automobile, real estate, and textiles, which raises global economic concerns [4]. However, to handle such issues, AI is a viable solution, as it efficiently involves finding a correlation, prediction, and connection among data. AI has played a crucial role earlier in quantitative finance and price predictions [5]. It has been foreseen that the GDP of the UK will rise up to 10.3% higher with the use of AI techniques [6]. AI facilitates to increase economic growth by bolstering the improvement of productivity and products, and the stimulation of new companies [7].

The COVID-19 data is collected in different forms such as text, images, and others, across the world and is used to study and analyze this disease [8]. It could be categorized as *Big Data* due to its nature and rate of generation, which necessitates big data analytics (BDA) for decision making to control the spread [9]. Fig. 2 shows the significance of BDA in terms of COVID-19, based on different characteristics (5 V’s) of big data, as follows.

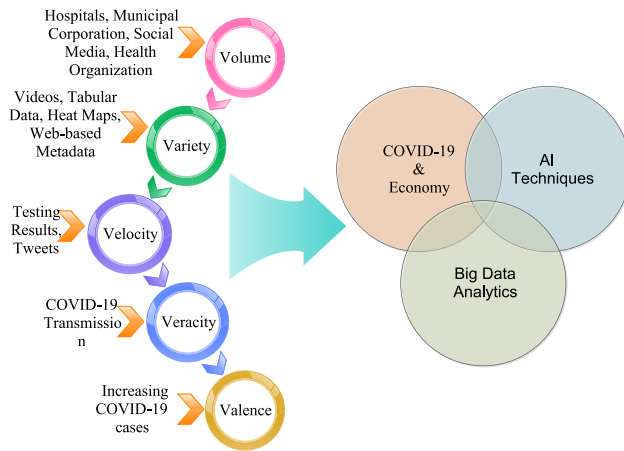


Fig. 2. Significance of big data in COVID-19 scenario.

- 1) *Volume*: It corresponds to the scale and size of the data. Currently, COVID-19 data is collected on an everyday basis through hospitals, municipal corporations, health organizations, and social media platforms. The accumulation of it is resulting in large volumes of data. Therefore, it is required to use big data-oriented frameworks like Hadoop or Spark to analyze it.
- 2) *Variety*: It corresponds to different forms of data and its complexity. The COVID-19 data is available in different structures and formats, for example, a structured tabular data set can show the number of cases in a specific country while a pictorial heat map of the world can show the same in the form of an image. Furthermore, videos are another source of COVID-19 data as they are used for classification and segmentation in AI-based approaches.
- 3) *Velocity*: It is the pace at which data is generated and speed at which it moves from one point to the next [10]. Here, every hour, information is added and updated on social media and government websites in the form of tweets, the location of infected persons, and the count of cases, recoveries, and demises. Thus, COVID-19 data corresponds to the aspects of velocity.
- 4) *Veracity*: It refers to the validity or volatility of COVID-19 data. During analytics, some information is valid and some turn out to be volatile. The government reports and daily counts are quite valid as they are generated based on the number of tests done. Different reports from World Health Organization (WHO) and other reputed organizations, related to the spread of this virus are changing very frequently in this outbreak. For instance, it was initially claimed that COVID-19 is transmitted through a touch of surface only. Later developments suggested that it is an air-borne disease. The data regarding the symptoms observed in the patients is also updating over time.
- 5) *Valence*: It refers to the connectedness of the data with each other, from different data sets [11]. For example, a video-based analysis of some state/region exhibiting social distancing violation is in accordance with the analytics of increasing cases on a structured tabular data set.

Massive amount of COVID-19 data is generated in a variety of forms and at a rapid rate, which can be crucial in deducing the impact of the outbreak on economic growth [12]. So, it requires intelligent data-driven decision making [13] and strategy planning to address this issue [14]. Hence, BDA plays a crucial role in understanding the trends of disease spread [15] envisioned with AI techniques. Moreover, the government cannot shut the whole nation for long to control the pandemic as it has a severe impact on the country's economy. The problem becomes more challenging as the solution should consider a revival of the economy with constraints of COVID-19 rise.

The ongoing/recent research, as described in the state-of-the-art section, have used several concepts of data collection, segmentation, detection for prevention and diagnosis of COVID-19, using various prediction techniques [16], [17] like mathematical, statistical, and ML models. But, they have not been explored to its full potential by predicting the economy and the total positive cases (individuals affected by the disease) in this pandemic and using the correlation between them for data-driven analysis [18]. Motivated from these dynamics, this article presents a scheme termed as $\xi boost$, an AI-based solution to predict COVID-19 cases and economy along with spark-based BDA for the economic development of healthy smart cities.

The analysis have been proposed by considering their practical applicability. $\xi boost$ scheme addresses the concerns for communication and data storage by suggesting wired and wireless connections and the use of cloud data servers. As COVID-19 data is categorized as big data, the scheme suggests to carry out deep learning (DL) computations on BDA frameworks like Spark. The governing bodies and health officials of the country are expected to be the enablers of this scheme, as it can provide them with COVID-19 and economy predictions to obtain optimal subsegments of days, through subsegment picker (SSP) algorithm to plan economy-boosting activities in the country.

A. Research Contributions

Following are the research contributions of this article.

- 1) Propose a novel $\xi boost$ scheme, which is a fusion of AI technique and BDA, providing long short-term memory (LSTM)-based predictions of COVID-19 cases and economy, which supplements the data-driven decisions made by the SSP algorithm for economy-boosting.
- 2) Design an SSP Algorithm to select the optimal subsegment of days for carrying out trade and commerce to boost the economy.
- 3) Spark-based daywise PPU analytics to co-relate it with the intensity of COVID-19 testing.
- 4) Performance evaluation of $\xi boost$ scheme compared to state-of-the-art approaches in terms of prediction accuracy.

B. Organization of This Article

The remainder of this article is organized as follows. Section II describes the state-of-the-art approaches, their critical analysis, and how the $\xi boost$ scheme provides solutions in areas which are yet untouched. Section III delineates the proposed architecture of the $\xi boost$ scheme, its components,

TABLE I
ABBREVIATIONS USED

Symbols	Description	Symbols	Description
ξ boost	Economy Boost	SSP	Sub Segment Picker
LSTM	Long Short Term Memory	RMSE	Root Mean Square Error
PPU	Pre and Post Unlock	MAPE	Mean Absolute Percentage Error
WHO	World Health Organisation	COVID	Coronavirus Disease
SEIR	Susceptible-Exposed-Infectious-Removed	GDP	Gross Domestic Product
LP	Linear Interpolation	MLP	Multi-layered Perceptron
ANFIS	Adaptive Network-based fuzzy inference system	INR	Indian rupee
RIC	Rate of Increase in Cases	RIR	Rate of Increase in Recovery
RID	Rate of Increase in Deaths	AI	Artificial Intelligence
ML	Machine Learning	DL	Deep Learning
CDS	Cloud Data Server	BDA	Big Data Analytics
USD	United States Dollar	$Pred$	Predictions
β	Predicted Economy	α	Predicted COVID-19 cases
χ	Economy Boosting Activities	ϕ	Economy Threshold Function
Ω	Economy Threshold Value	κ	Sub Segment Separator Function
ω	Set of optimal Sub Segment of days	λ	New Economy after using SSP
RNN	Recurrent Neural Network	DC^T	Transposed Data
DN	Data Normalisation	DI	Integrated Data
WLC	Weighted Limit Calculator	Lm	Weighted Limit

and usage in practical scenario. In Section IV, the workflow of the proposed scheme is described, which includes the flow of information and corresponding economy-boosting analysis. Next, in Section V, the experimental results, data sets, data-driven algorithms (SSP and PPU), and their outcomes are discussed, along with comparative analysis with other baseline approaches. Finally, we conclude this article in Section VI. Table I shows the abbreviations used in this article.

II. STATE OF THE ART

This section highlights the concepts and issues related to COVID-19 and exiting work to mitigate it. Several research works are on-going across the globe to analyze the trends of the pandemic, its future prediction, and approaches to reduce its spread, which are described in this section.

Various mathematical and statistical models have been presented for outbreak trend analytics [16], [19], [20]. For example, Mandal *et al.* [21] formulated a mathematical model to mitigate disease transmission. This approach uses a time-dependent variables like susceptible, exposed, hospitalized, infected, quarantined, and recovered populations to form an autonomous system using first-order differential equations. Similarly, Yang *et al.* [22] presented a susceptible-exposed-infectious-removed (SEIR) model to derive the epidemic curve in china. They predicted a peak in COVID-19 spread by late February and a gradual decline by the end of April month. In addition, Ahmar and del Val [23], used the SutteARIMA technique to forecast COVID-19 and Spanish stock market value. Even though they forecasted it with a mean absolute percentage error (MAPE) value of 3.6%, they provided only two days of predictions, which is not suitable for any practical applications, as long term predictions are expected for formulating prevention strategies [24], [25]. However, the mathematical models are prone to being an unfit approach for dynamic scenarios. The mathematical models are formulated linearly along with several assumptions. Given the static nature of mathematical

models [26], they can not be changed or tuned once they are modelled, and thus, are prone to inconsistent results.

