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

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
An AI-based disease detection and prevention scheme for COVID-19

Sudeep Tanwar a,*, Aparna Kumari a,*, Darshan Vekaria a, Neeraj Kumar b,c, Ravi Sharma d

a Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad, India
b Thapar Institute of Engineering and Technology, (Deemed to be University), Patiala, Punjab, India
c Department of Computer Science and Information Engineering, Asia University, Taichung, Taiwan
d Centre for Inter-Disciplinary Research and Innovation, University of Petroleum and Energy Studies, P.O. Bidholi Via-Prem Nagar, Dehradun, India

ARTICLE INFO

Keywords:
COVID-19
LSTM
AI
Disease prediction
ARIMA
Healthcare 4.0
Disease prevention

ABSTRACT

The proliferating outbreak of COVID-19 raises global health concerns and has brought many countries to a standstill. Several restrain strategies are imposed to suppress and flatten the mortality curve, such as lockdowns, quarantines, etc. Artificial Intelligence (AI) techniques could be a promising solution to leverage these restraint strategies. However, real-time decision-making necessitates a cloud-oriented AI solution to control the pandemic. Though many cloud-oriented solutions exist, they have not been fully exploited for real-time data accessibility and high prediction accuracy. Motivated by these facts, this paper proposes a cloud-oriented AI-based scheme referred to as D-espy (i.e., Disease-espy) for disease detection and prevention. The proposed D-espy scheme performs a comparative analysis between Autoregressive Integrated Moving Average (ARIMA), Vanilla Long Short Term Memory (LSTM), and Stacked LSTM techniques, which signify the dominance of Stacked LSTM in terms of prediction accuracy. Then, a Medical Resource Distribution (MRD) mechanism is proposed for the optimal distribution of medical resources. Next, a three-phase analysis of the COVID-19 spread is presented, which can benefit the governing bodies in deciding lockdown relaxation. Results show the efficacy of the D-espy scheme concerning 96.2% of prediction accuracy compared to the existing approaches.

1. Introduction

The global spread of mysterious COVID-19 started at the end of December 2019, identified as novel coronavirus/2019-nCOV. The genetic structure of COVID-19 is close to severe acute respiratory syndrome coronavirus 2 (SARS-COV2). The rapid spread of COVID-19 worldwide turned into an outbreak, and the World Health Organization (WHO) declared it a pandemic. Initial cases of this disease are associated with the live animal market in China (Wuhan City in December 2019). Still, within no time, human-to-human interaction has caused its presence in almost every part of the world with a variant of it, i.e., Omicron [1,2]. Fig. 1 shows the humongous count of COVID-19 cases worldwide [3]. The data shown in the bar graph is based on the records available by 3rd December 2021.
Any disease outbreak severely impacts the economy of a country as well as the lives of the people. It is evident from the given facts and figures regarding the emergence of COVID-19 that it has turned out to be a threat to humans, and it has to be contemplated as a vital domain for research to control the pandemic [4]. India is one of the countries which has been severely affected by COVID-19. Though the government imposed multiple lockdowns, quarantines, travel limits, and many other tactics to control the pandemic, the ceaseless rise in the cases proves the gravity of the situation and a need for efficient prediction and prevention planning to overcome the COVID-19 crisis.

To control this pandemic, AI is a viable solution for efficient prediction and prevention mechanism to assist healthcare operations [5,6]. AI bolster the use of Machine Learning (ML) and Deep Learning (DL) models for disease prediction, given their ability to learn about the trends and seasonality from the historical data [1]. Since several sources provide information about the disease and its rise, the government collects data from various relevant sources. Few data-driven research work has been carried out in this direction including statistical and mathematical model [7–10].

Most of the existing data-driven methods used linear methods and neglected the temporal components of the COVID-19 data. Then, the statistical models (e.g., Moving Average (MA), Auto-Regressive (AR), and Auto-Regressive Moving Average (ARIMA)) depend on assumptions that make it difficult to forecast real-time transmission rates [8,9]. These methods cannot fit perfectly in many cases, and the model’s accuracy is also quite low during prediction. Another challenge is the real-time access to the prediction result to all the stakeholders for future analysis and decision-making, which could be handled by cloud storage. Motivated by these facts, this paper proposes a cloud-oriented AI-based scheme referred to as D-espy (i.e., Disease-espy) for disease detection and prevention for the COVID-19 outbreak.

D-espy carries out a performance analysis of AI-based prediction techniques like ARIMA, Vanilla LSTM, and Stacked LSTM on the COVID-19 dataset for India. Therefore, a cloud-integrated AI-based prediction mechanism can be ideal for COVID-19 futuristic analysis. Then, the MRD mechanism has been designed based on the LSTM predictions to facilitate the distribution of medical resources, for instance, medicine, testing kits, masks, medical staff, and many more within the contaminated regions in the country.

With a limited workforce and medical resources, it is necessary to combat the COVID-19 spread optimally within the country. Therefore, AI plays a crucial role in this scheme for providing the three-week early predictions of the total COVID-19 cases, recoveries, and deaths. The AI-based Stacked LSTM prediction results will be used in the data-driven optimal MRD approach across various regions of the country. In contrast, the predictions result obtained from ARIMA, Vanilla, and Stacked LSTM models could be used for the phase analysis in lockdown scenarios. The proposed D-espy scheme will immensely benefit the government in deciding the proper distribution of medical resources (in a limited time period) to the people in more potential danger of the disease.

1.1. Motivation

Following are the motivation for this paper.