On the other side, a statistical and epidemiological model analyses the epidemic's curve, prevalence, and its expected lifetime. However, the COVID-19 data is time series and dynamic in nature, which requires the AI technique that is viable to extract the trends from it. Also, statistical and epidemiological models are prone to produce unreliable results. The AI-based machine learning (ML) and DL approach [27] have proved to be a good fit to combat these issues.

Here, the system learns automatically based on the previous computations, over preprocessed data. These techniques have been highly coveted for forecasting and analytics purpose. For example, Ardabili *et al.* [28] proposed forecasting of COVID-19 in Italy, China, Iran, Germany, and the USA through multilayered perceptron (MLP) model and adaptive network-based fuzzy inference system (ANFIS) model. They tabulated results for each country, showing root mean-square error (RMSE) and correlation coefficient value for each of them against consideration of different numbers of neurons for the models. The optimal value for the number of neurons was different for each country according to the rate of transmission of disease. Gupta *et al.* [29] proposed a weather data-driven case which described the relation between temperature and COVID-19 cases, based on the disease spread in the U.S. They showed that for the Indian scenario, there would be a decrease in the number of cases recorded during summers, but that didn't happen. Then, Tuli *et al.* [30] proposed a robust Weibull model, which is based on iterative weighing that outperforms iterative fitting over other distributions like Gaussian, Beta, Fisher-Tippet, and Log normal. Apart from ML, some DL-based models such as neural networks, LSTM, have been developed to predict COVID-19 trends.

Tomar and Gupta [31] proposed an LSTM-based COVID-19 total cases prediction for the next 30 days in India. Their predictions have been provided for April and measured per day error in the predictions ranging from -6.44% to 8% . They

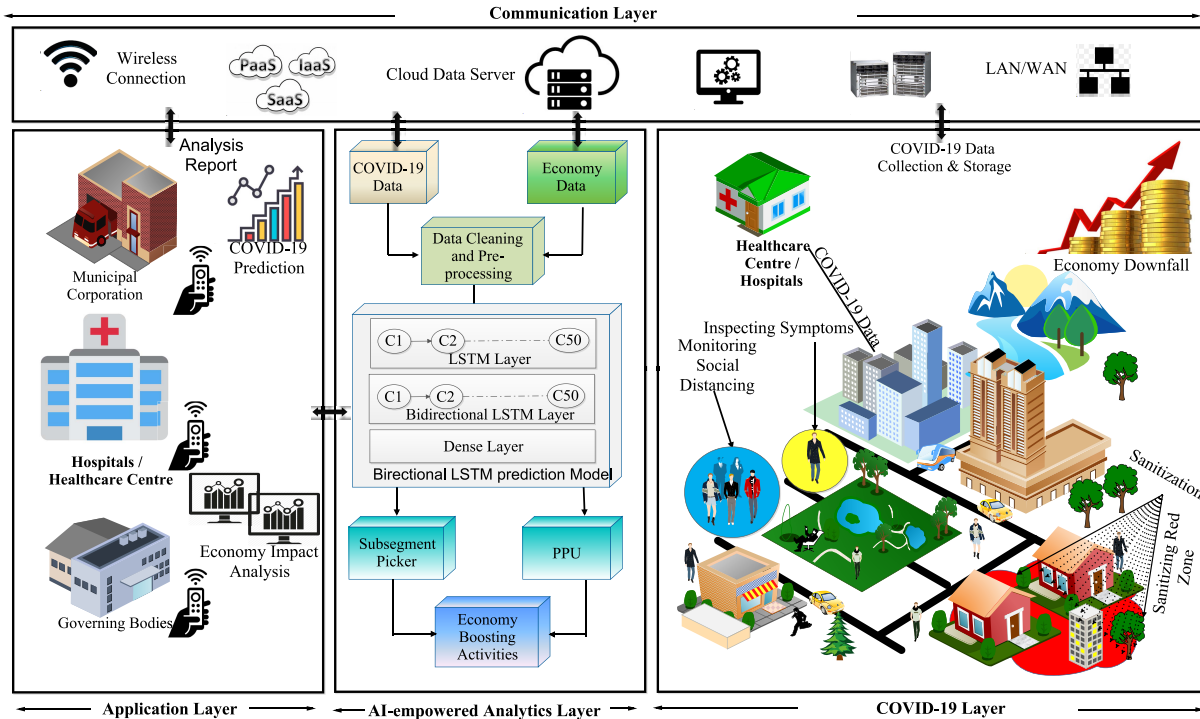


Fig. 3. System architecture of ξ_{boost} scheme.

validated the effects of social isolation and strict lockdown measure by verifying the spread for a transmission rate of $r \in [0.001, 2.3]$. However, apart from predictions, the work does not provide any cogent ideas on how the predictions can be utilized. Likewise, Arora *et al.* [32] presented an LSTM-based model for COVID-19 trend predictions. They carried out a comparative analysis of deep LSTM, convolutional LSTM, and bidirectional LSTM. Their best results obtained by the bidirectional LSTM model, predict the cases for 32 States/Union territories within India with an average MAPE value of 3.22%.

From the above discussion, it has been evident that long term prediction of COVID-19 trend with minimal error needs to be explored (with a considerable size of training data). Motivated by these factors, this article presents a novel IoT-oriented AI-based scheme, ξ_{boost} , to predict COVID-19 trend and its impact on the economy. As per our knowledge, no research work has been done so far, which has addressed the economic crisis of India amid the outbreak. The ξ_{boost} scheme is very crucial for the government as it provides an SSP algorithm to minimize the economic downfall by providing optimal subsegments of days to waive off lockdown restrictions and promote activities for an economic boost in a specific country. Furthermore, a spark-based PPU analysis (a feedback system) can give important feedback to the government regarding the development of COVID-19 and its correlation with the intensity of testing.

III. SYSTEM ARCHITECTURE AND PROBLEM FORMULATION

This section discusses the architecture of the proposed ξ_{boost} scheme along with problem formulation.

A. System Architecture

Fig. 3 shows the system architecture of the ξ_{boost} scheme. It is divided into four different layers, namely, 1) COVID-19 layer; 2) AI-empowered analytics layer; 3) application layer; and 4) communication layer, which are explained as follows.

1) *COVID-19 Layer:* The COVID-19 layer is focused on relevant data collection. WHO has also acknowledged the different means of spread of this disease in humans [33]. Thus, human-to-human interaction transmits the disease and expands it at an exponential rate. The data is collected from several sources like CCTV footage for social distancing, symptom identification from infected people, IoT-based body temperature sensors, IoT-based health monitoring devices, number of zones covered by municipal corporation in sanitization drives, and daily dynamics of the economy from the finance ministry. Different sources provide crucial data, which is stored on the cloud server [34]. From this layer, using wireless (e.g., ZigBee and WiFi) and wired (like LAN/WAN) connections, the complex communications are handled and historical/real-time data about COVID-19 and economy is forwarded or fetched from the cloud data server through communication layer. This data can be accessed by government officials, researchers, citizens as per security constraints or the platform at which it is released.

2) *AI-Empowered Analytics Layer:* In this layer, big data is collected from the COVID-19 layer to produce analytics for further decision making. As shown in Fig. 3, everyday COVID-19 cases and daily economy data are taken for analysis. As there might be many inconsistencies in this data, we furnish the quality of it by carrying out data cleaning

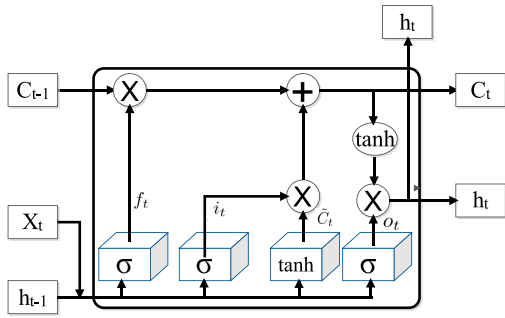


Fig. 4. LSTM cell.

and preprocessing, which includes data transformation and normalization.

After carrying out these activities, we obtain preprocessed data. It is required to predict the future trends of the disease as well as the economy, for decision making and future strategy planning, to develop healthy smart cities. Therefore, the proposed AI-based ξ boost scheme, comprises of a novel bidirectional LSTM model and BDA based on Spark. The data is inputted to the bidirectional LSTM model to obtain predicted values *Pred*. Fig. 3 shows that the prediction model is made of one LSTM layer, one bidirectional LSTM layer and one dense layer. Each of the LSTM layer consists of cells, as represented in Fig. 4

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \quad (2)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (3)$$

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \quad (4)$$

$$h_t = o_t * \tanh(C_t). \quad (5)$$

The above equations show the main operations of an LSTM cell [35]. Here, i is the input gate, which decides which information should be passed in the current timestamp, f is the forget gate which discards irrelevant data from previous timestamp, o is the output gate which controls the flow of information in the network. Moreover, W represents the weight matrices while σ is the sigmoid activation function. x represents the input, b is the biases, h is the output and c is the cell memory at time t .

Furthermore, PPU analyzer and SSP are proposed to make a productive analysis of the forecasts. As the Indian government declares different phases of lockdowns with restrictions, PPU is a pre and post unlock analyzer which shows the daywise development of COVID-19 cases, demises, and recoveries across different phases and can validate the intensity of testing on daily basis.

SSP is a subsegment picker, which picks multiple subsegments of days optimally for relaxing the lockdown restrictions. The resulting subsegments of days and the feedback obtained from PPU are used in carrying out economy-boosting activities such as trade, businesses, working of public facilities, etc., throughout the country. Next, this layer accepts the COVID-19 and economy data for time series forecasting through the communication layer. The data-driven decisions made by SSP and PPU components of this layer are stored in the cloud data

server for the governing bodies, so that they can implement the same in smart cities.

3) *Application Layer*: This layer is responsible to implement the feedback and results obtained from the AI-empowered Analytics Layer. The daily predictions of economy value $\{\beta_1, \beta_2, \dots, \beta_d\} \in \beta$ for d days are used by SSP to give the subsegment of days in which economy-boosting activities should be carried out by the government and municipal corporation along with complying protocols of social distancing and sanitization. This will ensure that economy is being boosted by resuming work for the given subsegment of days along with precautions that a drastic increase in COVID-19 cases should not happen. The PPU analysis can be used by the hospitals and healthcare centers to ensure that testing of COVID-19 is being carried out at full strength. PPU feedback can suggest that the current amount of testing is sufficient according to the predicted rise $\{\alpha_1, \alpha_2, \dots, \alpha_d\} \in \alpha$ in the COVID-19 cases for d days.

The decision made at the application layer will be the key aspects for developing healthy smart cities within the country. Through satellite communication, this layer broadcasts the decisions to the media platforms to notify the citizens about the relaxations and restrictions related to COVID-19. While, through wireless mobile communication, the governing bodies at this layer provides instruction to municipal corporations to implement the restrictions. The new decisions are stored on the cloud storage at communication layer for a record.

4) *Communication Layer*: The communication layer is equipped with various forms of wired and wireless communication such as LAN, WAN, MAN, ZigBee, WiFi, satellite and mobile communication. Each layer interacts with another layer through a data transfer from communication layer. The structure is like a Star topology, with the communication layer being the mediator for every flow of information. This helps in achieving centralized management of the network.

The readings of body temperature and medical checks being carried out by the sensors are collected through IoT-based communication. The cloud-based data server *CDS* has all the data stored and produced across all the layers. The *CDS* is updated in specific intervals when any new COVID-19 data is reported, new IoT-based sensor readings are found, or any new development regarding the after-effects of an outbreak (like economy drop and others) are noticed.

The *CDS* stores these data so that it can be used by the AI-empowered analytics layer to make a forecast on total positive cases and the economy. The results of SSP and PPU, are also stored and passed on to the application layer for strategizing the economy-boosting activities.

B. Problem Formulation

Let $\{d_1, d_2, \dots, d_n\} \in d$ be the set of n days, for which $\{\alpha_1, \alpha_2, \dots, \alpha_n\} \in \alpha$ are the corresponding COVID-19 cases and $\{\beta_1, \beta_2, \dots, \beta_n\} \in \beta$ are the corresponding economy values. Considering the Indian scenario, the economy value is considered in Indian Rupee (INR) equivalent to \$1. Let χ be the set of economy-boosting activities, such as trade, construction, agriculture, sales and services, etc. Let ϕ be the economy

threshold function, which can define an amount, above which, economy value is considered to be on a downside. The economy threshold Ω is defined as follows:

$$\Omega = \phi(\alpha, \beta) \quad (6)$$

where the functioning of ϕ is governed by the constraint: $\beta \propto \alpha$, i.e., for deciding the threshold, the value of β on the day d_i is weighted in proportion to the value of α on the same day. Let κ be the subsegment separator function, which can separate the optimal subsegment of days from the set d while taking into consideration the values in set β and the threshold value Ω . Let ω define the set of an optimal subsegment of days

$$\omega = \kappa(d, \beta, \Omega). \quad (7)$$

Carrying out economy-boosting activities χ will improve the economy, i.e., it will reduce the conversion value of 1 USD to INR, thus making the Indian economy value/prestige higher. Let λ be the set of an expected new economy worth defined by

$$\lambda = \chi(\omega) \quad (8)$$

i.e., the economy worth after carrying out economy-boosting activities χ over the Ω subsegment of days. In view of the above discussion, the objective function of $\xi boost$ can be defined as follows:

$$\varphi \leftarrow \text{MAXIMIZE}(\lambda) \quad (9)$$

subject to the constraints

$$\omega \in d \quad (10)$$

$$\alpha > 0, \beta > 0, \Omega > 0 \quad (11)$$

$$\lambda > \beta$$

$$\forall d_i \in \omega$$

$$\text{MIN}(\beta) \leq \Omega \leq \text{MAX}(\beta). \quad (12)$$

IV. $\xi boost$: THE PROPOSED APPROACH

Fig. 5 depicts the workflow of the proposed $\xi boost$ scheme. human-to-human transmission of COVID-19 is increasing at an exponential rate which brings four possibilities among the health status of citizens: 1) being infected; 2) recovered; 3) succumb; and 4) remain noninfected, from the disease. This marks the origin of COVID-19 data, describing the total cases, recoveries and deaths due to COVID-19. The restrictions imposed due to the pandemic brings an economic downfall, the data of which can be obtained by recording daily readings. This comprises the data used in the $\xi boost$ scheme.

Before carrying out any algorithmic computations, data preprocessing is performed to furnish the data for making predictions.

- 1) The collected data of pandemic and economy is preprocessed for data cleaning. Unwanted observations such as duplicate data cells and irrelevant information from the data are removed. Next, we filter out the unwanted outliers that can be observed in both the economy and COVID-19 data. For addressing the problems of null and zero value occurrence, we use linear interpolation LP .

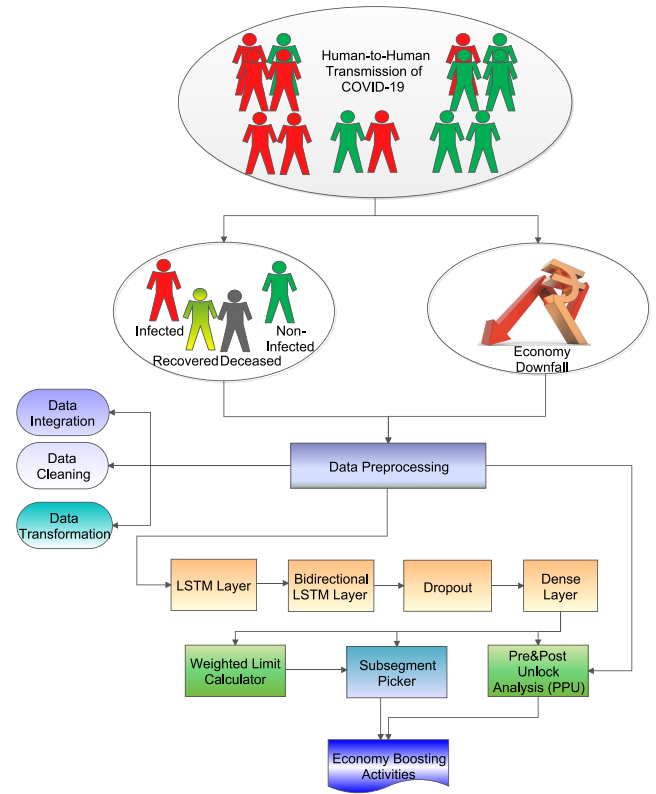


Fig. 5. Workflow of the proposed $\xi boost$ scheme.