- The fatal aspects and the exponential growth of COVID-19 worldwide is one of the key criteria to explore this field. Accurate and early prediction of a pandemic will allow the government to plan preventive strategies and lockdowns beforehand.
- Existing research work mostly focused on forecasting the outbreak’s trends and rise. However, prevention schemes based on the rise of COVID-19 in different regions have not been explored to their full potential. So, there is a need for a preventive strategy that can optimally redistribute medical resources across various regions of the country to control the spread of disease.
- The proposed scheme opens the door for studying the future rise of the disease and planning lockdown relaxations and preventive strategies accordingly.
### Table 1
Abbreviations used.

| Symbols | Description          | Symbols | Description          |
|---------|----------------------|---------|----------------------|
| $D - espy$ | Disease-espy       | $JHU$   | Johns Hopkins University |
| LSTM    | Long Short Term Memory | $RMSE$ | Root Mean Square Error |
| ARIMA   | Auto Regressive Integrated Moving Average | $MRD$ | Medical Resource Distribution |
| $R^2$   | Regression score     | $COVID$ | Coronavirus Disease  |
| $L_i_p$ | Linear Interpolation | $D_p$   | Preprocessed data    |
| $w$     | COVID Data Sources   | $x$     | Data Server          |
| $\varphi$ | Optimal solution    | $CD$    | Cloud Data Storage  |
| $DS$    | Cloud obtained Data  | $DS^T$  | Transposed Data      |
| $DN$    | Data Normalization   | $D_i$   | Furnished Data       |
| $Pred$  | Predicted Data       | $Hist$  | Historical Data      |
| $R$     | Region               | $\mu$   | Medical Resource Set |
| $T$     | Time Duration        | $\Delta$ | Rate of rise in cases |
| $\Omega$ | Optimal Medical Resource set | $a$ | Rate of rise in recovery |
| $\beta$ | Rate of rise in Demise | $P$ | Population |
| $\Theta$ | Relative Sorting function | $\tau$ | Incrementer Function |
| $\nu$   | Decrmenter Function  | $\Phi$  | Optimal Resource Distribution Function |
| $L$     | Locality             | $R$     | Region               |
| $d$     | Factor of Transfer   | $WHO$   | World Health Organization |
| AI      | Artificial Intelligence | $ROC$  | Rate of Growth       |

1.2. Research contributions

Following are the research contributions of this paper.

- Proposed an AI-based disease prediction model (consists of ARIMA, Vanilla LSTM, and Full Stacked LSTM models) that predicts the number of total cases, demises, and recovery up to three weeks.
- Designed an MRD scheme that can help the government allocate resources to different regions based on the expected rise in the cases.
- Performance evaluation of $D$-esy over prediction accuracy by comparing it with state-of-the-art approaches.
- Performed three-phase analysis of the proposed $D$-esy scheme in the Indian scenario.

1.3. Organization

The remaining section of the paper is organized as follows. Section 2 highlights the background of COVID-19 and existing data-driven work based on AI techniques. Then, system architecture and problem formulation of the proposed scheme $D$-esy are described in Section 3. Further, Section 4 explains the workflow of the proposed scheme. Next, Section 5 shows the experimental results obtained by $D$-esy and includes a comparative analysis with other baseline models. At last, we conclude the paper in Section 6. Table 1 lists all the acronyms used in the paper.

2. State-of-the-art

Currently, the restrain strategy to control the global threat, i.e., COVID-19 is a major concern worldwide. Various studies are being conducted to predict the feasible progression of this epidemic using various mathematical and statistical models. Several research works to predict the transmission of the COVID-19 virus came up using mathematical models [11,12]. Further, the mathematical models are subject to potential bias due to their static nature and cannot change the model once formulated. Another side, a statistical model predicts the epidemiological trend of the occurrence and prevalence of COVID-19. Covid-19 consists of a time series dynamic data set, which requires sequential networks to extract patterns from it. Therefore, results are often vague from statistical and epidemiological models [13].

In order to overcome the barriers of these approaches, AI techniques could be a prominent solution to predict the real-time transmission of the disease [5,14]. The AI-based approaches provided efficient supervised and unsupervised learning of collected data and resulted to analytics [15]. Benvenuto et al. [16] presented an ARIMA model on the COVID-19 dataset given by Johns Hopkins University (JHU). Then, Chimmula et al. [13] has proposed an LSTM-based COVID-19 transmission time series forecasting scheme for Canada with 92.67% of prediction accuracy.

The above comparison of state-of-the-art techniques bolsters the case of using DL models like LSTM since they can capture the trends in time series data and give a forecast using the same range of time periods. Therefore, DL models like LSTM seems to be the best fit for early prediction of COVID-19 cases [22,23]. Table 2 presents a comparative summary of AI-based prediction approaches for COVID-19 outbreak with the proposed $D$-esy scheme. From the above discussion, it is evident that there is no significant work carried out for long-term predictions to tackle and limit the spread of COVID-19 in real-time [24,25]. Motivated by these facts, we proposed an AI-oriented $D$-esy scheme, a cloud-based prediction model that can accurately predict the disease spread over three weeks.
3. System architecture and problem formulation

The disease prediction and prevention are some of the significant aspects of controlling the spread of COVID-19 using a data-driven approach. This section presents the proposed D-espy scheme and its problem formulation for COVID-19 disease prediction.

3.1. System architecture

Fig. 2 illustrates the architecture of D-espy, which is categorized into three layers: (i) Physical Layer (PL), (ii) Data Transfer and Communication Layer (DTCL), and (iii) Data Analytics and Decision Making Layer (DADML). The layer-wise description is as follows.

3.1.1. Physical layer

This layer is responsible for the collection and transmission of the COVID-19 data from the relevant sources. The COVID-19 data can be obtained from different sources (\( \psi \)), for instance, patients, hospitals, social media, government reports, and the manual entry of records by health workers, \( \{ \psi_1, \psi_2, \ldots , \psi_n \} \in \psi \). Each of these sources store the data at data servers associated to them \( \{ x_1, x_2, \ldots , x_n \} \in x \) for information storage.

These scattered data can benefit decision-making when aggregated in one place for analytics. The government uses its cloud data storage (\( CD \)) to track each infection case. Cloud-based storage plays a crucial role here as it has the potential to provide analytics reports to various stakeholders in real-time. Control the spread of COVID-19 is not an easy task in a developing country like India; various government departments (e.g., police department and hospital) handle it on their end. For example, during a lockdown, the Indian government categorized infected and non-infected regions into three categories: red (contaminated zone), green (no cases of infection), and orange (fewer cases of infection) [26].