- The null and zero values are replaced by the interpolated values and the cleaned data DC is prepared. The major goals of data cleaning were a) to remove errors; b) to make efficient memory use of resources; c) to remove redundancy; d) to increase reliability of the data and its corresponding analysis; and e) to make optimal analysis.
- 2) Then, we pass DC through the *data transformation* (as depicted in Fig. 5) stage to make it structured and of better quality for forecasting in later stages. The disease data sets included the countrywise daily cases, where the rows marked the country and the columns marked the date. Whereas, the data set for the economy was structured such that the rows marked the date and the column marked as economy value correspondingly. So, the disease data set is transposed to DC^T to get its structure in sync with the data set of the economy. Then, DC^T columns are filtered according to the selected country for which predictions are required. Then, the data normalization DN has been performed. The numeric columns of DC^T are normalized to a common scale, without distorting differences in the ranges of values. Data normalization is required since the features, i.e., total positive cases, and economy values for every date, are in different ranges. The data normalization is achieved through min-max normalization [36], which is done as shown in Eq. below, where, A is the attribute of the data, $\min(A)$ and $\max(A)$ is an absolute maximum of A , v is the old value of the data, v' is the new value of that data, and $\text{new_max}(A)$ and $\text{new_min}(A)$ being

Algorithm 1 Algorithm for LSTM Model— ξ boost

Input: Dataset DS for Attribute p , Dropout u , loss function lf , optimizer op , Epochs ep
Output: Predicted Attribute $Pred$
Initialization: From DS Consider list of d days with value of (Economy/Covid-19 cases), $\{k_1, k_2, \dots, k_d\}$. Let XT be the training dataset. Let N_i be the number of nodes in layer i and let RT be the return sequence.

- 1: **procedure** LSTM_DISEASE_AND_ECONOMY_FORECASTING(DS, p, u, lf, op, ep)
- 2: Identify Features from DS
- 3: $DS \leftarrow TRANSPOSE(DS)$
- 4: $TypeConvert(DS.Date, date - time)$
- 5: $TypeConvert(DS.attribute, Float)$
- 6: **for** $i \leftarrow 1$ to n **do**
- 7: $k_i \leftarrow INPUT(DS_d)$
- 8: **end for**
- 9: **for** $i \leftarrow 1$ to n **do**
- 10: Normalize k_i in range (0,1)
- 11: **end for**
- 12: **for** $i \leftarrow 1$ to n **do**
- 13: **if** $k_i[i].IsValid$ **then**
- 14: $k_i \leftarrow k_i.Interpolate()$
- 15: **end if**
- 16: **end for**
- 17: $I_{seq} \leftarrow INITIALIZE_SEQUENTIAL_LSTM_MODEL(DS)$
- 18: $x \leftarrow LSTM(R_i, N_1, RT = True, (XT.shape, 1))$
- 19: $y \leftarrow BiLSTM(N_2, x)$
- 20: $BiLSTM.DROPOUT(u)$
- 21: $I_{seq} \leftarrow Dense_L(y, N_3)$
- 22: $I_{seq}.TRAIN(lf, op, ep)$
- 23: $Pred \leftarrow I_{seq}.TEST(DS)$
- 24: **return** $Pred$
- 25: **end procedure**

the new range of the data ([0, 1] in this case)

$$v' = ((v - \min(A))/(\max(A) - \min(A))) * (\text{new_max}(A) - \text{new_min}(A)) + \text{new_min}(A). \quad (13)$$

- 3) After normalization, the data is integrated for both the outbreak and economy into one dataframe, thus, optimizing the accessibility of the data through a single dataframe. The integrated data DI is passed on to the proposed AI-based scheme incorporated with the bidirectional LSTM model for forecasting of COVID-19 cases and the economy.

The LSTM model includes a bidirectional LSTM layer because conventional neural network model carries out unidirectional processing of the inputs and neglect the developments made in the future [37]. This problem is addressed by the bidirectional-LSTM structure. The neurons in a layer are split into forward state and backward state, such that the i th neuron's forward state passes its computation to that of the $(i + 1)$ th neuron, and the $(i + 1)$ th neuron's backward state passes its computation to that of the i th neuron. Thus, input x_i at the i th neuron is processed to output y_i considering both the forward and backward states of the i th neuron.

Unidirectional LSTM layers process the information of past only and therefore, they might become prone to underfitting as it may not generalize to new data. Bidirectional LSTM layers preserve the information from both, past and future, but, adding more of its layers can cause overfitting, as it may learn the detail and noise in the training data to an extent that it negatively impacts the models ability to generalize. Thus, considering the stochastic nature of both COVID-19 and economy, we use one layer of each, unidirectional and bidirectional LSTM, to make it a good fit for stochastic natured

Algorithm 2 Algorithm for Weighted Limit Calculator

Input: Predicted COVID-19 cases α , Predicted Economy β
Output: Weighted limit for Healthy Economy Lm
Initialization: Let d be the total number of days for which the limit has to be decided. Since the economy value on day i is decided on factors and events taken place on days $i - 1$, so we will correlate cases on day i with the economy value on day $i + 1$.

- 1: **procedure** CALCULATE_WEIGHTED_LIMIT(α, β)
- 2: $\Delta \leftarrow 0$
- 3: **for** $i \leftarrow 1$ to d **do**
- 4: $\Delta \leftarrow \Delta + (\alpha_i) * (\beta_{i+1})$
- 5: **end for**
- 6: $\tau \leftarrow 0$
- 7: **for** $i \leftarrow 1$ to d **do**
- 8: $\tau \leftarrow \tau + \alpha_i$
- 9: **end for**
- 10: **if** $\tau == 0$ or $\Delta == 0$ **then**
- 11: $Lm \leftarrow 0$
- 12: **return** Lm
- 13: **end if**
- 14: $Lm \leftarrow \Delta/\tau$
- 15: **return** Lm
- 16: **end procedure**

data and to discard the possibilities of overfitting and underfitting. Algorithm 1 shows the structure of the LSTM model. In forward pass, weighted sum of the inputs is evaluated. In a backward pass, the errors are calculated and corresponding weights are updated [38]. Therefore, the computational complexity of the LSTM networks is $O(W)$, where W is the number of edges in the network [39] and is calculated as follows:

$$W = 4IH + 4H^2 + 3H + HK \quad (14)$$

where I is the number of inputs, K is the number of outputs and H is the number of cells in the hidden layer. The computations of LSTM layers are processed with the dropout technique. Dropout is required to prevent the model from over-fitting. The dense layer outputs the predictions α and β , i.e., the predicted values of COVID-19 cases and economy, respectively. α and β are passed through weighted limit calculator (WLC), as described in Algorithm 2, to obtain a limit Lm beyond which the economy value is categorized as steeply falling. WLC is an important and inseparable component of this scheme, which is essential for the functioning of SSP algorithm. WLC multiplies the daily economy value and COVID-19 cases and takes the sum of it for given n days. This sum is divided by the total of COVID-19 cases observed in these n days, which gives us the economy limit Lm . The time and space complexity of the WLC algorithm is linear, i.e., $O(d)$, where d is the number of days for which Lm is calculated. The complexity is linear as all operations are done with a for loop iterating upon d days.

β and Lm are inputted in the SSP. It aims to give an optimal subsegment of days in which several activities, occupations, and professions, should be given relaxation from the restrictions given by the government so that they can function normally while taking precautionary measures like using masks, workplace sanitization, and social distancing. If these activities will be given some days in a row to work normally along with taking precautionary measures, it will help in boosting the economy of the country and also restrict the economy on the selected subsegment of days to surpass the limit Lm . The working of SSP is defined in Algorithm 3.

Algorithm 3 Algorithm for SSP

Input: Predicted COVID-19 cases α , Predicted Economy β , Number of days d
Output: Set of Optimal Subsegment of days F
Initialization: Let LS be the list storing initial subsegment of day, $start$ be the starting day of a subsegment and $stop$ be the ending day of a subsegment, $LS \leftarrow []$, $start \leftarrow -1$, $stop \leftarrow -1$

```

1: procedure CALCULATE_SUB_SEGMENTS( $\alpha, \beta, d$ )
2:    $Lm \leftarrow CALCULATE\_WEIGHTED\_LIMIT(\alpha, \beta)$ 
3:   for  $i \leftarrow 0$  to  $d$  do
4:     if  $start == -1$  and  $\beta_i > Lm$  then
5:        $start \leftarrow i$ 
6:     end if
7:     if  $start! = -1$  and  $\beta_i \leq Lm$  then
8:        $stop \leftarrow i - 1$ 
9:        $LS.add(start, stop)$ 
10:       $start \leftarrow -1$ 
11:       $stop \leftarrow -1$ 
12:    end if
13:  end for
14:  if  $start! = -1$  and  $stop == -1$  then
15:     $stop \leftarrow d - 1$ 
16:     $LS.add(start, stop)$ 
17:  end if
18:   $n \leftarrow len(LS)$ 
19:   $F \leftarrow []$ 
20:  for  $i \leftarrow 1$  to  $n$  do
21:    if  $i == n - 1$  then
22:       $F.add(LS[i])$ 
23:      break
24:    end if
25:     $l1 \leftarrow LS[i][1] - LS[i][0]$ 
26:     $l2 \leftarrow LS[i + 1][1] - LS[i + 1][0]$ 
27:     $X \leftarrow LS[i + 1][0] - LS[i][1]$ 
28:     $min \leftarrow MINIMUM(l1, l2)$ 
29:    if  $X \leq min/2$  then
30:       $start = LS[i][0]$ 
31:       $stop = LS[i + 1][1]$ 
32:       $F.add(start, stop)$ 
33:       $i = i + 1$ 
34:      continue
35:    end if
36:    if  $X > min/2$  then
37:       $F.add(LS[i])$ 
38:    end if
39:  end for
40:  return  $F$ 
41: end procedure

```

As all logical operations in this algorithm are achieved with a for loop iterating over d days of economy-boosting, space and time complexity of the SSP algorithm is $O(d)$. The subsegment of days is collected for which the economy value is $\geq Lm$. The activities will benefit in case they are given a higher number of consecutive days for work-related relaxation, as it continues the momentum of work, thus providing an even greater boost to the economy. To maximize the number of days in the selected subsegments, we will merge them with some days which have an economy value less than the Lm . If H_k and H_{k+1} are two adjacent subsegments of days with an economy value higher than Lm , then the subsegment of days between them, X , will have an economy value lower than Lm . If the length of X is less than or equal to half of the minimum length among H_k and H_{k+1} , then we merge the subsegments X , H_k and H_{k+1} , resulting into a new subsegment F , which will be providing more number of days in a row for economy-boosting activities. The optimal merging factor is considered as 0.5 here, since keeping it lower will not allow us to add more number of days, and keeping it higher will add a significantly larger number of days, which will also put people the risk of getting infected, as they will be going out of their houses for more number of days.

Furthermore, α is passed on to PPU to verify uniform distribution for a rise in cases and recovery and validate the intensity of testing from the same. Therefore, the selected subsegment of days from SSP and feedback on testing intensity from PPU will help the government in planning the economy-boosting activities for the next unlock phase.

V. EXPERIMENTAL RESULTS AND DISCUSSION

The results shown in this section are in accordance with the $\xi boost$ architecture proposed in this article. This section includes the results, analysis, and discussions obtained by the implementation of $\xi boost$ scheme. Here, *Experimental setup and Predictions results* section comprises the results of the bidirectional LSTM model for the COVID-19 and economy data. Then the next section *Performance Evaluation and Comparative Analysis* deals with the discussion on the accuracy value of the proposed model and its comparison with the baseline models. Later, the section on *Subsegment Picker Analysis* talks about the results obtained by using the SSP algorithm on the actual economy data from January to June. Finally, *spark-based PPU analysis* section describes the PPU analysis of COVID-19 in India.

A. Experimental Set-Up and Prediction Results

1) *Data Set:* The $\xi boost$ scheme is implemented and tested against the data sets of COVID-19 cases [40] and the daily economy [41] for the Indian scenario. These data sets are highly stochastic in nature as an increase or decrease in the economy and COVID cases depend on several environmental and physical factors. The data set [40] is taken from the Johns Hopkins University Coronavirus Resource center. It consists of time-series data of confirmed COVID-19 cases from January 22, 2020 for 269 Countries/Regions. The economy data set [41] contains time series data of daily Indian economy value with respect to \$1. [40] is split into training data (January 22, 2020 to 28 May 2020) and testing data (May 29, 2020–June 28, 2020) to produce 30 days of predictions of COVID cases. Similarly, to study and forecast the dynamics of the changing economy amid the COVID outbreak in India, we have taken the economy data [41] from January 22, 2020 to June 21, 2020 for training purposes and that of June 22, 2020 to June 28, 2020 for testing to forecast seven days of the economy.

2) *Prediction Results:* For implementation of $\xi boost$ scheme, Python is used to interact with DL libraries. Open Source libraries such as Numpy v1.18.4, Pandas v1.0.4, Keras v2.3.1, and Spark v2.4 are used to perform various computations and analytics. Rigorous hyperparameter tuning of the bidirectional LSTM model has been done along with addressing the idiosyncratic nature of both the data sets. The model incorporates mean squared error loss, which has been brought to the optimal point by using the adam optimizer. The bidirectional LSTM model being implemented here is in accordance with that mentioned in the AI-empowered analytics layer, shown in Fig. 3.

Considering the different nature and features of both the data sets, the hyperparameters of the bidirectional LSTM model used for each of them are different. While training

TABLE II
 ξ boost PREDICTION MODEL

ξ boost Prediction Model Structure			
Attribute	Layer 1	Layer 2	Layer 3
Type	Input - Unidirectional LSTM	Hidden - Bidirectional LSTM	Output - Dense
No. of Nodes	50	50/100	1
Model type	Sequential		
Dropout	25%		
Loss function	Mean Squared Error		
Optimizer	Adam		
Validation Split	20%		

TABLE III
ACCURACY MEASURES USING THE PROPOSED ξ boost SCHEME

Attribute / Measure	MAPE	RMSE
COVID-19	1.27%	6308
Indian Economy	0.12%	0.149

where N is the number of samples, x_i and y_i are the actual and predicted value of the sample, respectively. While

$$MAPE = 100 * \frac{1}{N} \sum_{i=1}^N (|A_i - F_i|/A_i) \quad (16)$$

where N is the number of samples, A_i and F_i are the actual and predicted value of the sample respectively. Table III shows the obtained MAPE and RMSE values for COVID-19 and Indian economy predictions. The COVID-19 data consists values in Lakhs whereas the RMSE score calculated by ξ boost scheme is in some thousand only, while the MAPE error is as low as 1.27%. Similarly for Indian economy results deviate from the actual values by few decimal places only. Moreover, the model also outshines while comparing it with state-of-the-art approaches.

Arora *et al.* [32] (Baseline 1) have predicted the statewide COVID-19 cases for India. After comparing the results of convolutional, stacked, and bidirectional LSTM, their bidirectional LSTM model gives an average MAPE value of 3.22% for prediction of COVID-19 in 32 states. Then, Tomar and Gupta *et al.* [31] (Baseline 2) have also used the LSTM model to predict the COVID-19 cases in India for 30 days. Even though there is no mention of the error percentage for 30 days of prediction, they have given the error reports for 5 days of prediction which averages 4.96%. Fig. 7(c) illustrate the comparison of percentage prediction error of proposed ξ boost scheme with respect to the baseline model [31] and [32], which justifies the better performance of ξ boost scheme over state-of-the-art approaches in terms of prediction accuracy. Table IV summarizes the performance of ξ boost with other schemes, validating that it makes a prediction for higher number of days, with less error and provides practically useful economy boosting SSP algorithm.

C. Subsegment Picker Analysis

It is important to take preventive measures to stop disease transmission, whereas, getting the economy back to a stable state is also equally important. So, ξ boost scheme also include a subsegment picker that considers both the above factors and proposes a subsegment of days in which economy-boosting activities can be carried out. Several basic facilities and businesses like transportation, construction, agriculture, restaurants, shops, import and export of goods, sales and services, manufacturing, etc., could work for a considerable amount of time to support the economy growth of the country as well as to fulfill public needs.

The SSP algorithm and its results shown here are in accordance with Subsegment picker suggested in the System Architecture, shown in Fig. 3. As described in Section IV, SSP resulted in the optimal subsegment of days along with taking

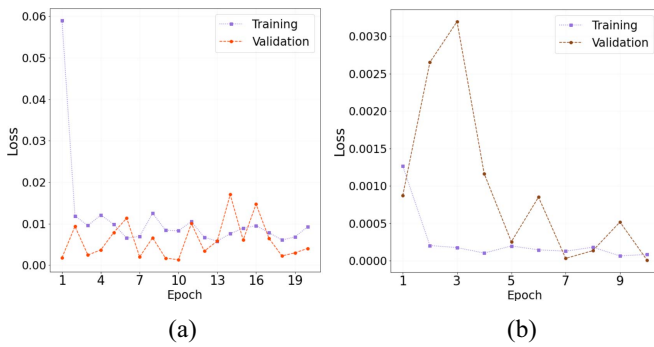


Fig. 6. Bidirectional LSTM model evaluation. (a) Loss comparison for COVID-19 data. (b) Loss comparison for economy data.

the COVID-19 data, the bidirectional LSTM layer uses 50 nodes and the model gets trained over 10 epochs. However, since economy values differ by a small magnitude, the training for economy data involves using the bidirectional LSTM layer of 100 nodes and training it over 20 epochs. Models are trained with a validation split of 20%. Table II summarizes the prediction model structure of ξ boost. Fig. 6(a) and (b) shows the plot of training and validation loss for the bidirectional LSTM models over the COVID-19 and economy data.