3.1.2. Data transfer and communication layer

Once data is collected in PL, it is transferred to the DTCL for processing. The government provides read access of \( CD \) for the people to carry out research and development to tackle the growth of the disease. The data obtained from PL is in raw format, which needs to be preprocessed for the user’s needs. Thus, several data preprocessing and cleaning actions are performed at this layer before transferring it to the next layer. The data is re-transformed to the required structure using transposition, the typecast to float from objects (if required), and then linearly interpolated to fill the missing or inconsistent value (if any). Finally, this data is normalized to a smaller range and converted in a fully furnished form \( D_t \) to make reliable analytics results. Here, the communication between PL and DADML layer is established using various existing communication technology such as wired connection (LAN/WAN) or wireless connection (Zigbee, 3G, 4G, etc.). COVID-19 data is labeled properly to avoid confusion when handling interactional complexity among layers.

| Approaches         | Year | Short Description                                                                 | Merits                                      | Demerits                                      |
|--------------------|------|-----------------------------------------------------------------------------------|---------------------------------------------|-----------------------------------------------|
| Benvenuto et al. [16] | 2020 | Presented an ARIMA model on the COVID-19 dataset given by JHU.                     | Forecast the epidemiological trend of COVID-19 prevalence and incidence | Application of proposed approach needs to be tested on other datasets too |
| Zheng et al. [17]   | 2020 | Hybrid approach using AI is presented for COVID-19 prediction                       | The prediction focused on impacted cities of China with significant MAPE values (i.e. with 0.52%, 0.38%, 0.05%, and 0.86%) | Emphasis short-term prediction only |
| Ribeiro et al. [18] | 2020 | Employed regression mechanism to predict COVID-19 cases in the Brazil.              | Multi-step-ahead predictions are made        | No mechanism is included to control the pandemic |
| Chimmula et al. [13] | 2020 | Time-series prediction of COVID-19 outbreak using LSTM                              | Patterns of COVID-19 data helps to control its transmission | Real-time data accessibility not included |
| Zandavi et al. [19] | 2021 | Dynamic hybrid approach to predict the COVID-19 spread using LSTM                   | Behavioral models is used under uncertainty  | Cloud-based real-time accessibility of data are not discussed |
| La Gatta et al. [20] | 2021 | Presented a neural network-based dynamic graph to predict COVID-19 outbreak         | Analyzed lock-down strategies for diffusion of epidemic | Real-time data availability is missing |
| Gautam et al. [21]  | 2021 | Transfer learning-based approach for COVID-19 cases and deaths prediction using LSTM| Single-step and multi-step predictions both are included in this approach | Presented approach performance need to be improved for US and Italy dataset |
| The proposed D-espy scheme | 2021 | Proposed a stacked LSTM model for COVID-19 outbreak prediction and prevention mechanism for it | Analysis for optimal distribution of medical resources | - |
3.1.3. Data analytics and decision making layer

This layer deals with the processing and analysis of COVID-19 data for its prediction and prevention. The preprocessed and fully furnished data $D_f$ is passed on to the Stacked LSTM prediction model for future predictions. The proposed LSTM model consists of different layers in the following order: Input LSTM Layer, Dropout layer, Hidden LSTM Layer, and Dense Output Layer. As the stacked LSTM model generates the predictions $\text{Pred}$, the MRD mechanism takes care of designing a prevention scheme to benefit the government in planning by considering the predictions $\text{Pred}$ as well as the historical data $\text{Hist}$.

Each country (like India) consists of several regions $\{R_1, R_2, \ldots, R_j\} \in R$. Each region is allocated with resources $\{\mu_1, \mu_2, \ldots, \mu_j\} \in \mu$ which is a sum of various elements like a number of testing kits, masks, medicines, government officials, doctors, testing teams, etc. Also, each $R_j$ will have a $\Delta_j$ percentage growth in the number of cases. Therefore, $\{\Delta_1, \Delta_2, \ldots, \Delta_j\} \in \Delta$ makes up the records of percentage growth of disease in each region. Each of the $\Delta_j$ will be computed based on the $\text{Pred}$ and $\text{Hist}$ values corresponding to the $R_j$. By considering $\Delta_j$ for each $R_j$, MRD re-evaluates the resource count value to $\Omega_j$ so that the regions with higher requirements and worse conditions in the disease outbreak get more resources to its help compared to the regions that have a relatively lower rate of growth and better conditions.
3.2. Problem formulation

Consider various localities \( \{L_1, L_2, \ldots, L_w\} \in L \) within a region \( R_j \). Each locality of a specific region is accomplished with various medical resources within it. Let \( \{S_1, S_2, \ldots, S_c\} \in S \) be the medical resources such as mask, medicines, testing kits, doctors, etc., which have been presented in each locality \( L_w \) of a region \( R_j \). Let \( S^d_{ij} \) and \( S^p_{ij} \) be the total doctors and miscellaneous medical resources, respectively, within the locality \( w \). Therefore, the total doctors and varied resources in each region \( R_j \) can be defined by:

\[
\mu^d_j = \sum_{i=1}^{w} S^d_{ij}. \tag{1}
\]

\[
\mu^p_j = \sum_{i=1}^{w} S^p_{ij}. \tag{2}
\]

Thus, total resources within Region \( R_j \) can be calculated as:

\[
\mu_j = \mu^d_j + \mu^p_j. \tag{3}
\]

Let \( \{\Delta_1, \Delta_2, \ldots, \Delta_j\} \in \Delta \) be the rate of growth in the number of total cases, \( \{a_1, a_2, \ldots, a_j\} \in \alpha \) be the total recovery, \( \{\beta_1, \beta_2, \ldots, \beta_j\} \in \beta \) be the demise rate and \( \{P_1, P_2, \ldots, P_j\} \in P \) be the set of population in each region. By taking the predicted disease growth \( \text{Pred}_j \) and current/previous disease record \( \text{Hist}_j \), \( \Delta_j \) is calculated as follows.

\[
\Delta_j = 100 \ast (\text{Pred}_j - \text{Hist}_j)/\text{Hist}_j. \tag{4}
\]

Let \( \Theta(a,b) \) be the relative sorting function such that it sorts elements at each index \( i \) in list \( a \) with respect to increasing order of corresponding index elements in \( b \). Therefore, we sort the list of \( \mu^d \) where its elements are sorted for the increasing order of their respective \( \Delta \), i.e., \( \Theta(\mu^d, \Delta) \). Here, \( \Delta \) shows the percentage increase in COVID-19 spread, which necessitates more medical resources to be supplied to control the pandemic.