The training and validation loss values are very close to each other (with a difference of decimals only), indicating the exceptional prediction capabilities of the ξ boost scheme using bidirectional LSTM. Fig. 7(a) and (b) shows the predictions of COVID-19 cases and Indian economy. Here, we make 30 days of prediction on COVID-19 cases and 7 days of prediction for economy (in INR equivalent of \$1).

B. Performance Evaluation and Comparative Analysis

The prediction results of the proposed ξ boost scheme have been evaluated on the RMSE and MAPE values. Equations (15) and (16) show the mathematical expressions for calculating the RMSE and MAPE value, respectively

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (15)$$

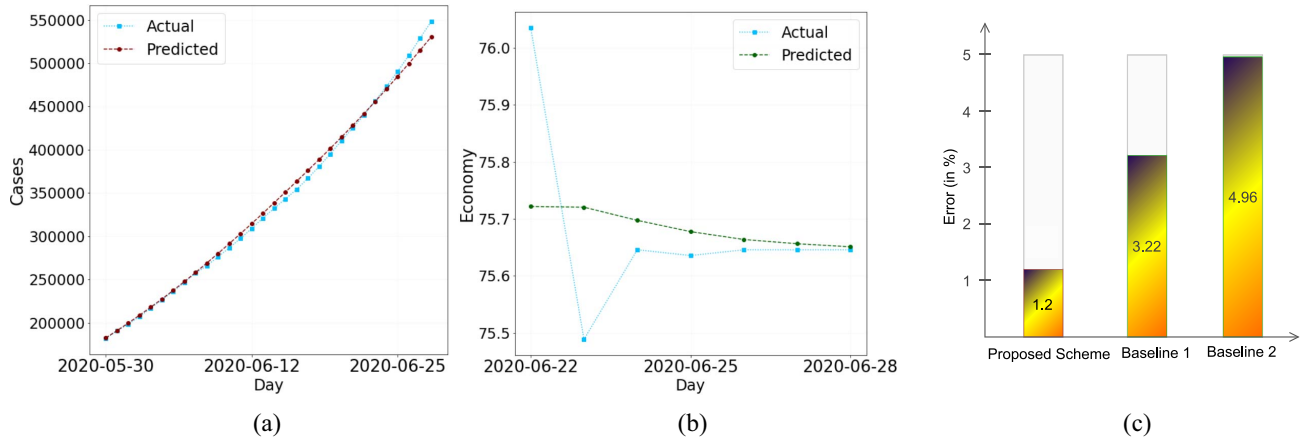


Fig. 7. AI-based COVID-19 cases and economy prediction. (a) COVID-19 cases prediction. (b) Economy prediction. (c) MAPE comparison.

TABLE IV
PERFORMANCE EVALUATION OF ξ boost SCHEME WITH STATE-OF-THE-ART APPROACHES

Reference	Method/Approach	% Error (in RMSE)	No. of days predicted	Data-driven decisions
Tomar <i>et al.</i> [31]	LSTM.	4.96 %	5	Validating transmission rate of the disease amid preventive measures
Arora <i>et al.</i> [32]	Deep, Convolutional and Bidirectional LSTM	3.22 %	7	State wise spread analysis and classification into mild, moderate and severe zone
Chimmula <i>et al.</i> [42]	LSTM	6.2%	28	Trend analysis of COVID-19, estimating the pandemic to end by 2020
ξ boost (Proposed scheme)	LSTM (Unidirectional + Bidirectional layer)	1.2%	30	Subsegment picker analysis for providing optimal days for economy boosting activities

into consideration the facts that 1) activities need a considerable amount of days to carry the momentum of work and be productive towards economy-growth and 2) not many days of relaxation for work should be given as it will expose people to travel and interact with others, thus increasing the risk of getting infected. Taking into account growing COVID-19 cases and the economy, the WLC gives the limit value of the economy. All the subsegments of days for which the forecasted economy crosses the WLC value are noted. We try to maximize the number of days in the selected subsegments. Between any two selected subsegments of days, say $H1$ and $H2$, there will be a subsegment of days X whose economy is less than the limit. To provide more working days for the activities in a row, we merge $H1$, $H2$, and X if the length of X is less than half of the minimum length among $H1$ and $H2$.

Fig. 8 visualize the Indian economy changes for 157 days taken from January to June. The subsegments of days highlighted by the black color are selected by the SSP approach for economy boosting activities. The red-colored horizontal line displays the limit calculated by the WLC. SSP makes sure that there are optimal days allotted for the normal functioning of economy-boosting activities as well as enough days to stick to the restrictions regarding abandonment of the activities by means to control COVID-19. In this way, it can solve the economy concerns in the country in the times of COVID-19.

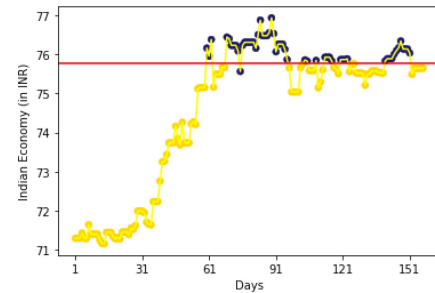


Fig. 8. SSP analysis of ξ boost scheme.

D. Spark-Based PPU Analysis

A spark-based PPU analysis on the daywise rise of total positive cases, recoveries, and demises in India is proposed to study and validate if regular testing is being carried out with maximum potential on a daily basis in the country [43]. Fig. 9 shows the scenario of the Pre Unlock days. As seen from the charts, while rate of increase in deaths (RID) is uniform, the rate of increase in cases (RIC), and rate of increase in recoveries (RIR) is not distributed equally among the seven days. This is because, during the preunlock period, the testing was not uniform and was not rapid during the initial days of the outbreak due to lockdown, as evident from [44].

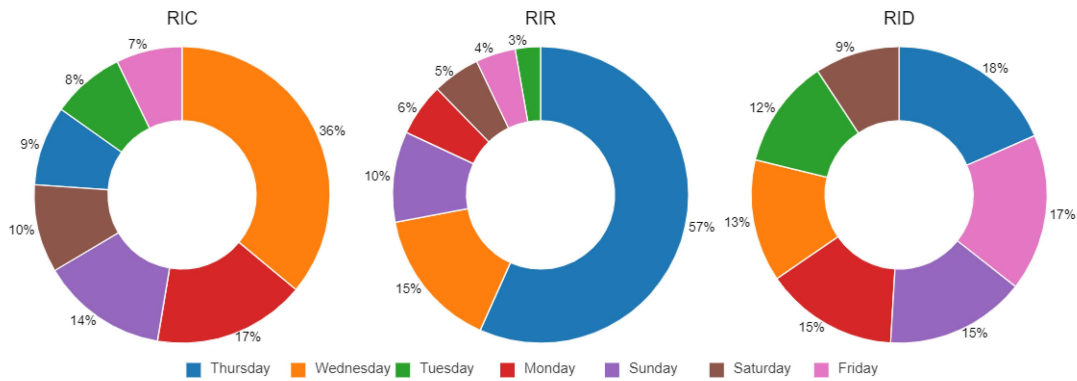


Fig. 9. PPU analysis before unlock.

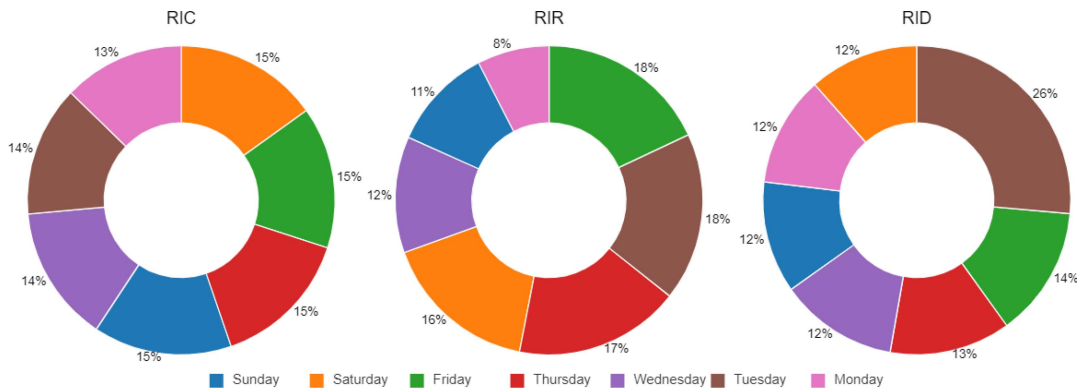


Fig. 10. PPU analysis after unlock.