Let \( \tau \) and \( \nu \) be the incremeneter and decremeneter functions, respectively. Let \( \Phi \) be the optimal redistribution function. Thus, the newly distributed medical resource set \( \Omega \) can be defined as follows.

\[
\Omega_x, \Omega_y = \Phi(\mu_x, P_x, \mu_y, P_y)), \tag{5}
\]

where, \( p < q, p < j, q < j \)

with the working of \( \Phi \) defined as:

\[
\Phi(\mu_x, P_x, \mu_y, P_y) = o(\mu^d_x \mu^d_y), \tau(\mu^p_x), \nu(\mu^p_y), \tag{6}
\]

\[
\forall(\Delta_j/P_x) < (\Delta_j/P_y) \]

subject to the constraints:

\[
S_{ij}, r, w, \Delta > 0 \tag{7}
\]

\[
\Omega \propto \alpha \tag{8}
\]

\[
\Omega \propto \Delta \tag{9}
\]

i.e., the new distribution of medical resources should cause an increase in the recovery rate and a decrease in the growth rate of total cases and demise. When we observe a decrease in the rate of growth in the cases for some regions, i.e., \( \Delta < 0 \), no medical resource distribution considering this region will take place as per the constraint shown in Eq. (8). When we decrease cases across all the regions, every region will have \( \Delta < 0 \). Thus no resource distribution takes place among any regions, as they all can decrease the rise in cases from the resource allotted to them in the previous iteration. Given the above discussion, the objective function for the \( D \)-espy is defined as follows.

\[
\varphi = \max \sum_{j=0}^{N} R_{\text{Distribution}}[\Omega_j] + \min(T)
\]

\[
\forall j \in [0, N]
\]

where \( R_{\text{Distribution}} \rightarrow \{\Omega_j\} \) denotes the maximum resource distribution in contaminated localities \( (L_w) \) using \( D \)-espy scheme in quickest time duration denoted as \( \min(T) \) in terms to prevent the disease to spread in the specific region \( R_j \).

4. \( D \)-espy : The proposed scheme

Fig. 3 shows the complete workflow of the proposed \( D \)-espy scheme. The features of the deadly disease COVID-19 are extracted from data of several sources to predict total cases, total demises, and total recoveries. One of the basic ways to collect data is from infected patients, who can provide a record of symptoms seen in the disease. One can get information about which other parts of the world are dealing with the same type of disease from social media by following various posts and tweets. Hospitals get the patients admitted and thus keep a daily count of the observed cases. Health workers, doctors, and nurses are producing the data related to
Algorithm 1 Algorithm for Stacked LSTM-based prediction model

**Input:** Data-set DS, Dropout x  
**Output:** Predicted Disease Growth Pred

**Initialization:** From DS Consider list of n Infected Regions, \( (R_1, R_2, \ldots, R_n) \). Let \( XT \) be the training dataset. Let \( N_i \) be the number of nodes in layer \( i \) and let \( RF \) be the return sequence.

1: procedure LSTM\_DISEASE\_FORECASTING( DS, x)
2:  
3:    \( DS \leftarrow TRANSPOSE(DS) \)
4:    \( TYPE\_CONVERT(DS.\text{Date},\text{date}−\text{time}) \)
5:    \( TYPE\_CONVERT(DS.\text{cases},\text{Float}) \)
6:    for \( i \leftarrow 1 \) to \( n \) do
7:      \( R_i \leftarrow INPUT(DS:\text{Region}_i) \)
8:    end for
9:    for \( i \leftarrow 1 \) to \( n \) do
10:      Normalize \( R_i \) in range (0,1)
11:    end for
12:    for \( i \leftarrow 1 \) to \( n \) do
13:      if \( R_i.\text{IsNotValid}() = \text{True} \) then
14:          \( R_i \leftarrow R_i.\text{Interpolate}() \)
15:      end if
16:    end for
17:    \( I\_seq \leftarrow \text{INITIALIZE\_SEQUENTIAL\_LSTM\_MODEL}(DS) \)
18:    \( x \leftarrow \text{LSTM}_1(R_i.N_1.RF = \text{True},\text{XT}\_\text{shape},1) \)
19:    \( \text{LSTM}_1.\text{DROPOUT}(x) \)
20:    \( y \leftarrow \text{LSTM}_2(x.N_2, RF = \text{False}) \)
21:    \( Pred \leftarrow \text{Dense}_1(y.N_1) \)
22:    return \( Pred \)
23: end procedure

---

**Fig. 3.** Workflow of D-espy.
the preventive measures that can be followed to minimize the spread of the disease and the medicines which can work towards curing the disease. The government employees keep track of the availability of resources like masks, testing kits, etc., which will be required in large quantities to deal with the outbreak.

Once the entire data is collected, it is stored in the cloud for analytics purposes. So, the relevant data from the cloud is taken ahead for preprocessing, which is as follows.

- The first step of data preprocessing involves cleaning the data. Duplicate cells in the dataset and irrelevant information are dropped and removed. Next, we filter out the information which turns out to be an outlier or an incorrect input in the dataset.
- Next, Data transformation is done to make the data structured in the desired format. We take the transpose of this dataset, i.e., \( D^T \), and drop the country column.
- The next step involves type conversion of the date-time attribute. Then, null and zero values are interpolated using Linear interpolation \( L_P \).
- Next, Data transformation is done to make the data structured in the desired format. We take the transpose of this dataset, i.e., \( D^T \) and drop the country column.
- Next, \( D^T \) is passed through data normalization \( D_N \), which converts all the data points in between the range \([0,1]\).

The preprocessed COVID-19 data is received by an AI-based prediction model. Here, based on the prediction accuracy, we have analyzed three AI-based prediction models, i.e., ARIMA, Vanilla LSTM, and Stacked LSTM. We select the model for the disease prevention mechanism. The model which can show the maximum accuracy in predicting the COVID-19 attributes is selected for further processing.

Firstly, we explored all the stated three models, i.e., ARIMA, Vanilla LSTM, and Stacked LSTM. The ARIMA is a statistical model used for time-series data to understand the COVID-19 data better and predict future trends of COVID-19. It is autoregressive in case it predicts future trends of COVID-19 based on past cases of COVID-19. Then, in stacked LSTM, layers are stacked together, and the outputs (cell states) of the first layer of LSTM cells are passed to the second layer of LSTM cells as inputs. Here, LSTM architectures comprise several hidden layers, can learn complex COVID-19 patterns efficiently, and increasingly create higher levels of representations of the input data sequence. Then, Vanilla LSTM is used, which is relatively simple to configure for the time-series data prediction, such as COVID-19 cases prediction, with a high accuracy rate compared to the existing approaches.