Post-Unlock, the testing has been uniform and rigorous and is also increasing gradually with the rise of cases, as a result of which the cases and recoveries are detected daily with the maximum potential of testing. This is evident from the Fig. 10. All the days have almost equally distributed rise in cases and recoveries. Therefore, these spark-based analytics can be helpful for governing bodies in validating if the testing is being done continuously along with the rise of infection, by checking for uniform distribution of disease rise on a daily basis. This will help in maintaining rigorous testing, involving more economy-boosting activities as discussed earlier as part of ξ boost scheme and keeping the health quotient high for citizens in smart cities during the pandemic.

VI. CONCLUSION

Several countries are facing health and financial crisis amid the uncertainty of the COVID-19 pandemic. In this article, we proposed a ξ boost scheme, which is a fusion of AI technique and BDA, which provides early prediction of total positive cases & economy to promote economy-boosting activities. The ξ boost scheme facilitates the economy-boosting using AI and Spark-based analytics while considering the fact to control the pandemic at the same time. In this research work, we have presented a bidirectional LSTM-based DL model that can predict the total positive cases in India and the Indian economy. The model outshines when compared to the state-of-the-art approaches w.r.t. high prediction accuracy, in terms of a low MAPE value, i.e., 1.27%. Furthermore, this article

describes a novel SSP approach, for optimal selection of subsegment of days in which economy-boosting activities can be carried out to prevent a sharp downfall of the economy in the near future. Next, the spark-based PPU mechanism is presented for the Indian scenario to validate a consistent and rigorous increase in disease testing throughout the country. The fusion of AI and BDA (prediction and analytics) are beneficial for state and national governing bodies to devise relaxations for economy-boosting activities in smart cities on the optimal subsegment of days.

In the future, we will expand the ξ boost scheme for other countries' COVID-19 scenarios like the USA, China, and more, as every country has different impacts of the global pandemic. Thus, we shall extend the proposed SSP algorithm to identify if the relaxations for economy-boosting activities will be viable for a specific state considering its population density.

REFERENCES

- [1] M. Nicola *et al.*, "The socio-economic implications of the coronavirus pandemic (COVID-19): A review," *Int. J. Surg.*, vol. 78, pp. 185–193, Jun. 2020.
- [2] S. Malhotra. *Restoring the World Economy With AI and Machine Learning Post COVID-19*. Accessed: May 15, 2020. [Online]. Available: <https://artificialintelligence.oodles.io/blogs/world-economy-with-ai-machine-learning/>
- [3] *Explained: How COVID-19 Has Affected the Global Economy*. Accessed: Aug. 27, 2020. [Online]. Available: <https://indianexpress.com/article/explained/explained-how-has-covid-19-affected-the-global-economy-6410494/>

- [4] D. Zhang, M. Hu, and Q. Ji, "Financial markets under the global pandemic of COVID-19," *Finan. Res. Lett.*, vol. 36, Oct. 2020, Art. no. 101528.
- [5] J. W. Goodell, "COVID-19 and finance: Agendas for future research," *Finan. Res. Lett.*, vol. 35, Jul. 2020, Art. no. 101512.
- [6] *Machine Learning in Economics—How Is It Used?*, Addepto, Warsaw, Poland. Accessed: 20-Aug-2020. [Online]. Available: <https://addepto.com/machine-learning-in-economics-how-is-it-used/#:~:text=Forecasts>
- [7] S. R. Baker, N. Bloom, S. J. Davis, K. Kost, M. Sammon, and T. Viratyosin, "The unprecedented stock market reaction to COVID-19," *Rev. Asset Pricing Stud.*, vol. 10, no. 4, pp. 742–758, 2020.
- [8] D. S. W. Ting, L. Carin, V. Dzau, and T. Y. Wong, "Digital technology and COVID-19," *Nat. Med.*, vol. 26, no. 4, pp. 459–461, 2020.
- [9] R. Vaishya, M. Javadi, I. H. Khan, and A. Haleem, "Artificial intelligence (AI) applications for COVID-19 pandemic," *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 14, no. 4, pp. 337–339, 2020.
- [10] N. Rastogi, S. K. Singh, and P. K. Singh, "Privacy and security issues in big data: Through Indian prospective," in *Proc. 3rd Int. Conf. Internet Things Smart Innovat. Usages (IoT-SIU)*, Bhimtal, India, 2018, pp. 1–11.
- [11] A. Srivastava, S. K. Singh, S. Tanwar, and S. Tyagi, "Suitability of big data analytics in Indian banking sector to increase revenue and profitability," in *Proc. 3rd Int. Conf. Adv. Comput. Commun. Autom. (ICACCA) (Fall)*, Dehradun, India, 2017, pp. 1–6.
- [12] J. M. Grant-Kels, B. Sloan, J. Kantor, and D. M. Elston, "Big data and cutaneous manifestations of COVID-19," *J. Amer. Acad. Dermatol.*, vol. 83, no. 2, pp. 365–366, 2020.
- [13] X. Jiang *et al.*, "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity," *Comput. Mater. Continua*, vol. 63, no. 1, pp. 537–551, 2020.
- [14] N. L. Bragazzi, H. Dai, G. Damiani, M. Behzadifar, M. Martini, and J. Wu, "How big data and artificial intelligence can help better manage the COVID-19 pandemic," *Int. J. Environ. Res. Public Health*, vol. 17, no. 9, p. 3176, 2020.
- [15] Z. Zhao, X. Li, F. Liu, G. Zhu, C. Ma, and L. Wang, "Prediction of the COVID-19 spread in african countries and implications for prevention and control: A case study in South Africa, Egypt, Algeria, Nigeria, Senegal and Kenya," *Sci. Total Environ.*, vol. 729, Aug. 2020, Art. no. 138959.
- [16] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of COVID-19 infection: Systematic review and critical appraisal," *Brit. Med. J.*, vol. 369, p. m1328, Apr. 2020.
- [17] I. M. Ibrahim, D. H. Abdelmalek, M. E. Elshahat, and A. A. Elfiky, "COVID-19 spike-host cell receptor GRP78 binding site prediction," *J. Infect.*, vol. 80, no. 5, pp. 554–562, 2020.
- [18] N. Vankadari and J. A. Wilce, "Emerging WuHan (COVID-19) coronavirus: Glycan shield and structure prediction of spike glycoprotein and its interaction with human CD26," *Emerg. Microbes Infect.*, vol. 9, no. 1, pp. 601–604, 2020.
- [19] A. Kumari, S. Tanwar, S. Tyagi, N. Kumar, M. Maasberg, and K.-K. R. Choo, "Multimedia big data computing and Internet of Things applications: A taxonomy and process model," *J. Netw. Comput. Appl.*, vol. 124, pp. 169–195, Dec. 2018.
- [20] M. H. D. M. Ribeiro, R. G. da Silva, V. C. Mariani, and L. dos Santos Coelho, "Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil," *Chaos Solitons Fract.*, vol. 135, Jun. 2020, Art. no. 109853.
- [21] M. Mandal, S. Jana, S. K. Nandi, A. Khatua, S. Adak, and T. K. Kar, "A model based study on the dynamics of COVID-19: Prediction and control," *Chaos Solitons Fract.*, vol. 136, Jul. 2020, Art. no. 109889.
- [22] Z. Yang *et al.*, "Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions," *J. Thorac. Dis.*, vol. 12, no. 3, pp. 165–174, 2020.
- [23] A. S. Ahmar and E. B. del Val, "SutteARIMA: Short-term forecasting method, a case: COVID-19 and stock market in Spain," *Sci. Total Environ.*, vol. 729, Aug. 2020, Art. no. 138883.
- [24] T. Liang *et al.*, "Handbook of COVID-19 prevention and treatment: Compiled according to clinical experience," in *The First Affiliated Hospital*, vol. 68, Dept. School Med., Zhejiang Univ., Hangzhou, China, 2020.
- [25] E. M. Bloch *et al.*, "Deployment of convalescent plasma for the prevention and treatment of COVID-19," *J. Clin. Invest.*, vol. 130, no. 6, pp. 2757–2765, 2020.
- [26] J. Cao *et al.*, "Mathematical modeling and epidemic prediction of COVID-19 and its significance to epidemic prevention and control measures," *J. Biomed. Res. Innovat.*, vol. 1, no. 1, pp. 1–19, 2020.
- [27] X. Xu *et al.*, "A deep learning system to screen novel coronavirus disease 2019 pneumonia," *Engineering*, vol. 6, no. 10, pp. 1122–1129, 2020.
- [28] S. F. Ardabili *et al.*, "COVID-19 outbreak prediction with machine learning," vol. 13, no. 10, p. 249, 2020.
- [29] S. Gupta, G. S. Raghuvanshi, and A. Chanda, "Effect of weather on COVID-19 spread in the US: A prediction model for India in 2020," *Sci. Total Environ.*, vol. 728, Aug. 2020, Art. no. 138860.
- [30] S. Tuli, S. Tuli, R. Tuli, and S. S. Gill, "Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing," *Internet Things*, vol. 11, Sep. 2020, Art. no. 100222.
- [31] A. Tomar and N. Gupta, "Prediction for the spread of COVID-19 in India and effectiveness of preventive measures," *Sci. Total Environ.*, vol. 728, Aug. 2020, Art. no. 138762.
- [32] P. Arora, H. Kumar, and B. K. Panigrahi, "Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India," *Chaos Solitons Fract.*, vol. 139, Oct. 2020, Art. no. 110017.
- [33] *Who Acknowledges 'Evidence Emerging' of Airborne Spread of COVID-19*, HealthWorld, Noida, India. Accessed: Aug. 26, 2020. [Online]. Available: <https://health.economictimes.indiatimes.com/news/industry/who-acknowledges-evidence-emerging-of-airborne-spread-of-covid-19/76847993>.
- [34] A. Kumari, S. Tanwar, S. Tyagi, and N. Kumar, "Fog computing for healthcare 4.0 environment: Opportunities and challenges," *Comput. Elect. Eng.*, vol. 72, pp. 1–13, Nov. 2018.
- [35] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D, Nonlinear Phenom.*, vol. 404, Mar. 2020, Art. no. 132306.
- [36] S. Jain, S. Shukla, and R. Wadhvani, "Dynamic selection of normalization techniques using data complexity measures," *Expert Syst. Appl.*, vol. 106, pp. 252–262, Sep. 2018.
- [37] M. Xu, T. Li, Z. Wang, X. Deng, R. Yang, and Z. Guan, "Reducing complexity of HEVC: A deep learning approach," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5044–5059, Oct. 2018.
- [38] T. Ergen and S. S. Kozat, "Efficient online learning algorithms based on LSTM neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3772–3783, Aug. 2018.
- [39] S. Raja. *FNNs, RNNs, LSTM and BLSTM*. Accessed: Nov. 22, 2020. [Online]. Available: <http://cse.iitkgp.ac.in/~psraja/FNNs>
- [40] *COVID-19 Dashboard By the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)*, Johns Hopkins Univ. Med., Baltimore, MD, USA. Accessed: Jul. 30, 2020. [Online]. Available: <https://coronavirus.jhu.edu/map.html>
- [41] *Download Historical Rates for USD—INR to Excel*, Excelrates. Accessed: Aug. 20, 2020. [Online]. Available: <https://excelrates.com/historical-exchange-rates/USD-INR>
- [42] V. K. R. Chimmula and L. Zhang, "Time series forecasting of COVID-19 transmission in Canada using LSTM networks," *Chaos Solitons Fract.*, vol. 35, Jun. 2020, Art. no. 109864.
- [43] J. Koshy. *News Analysis, India May Have Undercounted Cases*. Accessed: Jun. 16, 2020. [Online]. Available: <https://www.thehindu.com/news/national/news-analysis-india-may-have-undercounted-cases/article31822079.ece>
- [44] *Is India Flattening the Testing Curve?*, Sci. THEWire, New Delhi, India. Accessed: Jun. 18, 2020. [Online]. Available: <https://science.thewire.in/health/covid-19-india-testing-rate-growth-states-data/>