According to the results described in Section 5, the Stacked LSTM model gives the most accurate results. At the same time, Vanilla LSTM results are comparable to Stacked LSTM. Unlike the Stacked LSTM structure, Vanilla LSTM has only 1 LSTM layer and therefore takes lesser training and validation time. Therefore, in the near future, when the COVID-19 data becomes comparatively

---

**Algorithm 2 Algorithm for ROG Calculator**

**Input:** Predicted Disease growth \( Pred \), Historical/Current disease data \( Hist \), Attribute Name \( n \)  

**Output:** Set for Rate of growth of attribute \( n \) in each region

**Initialization:** Let \( j \) be the total no. of regions in the country

1: **procedure** CALCULATE_RATE_OF_GROWTH( \( Pred \), \( Hist \), \( n \) )  
2: if \( n \).equals(\( Cases \)) then  
3: \( P \leftarrow Pred_{Cases} \)  
4: \( H \leftarrow Hist_{Cases} \)  
5: for \( i \leftarrow 1 \) to \( j \) do  
6: \( \Delta_i \leftarrow (P_i - H_i)/H_i \)  
7: end for  
8: for \( i \leftarrow 1 \) to \( j \) do  
9: \( \Delta_i \leftarrow \Delta_i + 100 \)  
10: end for  
11: return \( \Delta \)  
12: end if  
13: if \( n \).equals(\( Demises \)) then  
14: \( P \leftarrow Pred_{Demises} \)  
15: \( H \leftarrow Hist_{Demises} \)  
16: for \( i \leftarrow 1 \) to \( j \) do  
17: \( \beta_i \leftarrow (P_i - H_i)/H_i \)  
18: end for  
19: for \( i \leftarrow 1 \) to \( j \) do  
20: \( \beta_i \leftarrow \beta_i + 100 \)  
21: end for  
22: return \( \beta \)  
23: end if  
24: if \( n \).equals(\( Recovery \)) then  
25: \( P \leftarrow Pred_{Recovery} \)  
26: \( H \leftarrow Hist_{Recovery} \)  
27: for \( i \leftarrow 1 \) to \( j \) do  
28: \( a_i \leftarrow (P_i - H_i)/H_i \)  
29: end for  
30: for \( i \leftarrow 1 \) to \( j \) do  
31: \( a_i \leftarrow a_i + 100 \)  
32: end for  
33: return \( a \)  
34: end if  
35: **end procedure**
Algorithm 3 Algorithm for MRD

Input: Predicted Disease Growth \(\text{Pred}\), Historical/Current Data of Disease \(\text{Hist}\), Set of medical resources per region \(\mu\), Population Set \(P\)

Output: Set of Optimal Distributed Medical Resource \(\Omega\)

Initialization: Let \(j\) be the no. of regions in the country, and let \(\mu\) be the initial allotment of medical resources in all the regions.

1: procedure CALCULATE_OPTIMAL_DISTRIBUTION(\(\text{Pred, Hist, }\mu, P\))
2: \(\Delta \leftarrow \text{ROG_calculator}(\text{Pred, Hist, Cases})\)
3: INCREASING_SORT(\(\Delta\))
4: \(x = 1, y = j\)
5: while \(x < y\) do
6: \(\text{if } (\Delta_x < 0 \& \& \Delta_y < 0) \text{ then} \)
7: \(\Delta_x \leftarrow \mu_x\)
8: \(\Delta_y \leftarrow \mu_y\)
9: \(x \leftarrow x + 1\)
10: \(y \leftarrow y - 1\)
11: end if
12: \(\text{if } (\Delta_x < 0 \& \& \Delta_y > 0) \text{ then} \)
13: \(\text{if } (\Delta_x / P_x) \leq (\Delta_y / P_y) \text{ then} \)
14: \(\Delta_x^{\text{inc}} \leftarrow \mu_x^{\text{inc}} + 0.5 \times (\mu_y^{\text{inc}})\)
15: \(\Delta_y^{\text{inc}} \leftarrow 0.5 \times (\mu_x^{\text{inc}})\)
16: \(\Delta_x^{\text{inc}} = 1 \times \text{INCREASED} - \text{FUN_D}\(\mu_x^{\text{inc}})\)
17: \(\Delta_y^{\text{inc}} = \mu_y^{\text{fund}}\)
18: \(\Delta_x = \Delta_x^{\text{inc}} + \Delta_x^{\text{fund}}\)
19: \(\Delta_y = \Delta_y^{\text{inc}} + \Delta_y^{\text{fund}}\)
20: end if
21: \(x \leftarrow x + 1\)
22: \(y \leftarrow y - 1\)
23: end if
24: \(\text{if } (\Delta_x > 0 \& \& \Delta_y > 0) \text{ then} \)
25: \(\text{if } (\Delta_x / P_x) \leq (\Delta_y / P_y) \text{ then} \)
26: \(d = \Delta_x - \Delta_y\)
27: \(d = d/100\)
28: \(d = M I N I M U M (0.5, d)\)
29: \(\Delta_x^{\text{inc}} = \mu_x^{\text{inc}} + d \times (\mu_y^{\text{inc}})\)
30: \(\Delta_y^{\text{inc}} = (1 - d) \times (\mu_x^{\text{inc}})\)
31: \(\Delta_x^{\text{inc}} = 1 \times \text{INCREASED} - \text{FUN_D}\(\mu_x^{\text{inc}})\)
32: \(\Delta_y^{\text{inc}} = \mu_y^{\text{inc}}\)
33: \(\Delta_x = \Delta_x^{\text{inc}} + \Delta_x^{\text{fund}}\)
34: \(\Delta_y = \Delta_y^{\text{inc}} + \Delta_y^{\text{fund}}\)
35: end if
36: \(x \leftarrow x + 1\)
37: \(y \leftarrow y - 1\)
38: end if
39: end while
40: return \(\Omega\)
41: end procedure

higher in volume, Vanilla LSTM can be beneficial in producing good results in lesser computational time. ARIMA predictions can be useful in the case where the accuracy of the predictions is not a factor, and we wish to know only about the trend of COVID-19 rise in the future. Algorithm 1 shows the working principle of the proposed Stacked LSTM-based prediction model. Then, the predicted values \(\text{Pred}\) and the previously available historical data \(\text{Hist}\) regarding the disease are fed into the Rate of Growth (ROG) calculator.

Algorithm 2 illustrates the rate of growth calculator to compute the spread rate. We select \(\Delta\) as the parameter to develop prevention schemes and analytics since the rise in cases highly defines the intensity of the disease in every region throughout the country. Therefore ROG passes \(\Delta\) to the MRD. According to the Journal of Family Medicine and Primary Care, WHO has recommended a doctor: population ratio to be 1:1000 [27]. Therefore, it is necessary to consider a region’s \(P\) and \(\Delta\) while distributing the medical resources optimally within the regions using MRD.

Algorithm 3 shows the sorted list of regions corresponding to \(\Delta\) increment in the infected cases. Therefore, the list will have the first region with the smallest ROG and the last region with the largest ROG. If \(\Delta\) for both of them is negative, i.e., there has been no rise in the cases, we make no change in their resources and go to the next pair, i.e., the second-lowest and the second-highest regions for \(\Delta\). If the \(\Delta\) for one of the regions is negative and its \(\Delta\): \(P\) ratio is also lower than the other region, half of the resources will be transferred from the lower \(\Delta\) region to the higher \(\Delta\) region and the funds for other miscellaneous medical resources will be proposed to be increased by the government. In case both the regions have a rise in cases, i.e., \(\Delta > 0\), the region with lower \(\Delta\) will contribute resources to the region with higher \(\Delta\), given the contributing region has a lower \(\Delta\): \(P\) ratio. The factor of transfer \(d\) is decided by the difference \(\Delta\) of both the regions. The limit of \(d\) is designed to be 0.5, i.e., 50%, since transferring more resources can lead to a rise in the cases later, where we initially saw lower \(\Delta\).

5. Experimental results and discussion

This section discusses experimental results, datasets, and phase analysis of \(D\)-espy scheme. Firstly, the Experimental Setup and Prediction of attributes subsection consist of details of the dataset and experimental environment setup and results of all three AI
models used in the proposed D-espy scheme. Then, the following subsection highlights the Performance evaluation and comparative analysis, which deals with the discussion on the accuracy values of the model and its comparison with a baseline model. Finally, the last subsection Phase analysis discussion shows the benefits of using predictions of all three models in strategic planning to control the transmission of COVID-19.

5.1. Experimental setup and prediction of attributes

A real-world dataset provided by the JHU has been taken, and the relevant information related to India is extracted from it [28,29]. The COVID-19 data from 22 January 2020 to 28 May 2020 has been considered for testing the D-espy scheme. Here, JHU data is divided into an 80:20 ratio to evaluate the performance of the proposed D-espy scheme [28]. Then, 22 days (7 May 2020 to 28 May 2020) of predictions were generated from Stacked LSTM, Vanilla LSTM models, and ARIMA models. The JHU dataset is preprocessed through Pandas v1.0.4 and NumPy v1.18.4 library packages. After preprocessing COVID-19 data, the D-espy scheme is implemented using Keras v2.3.1 with the Tensorflow package. The AI models, i.e., ARIMA, Vanilla LSTM, and Stacked LSTM (part of D-espy scheme) are being trained using 12 GB NVIDIA Tesla K80 GPU.

Here, we used the COVID-19 data of India for training and testing purposes when the lockdown was imposed on the entire nation. The infected cases would be less during lockdown compared to the non-lockdown time. Therefore, the result set refers to when the lockdown was still on. The results of these models might differ in terms of accuracy when tested upon data from a situation where the lockdown has been uplifted from the country. Here, we considered the total cases, demises, and recovery cases as the attributes for COVID-19 disease forecasting. The count of cases, recovery, and deaths plotted in Fig. 5 signifies the total count of cases/recoveries/deaths recorded till the given date(in thousands). They are not the active count of the attributes on a particular day. Details of each attribute are as follows:

- **Total Active Cases**: Fig. 4 shows the prediction of a total number of cases using ARIMA, Vanilla LSTM, and Stacked LSTM. The Fig. 4(a) comprises detection of the trend, but its accuracy is below par, which makes it difficult to implement in real-life scenarios. Thus, the ARIMA model can be used if we only want to know the future trend of the disease and its related attributes. Then, Figs. 4(b) and 4(c) shows quite accurate prediction of the total cases using Vanilla and stacked LSTM. Prediction accuracy of Vanilla LSTM lags behind that of stacked LSTM on a minor basis. One key difference between the Stacked LSTM model and the Vanilla LSTM model is that the latter has no hidden LSTM layer of 50 nodes, which is there in the Stacked LSTM model. So, the Vanilla LSTM model takes a little less computation time to produce the forecast regarding Stacked LSTM. But, this comes at the cost of accuracy. With fewer data available for the disease in the initial stages, Stacked LSTM will be a wiser choice since computation time is not an issue, whereas Vanilla LSTM can be incorporated for large datasets.

- **Total Demises**: Fig. 5 shows the prediction of demises using ARIMA, Vanilla LSTM and Stacked LSTM. Again, based on the COVID-19 time-series data, the Stacked LSTM obtained a considerably high accuracy compared to the other two models.

- **Total Recovery**: Fig. 6 shows the prediction of demises using ARIMA, Vanilla LSTM and Stacked LSTM. Again, the Stacked LSTM-based prediction model on the COVID-19 time-series data obtained higher accuracy when compared to the other two models. Thus, the Stacked LSTM model is the best of all and gives accurate and precise results. Therefore, we consider these results highly useful for predicting the attributes for the next 22 days.
5.2. Performance evaluation and comparative analysis

RMSE value (Root-Mean-Square error) and $R^2$ (coefficient of determination) regression score are the two parameters for evaluating all the models. RMSE calculation is done as:

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

(10)

where $N$ is the number of samples, $x_i$ and $y_i$ are the actual and predicted value of the sample. Table 3 shows the RMSE values of all three AI models against each attribute.