Darshan Vekaria received the B.Tech. degree in information technology from the Institute of Technology, Nirma University, Ahmedabad, India, in 2020.

He has been a part of several research projects during his undergraduate studies which led to publications in leading conferences and journals of IEEE and Elsevier. His research interests include artificial intelligence, deep learning, big data analytics, and smart grids.



Aparna Kumari (Member, IEEE) received the Bachelor of Computer Application degree from the Birla Institute of Technology, Mesra, India, in 2006, and the master's degree in technology from Jawaharlal Nehru University, New Delhi, India, in 2012.

She is a full-time Ph.D. Research Scholar with the Computer Science and Engineering Department, Nirma University, Ahmedabad, India. She has authored/coauthored 11 publications (including nine papers in SCI Indexed Journals and two papers in IEEE ComSoc sponsored International Conferences). Some of her research works published in top-cited journals, for instance, the *Computer and Electrical Engineering*, the IEEE NETWORKS, the *Computer Communications* (Elsevier), and the *International Journal of Communication Systems* (Wiley). Her research interests include big data analytics, smart grid system, blockchain technology, and deep learning.

Ms. Kumari is a recipient of a Doctoral Scholarship from Nirma University, under the Ph.D. scheme.



Sudeep Tanwar (Member, IEEE) received the B.Tech. degree from Kurukshetra University, Kurukshetra, India, in 2002, the M.Tech. degree (Hons.) from Guru Gobind Singh Indraprastha University, New Delhi, India, in 2009, and the Ph.D. degree with specialization in wireless sensor network from the Faculty of Engineering and Technology, Mewar University, India, in 2016.

He is an Associate Professor with the Computer Science and Engineering Department, Institute of Technology, Nirma University, Ahmedabad, India.

He is a Visiting Professor with Jan Wyzykowski University, Polkowice, Poland, and the University of Pitesti, Pitesti, Romania. He has authored or coauthored more than 200 technical research papers published in leading journals and conferences from the IEEE, Elsevier, Springer, and Wiley. Some of his research findings are published in top-cited journals, such as the IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, the *Transactions on Emerging, Telecommunications Technologies*, the IEEE WIRELESS COMMUNICATIONS, the IEEE NETWORKS, the IEEE SYSTEMS JOURNAL, the IEEE ACCESS, the *IET Software*, the *IET Networks*, the *Journal of Information Security and Applications*, the *Computer Communication*, the *Applied Soft Computing*, the *Journal of Parallel and Distributed Computing*, the *Journal of Network and Computer Applications*, the *PubMed Central*, the *Sustainable Computing: Informatics and Systems*, the *Computer & Electrical Engineering*, the *International Journal of Computer Systems*, the *Software: Practice and Experience*, the *Multimedia Tools and Applications*, and the *Telecommunication System*. He has also edited/authored thirteen books with International/National Publishers like IET and Springer. One of the edited text-book titled, *Multimedia Big Data Computing for IoT Applications: Concepts, Paradigms, and Solutions* (Springer 2019) is having 3.7 million downloads till September 10, 2020. It attracts attention of the researchers across the globe. (<https://link.springer.com/book/10.1007/978-981-13-8759-3>). He has guided many students leading to M.E./M.Tech. and guiding students leading to Ph.D. His current interests include wireless sensor networks, fog computing, smart grid, IoT, and blockchain technology.

Dr. Tanwar is an Associate Editor of the *International Journal of Communication Systems* and the *Security and Privacy Journal* (Wiley). He was invited as the guest editor/editorial board member of many international journals, invited as keynote speaker in many international conferences held in Asia and invited as program chair, publications chair, publicity chair, and session chair in many international conferences held in North America, Europe, Asia, and Africa. He has been awarded best research paper awards from IEEE GLOBECOM 2018, IEEE ICC 2019, and Springer ICRIC-2019.



Neeraj Kumar (Senior Member, IEEE) received the Ph.D. degree in CSE from Shri Mata Vaishno Devi University, Katra, India, in 2009.

He was a Postdoctoral Research Fellow with Coventry University, Coventry, U.K., where he is a Visiting Professor. He is currently a Full Professor with the Department of Computer Science and Engineering, Thapar University, Patiala, India, and also with Asia University, and King Abdul Aziz University, Saudi Arabia. His research is supported by fundings from Tata Consultancy Service, Council

of Scientific and Industrial Research, and Department of Science Technology. He has published more than 400 technical research papers in leading journals and conferences from IEEE, Elsevier, Springer, and Wiley. Some of his research findings are published in top-cited journals, such as the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, the IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, the IEEE TRANSACTIONS ON CLOUD COMPUTING, the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the IEEE TRANSACTIONS ON CONSUMER ELECTRONICS, the IEEE NETWORK, the *IEEE Transactions on Communications*, the IEEE WIRELESS COMMUNICATIONS, the IEEE INTERNET OF THINGS JOURNAL, the IEEE SYSTEMS JOURNAL, the *Future Generation Computing Systems*, the *Journal of Network and Computer Applications*, and the *Computer Communications*. He has guided many Ph.D. and M.E./M.Tech.

Prof. Kumar has awarded Best Research Paper Awards from IEEE ICC 2018 and IEEE SYSTEMS JOURNAL 2018. He is leading the Research Group Sustainable Practices for Internet of Energy and Security, where group members are working on the latest cutting edge technologies. He is a TPC member and reviewer of many international conferences across the globe.