We have also taken into account the $R^2$ score for predictions. For example, if $y$ is the set of $n$ original data points and $f$ is the set of corresponding predictions, then the $R^2$ is calculated as follows.

$$y_{avg} = \frac{1}{N} \sum_{i=1}^{N} y_i.$$  

(11)

$$Dif_{tot} = \sum_{i=1}^{k} (Y_i - y_{avg}).$$

(12)
The best possible value of the $R^2$ score is 1, which usually ranges between negative values and 1. The major benefit of using the $R^2$ score is that it can prevent the model from overfitting. The overfit model will contain too many predictors and start considering random noise. Since it is difficult to predict random noise, an overfit model will drop the $R^2$ score [30]. Table 4 shows the high accuracy of the models regarding the $R^2$ score (by multiplying with 100).

Fig. 7 shows the comparison between the proposed scheme and baseline model [13]. We compare D-espy scheme with other state-of-the-art approaches in terms of prediction accuracy. The baseline model obtained the prediction accuracy of 93.4% using the LSTM model for short-term forecasting, while D-espy scheme outshines with aggregated prediction accuracy of 96.2%. The baseline approach has considered their training data to span until 31st March 2020. Since COVID-19 rises are dynamic, the larger the training data size, the better the reliability of the results. D-espy has a larger size of training data, spanning until 30th April 2020. Thus, D-espy results can be considered more reliable as they teach the latest rising trends in COVID-19. In terms of complexity, D-espy outshines the existing methods as it can predict COVID-19 cases and recoveries and deaths using a single Stacked LSTM Model for three weeks.

5.3. Phase analysis discussion

Detecting and preventing any disease is an integral part of government planning during an outbreak like COVID-19. The prevention mechanism required proper planning, for instance, imposing lockdown, restricted travel limits, social distancing, shutting down public places, and many more to control the spread of this COVID-19. The essential medical resources like medicines, hospitals, and other daily-life supporting products/services must be available to control the transmission of COVID-19. Therefore, we perform a three-phase analysis for a rise in the total cases, recoveries, and demises by taking the data from 8 May to 28 May 2020 (data belongs to the lockdown period). Figs. 8, 9 and 10 shows the three-phase analysis, where each phase spans for 7 days. We have considered Phase-1 to span from 8 May to 14 May, Phase-2 from 15 May to 21 May, and Phase-3 from 22 May from 28 May. It includes an original count of the attributes and the corresponding predicted count by all three AI Models on the last day of each phase. Since this phase analysis uses the values of the prediction from all models in a phase-wise manner, it would be easy to conclude which model is the best in predicting a particular attribute in a given phase.

Phase-1 consists of maximum restriction (i.e., strict lockdown) on all the services. People must maintain social distancing and avoid being in contact with other people. Next, phase-2 comprises analysis with some relaxation after observing disease rise during phase-1. The government gives relaxation from lockdown (waived off a few lockdown restrictions) for basic and moderately essential facilities. The rise of disease in phase 2 plays an important role in the relaxation decision in phase 3. If the transmission speed shows a low trend, then Phase-3 brings more relaxation to the basic facilities and activities.

### Table 4

| Attributes/Features | ARIMA | Vanilla LSTM | Stacked LSTM |
|--------------------|-------|--------------|--------------|
| Demise             | 82.2  | 92.0         | 94.1         |
| Recovery           | 57.8  | 94.8         | 96.9         |
| Total Cases        | 58.2  | 94.2         | 97.6         |
| Average            | 66.1  | 93.6         | 96.2         |
The prediction models proposed in this paper can be used to plan the phases. Figs. 8, 9 and 10 shows actual statistics of COVID-19 in India and the corresponding predictions by the models. Although we saw earlier that the Stacked LSTM model is the most accurate overall, the predictions from all three models can be beneficial for the phase analysis.

Considering a scenario here, the government wants to decide on a relaxation of service $x$ in phase-2 during a lockdown. Let us say the relaxation on $x$ highly depends on the number of recoveries. According to the Fig. 9, the Vanilla LSTM model gives the most accurate result for predicting the recoveries in Phase 2. Therefore, the results of Vanilla LSTM can be followed whenever relaxation on $x$ services is decided for phase 2 of lockdown. So, even though the results of Section 5 displayed dominance of Stacked LSTM results over Vanilla LSTM and ARIMA, we will consider the results of Vanilla LSTM when it comes to making decisions related to recoveries in Phase 2 of any lockdown. It justifies that the predictions of all three models are highly useful, as we can map which model gives the best results for every attribute in every phase and use the results of that model accordingly. Thus, these analytical results benefit the governing bodies in planning the degree of relaxation on important services during each lockdown phase.

6. Conclusion

The global threat of COVID-19 necessitates predicting the feasible progression of this outbreak to control it. In this paper, we proposed an AI-driven disease prediction and prevention scheme, i.e., $D$-espy based on the cloud. In this work, we have presented
and tested three prediction models: ARIMA, Vanilla LSTM, and Stacked LSTM, for predicting disease growth in India using the JHU dataset. The major attributes of the COVID-19 data comprise a total number of cases, recovery, and demise. Then, disease growth is predicted based on these attributes using the AI model. The stacked LSTM model obtained a considerably high accuracy with an average accuracy of 96.2%. These results outclass several other conventional prediction algorithms. Also, to the best of our knowledge, none of the existing work predicts the COVID-19 attributes for such a considerable time span along with the mentioned accuracy. Furthermore, the paper describes a novel D-espy scheme for COVID-19 detection and prevention through the optimal distribution of crucial medical resources (including medical staff) among the contaminated regions of India. It could help to control the COVID-19 spread on a national and international basis. The paper also discusses the three-phase analysis of a lockdown during COVID-19 in India and the potential benefits of the predictions of all three models in deciding the relaxations on important services. This analysis can benefit the government in making timely decisions to restrain COVID-19. Shortly, this work will be extended in the future to provide a more accurate prediction of COVID-19 by tuning the hyperparameters. Finally, we hope that this study gives some references for future research to control the epidemic.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.compeleceng.2022.108352.

Data availability

No data was used for the research described in the article.

References

[1] Mohamadou Y, Halidou A, Kapen PT. A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19. Appl Intell 2020;50(11):3913–25.
[2] Xu X, Jiang X, Ma C, Du P, Li X, Lv S, et al. A deep learning system to screen novel Coronavirus disease 2019 pneumonia. Engineering 2020;6(10):1122–9.
[3] Worldometers. 2021, https://www.worldometers.info/coronavirus/ (Online; Accessed: 3-December-2021).
[4] Gupta R, Kumari A, Tanwar S, Kumar N. Blockchain-envisioned softwareized multi-swarming UAVs to tackle COVID-19 situations. IEEE Netw 2020;1–8.
[5] Talati S, Vekaria D, Kumari A, Tanwar S. An AI-driven object segmentation and speed control scheme for autonomous moving platforms. Comput Netw 2021;186:107783.
[6] Kumari A, Tanwar S, Tyagi S, Kumar N. Fog computing for healthcare 4.0 environment: Opportunities and challenges. Comput Electr Eng 2018;72:1–13.
[7] Kumari A, Tanwar S, Tyagi S, Kumar N. Verification and validation techniques for streaming big data analytics in internet of things environment. IET Netw 2018;8(2):92–100.
[8] Alimadadi A, Aryan S, Manandhar I, Munroe PB, Joe B, Cheng X. Artificial intelligence and machine learning to fight COVID-19. Physiol Genomics 2020;52(4):200–2, PMID: 32216577.
[9] Wynants L, Van Calster B, Bonten MM, Collins GS, Debray TP, De Vos M, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. Bmj 2020;369.
[10] Roda WC, Varughese MB, Han D, Li MY. Why is it difficult to accurately predict the COVID-19 epidemic? Infect Dis Model 2020.
[11] Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. Int J Infect Dis 2020;92:214–7.
[12] Shim E, Tariq A, Choi W, Lee Y, Chowell G. Transmission potential and severity of COVID-19 in South Korea. Int J Infect Dis 2020;93:339–44.
[13] Chimnula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos Solitons Fractals 2020;135:109864–72.
[14] Kumari A, Tanwar S. Secure data analytics for smart grid systems in a sustainable smart city: Challenges, solutions, and future directions. Sustain Comput Inf Syst 2020;28:100427.
[15] Tanwar S, Bhatia Q, Patel P, Kumari A, Singh PK, Hong W-C. Machine learning adoption in blockchain-based smart applications: The challenges, and a way forward. IEEE Access 2019;8:474–88.
[16] Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M. Application of the ARIMA model in forecasting COVID-19 epidemic data. Brief 2020;29:105340–51.
[17] Zheng N, Du S, Wang J, Zhang H, Cui W, Kang Z, et al. Predicting COVID-19 in China using hybrid AI model. IEEE Trans Cybern 2020;50(7):2891–904.
[18] Ribeiro MHDM, da Silva RG, Mariani VC, dos Santos Coelho L. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. Chaos Solitons Fractals 2020;135:109864–72.
[19] Zandavi SM, Rashidi TH, Valae F. Dynamic hybrid model to forecast the spread of COVID-19 using LSTM and behavioral models under uncertainty. IEEE Trans Cybern 2021;1–13.
[20] La Gatta V, Moscato V, Postiglione M, Sperli G. Artificial neural network exploiting dynamic graph structured data applied to the COVID-19 outbreak. IEEE Trans Big Data 2021;7(1):45–55.
[21] Gautam Y. Transfer learning for COVID-19 cases and deaths forecast using LSTM network. ISA Trans 2021.
[22] Vekaria D, Kumari A, Tanwar S, Kumar N. Boost: An AI-based data analytics scheme for COVID-19 prediction and economy boosting. IEEE Internet Things J 2020;8(21):15977–89.
[23] Gupta R, Kumari A, Tanwar S, Kumar N. Blockchain-envisioned softwareized multi-swarming uavs to tackle covid-19 situations. IEEE Access 2020;52(4):200–2, PMID: 32216577.
[24] Wang L, Lin ZQ, Wong A. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. Sci Rep 2020;10(1):1–12.
[25] Chowdhury ME, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, et al. Can AI help in screening viral and COVID-19 pneumonia? IEEE Access 2020;8:132665–76.
[26] COVID-19 containment plan: What are red, orange, green zones?. 2021, https://www.timesnownews.com/india/article/coronavirus-zones-and-their-meanings-covid-19-containment-plan-what-are-red-orange-green-zones-580094 (Online; Accessed: 5-December-2021).
[27] Raman K, Ranabir P. India achieves WHO recommended doctor population ratio: A call for paradigm shift in public health discourse!. 2021, https://www.jfmpc.com/text.asp?2018/7/5/841/245755 (Online; Accessed: 7-December-2021).
[28] JHU CSSE. 2021, https://github.com/CSSEGISandData/COVID-19 (Online; Accessed: 30-November-2021).
[29] India ministry of health and family welfare. 2022, https://www.mohfw.gov.in/ (Online; Accessed: 30-July-2022).
[30] Regression analysis tutorial and examples. 2022, https://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-tutorial-and-examples (Online; Accessed: 29-July-2022).

Sudeep Tanwar (M’15, SM'21) is working as a full Professor at Nirma University, India. He received his Ph.D. in computer science and engineering from Mewar University, India. His research interests include WSN, blockchain technology, fog computing, and smart grid. He has authored/co-authored more than 270 research papers in leading journals and conferences of repute and has edited/authored more than 24 books published in leading publication houses.

Aparna Kumari is working as an Assistant Professor in the Computer science and Engineering Department at Nirma University, India. She has authored/co-authored some SCI Indexed Journals publications and IEEE ComSoc-sponsored International Conferences. Her research interests include smart grid, blockchain, and data analytics.

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

Neeraj Kumar (Senior Member, IEEE) received a Ph.D. degree in CSE from Shri Mata Vaishno Devi University, Katra (Jammu and Kashmir), India. He was a Postdoctoral Research Fellow with Coventry University, Coventry, U.K. He is currently a Full Professor with the Department of Computer Science and Engineering, Thapar University, Patiala (Pb.), India. He has published over 800 technical research papers in leading journals and conferences.

Ravi Sharma is working as a Professor in the Centre for Inter-Disciplinary Research and Innovation, University of Petroleum and Energy Studies, Dehradun, India. He has contributed various articles on business analytics, prototype building for a startup, and artificial intelligence. He is one of the leading academic consultants to uplift research activities in inter-disciplinary domains.