IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 August 2023, accepted 20 August 2023, date of publication 29 August 2023, date of current version 7 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3309601

SURVEY

Short-Term Traffic Prediction Using Deep Learning Long Short-Term Memory: Taxonomy, Applications, Challenges, and Future Trends

ANWAR KHAN[1], MOSTAFA M. FOUDA[2], (Senior Member, IEEE), DINH-THUAN DO[3], (Senior Member, IEEE), ABDULAZIZ ALMALEH[4], AND ATIQ UR RAHMAN[5]

[1]Department of Electronics, University of Peshawar, Peshawar 25120, Pakistan
[2]Department of Electrical and Computer Engineering, Idaho State University, Pocatello, ID 83209, USA
[3]School of Engineering, University of Mount Union, Alliance, OH 44601, USA
[4]College of Computer Science, Information Systems Department, King Khalid University, Abha 61421, Saudi Arabia
[5]Faculty of Computer Information Science, Higher Colleges of Technology, Ras Al Khaimah Campus, Ras Al Khaimah, United Arab Emirates

Corresponding author: Anwar Khan (arkhan@uop.edu.pk)

ABSTRACT This paper surveys the short-term road traffic forecast algorithms based on the long-short term memory (LSTM) model of deep learning. The algorithms developed in the last three years are studied and analyzed. This provides an in-depth and thorough description of the algorithms rather than their marginal description as performed in the existing surveys that focus on general deep learning algorithms. The chosen algorithms are classified depending upon the use of LSTM in combination with other techniques for processing input data features towards a final traffic forecast. The operational strategies of the algorithms are described with merits and limitations. Moreover, a comparative analysis of the compared classes of algorithms is also provided. These strategies are helpful in selection of the right algorithms and their classes for the diverse traffic conditions and their future investigation for improvement. Besides, the applications of these classes of algorithms to traffic forecast in various networks for the latest decade is graphically depicted. Moreover, the applications of the LSTM in other fields involving a forecast are provided. Finally, the challenges associated with the short-term traffic forecast using the LSTM are described and strategies are highlighted for their future investigation.

INDEX TERMS Short-term traffic prediction, long short-term memory, LSTM, deep learning, intelligent transportation.

I. INTRODUCTION

In intelligent transportation systems (ITS), short-term traffic prediction is one of the major disciplines of research that aims to ensure efficient traffic management by overcoming or minimizing the traffic challenges [1]. Specifically, in road traffic management, it helps in ensuring traffic lane management, predicting traffic congestion, ensuring selection of the optimal traffic paths and road safety by defining the optimal traffic speed, prevention and detection of accidents, adaptive identification of surrounding traffic and estimating the travel time to destination, to mention a few [2]. In particular, these strategies are helpful and effective in governing the journey and interaction of human-driven as well as autonomous vehicles [3].

The long short-term memory (LSTM) model is one of the research paradigms of deep learning that has been recently used for traffic prediction in ITS [4], [5], [6], [7]. It has the inherent capability of modeling the stochastic nature of traffic data and identifying their spatio-temporal characteristics. The LSTM-based traffic networks retain the long and short-term data in memory and use them for a prediction decision at the current instant of time. This is contrary to the conventional use of deep learning methods where no memory is involved

The associate editor coordinating the review of this manuscript and approving it for publication was Yanli Xu.

in performing an output decision. In addition to traffic prediction, the LSTM algorithms have also been recently used in stock prediction [8], classification of plant diseases [9], prediction of common diseases in humans [10], health and effective lifetime forecast of electric vehicles [11], prediction of carbon dioxide based on sensory data from vehicles [12], weather forecasting [13] and channel estimation in wireless communications [14], to mention a few.

This paper conducts a survey of the short-term road traffic forecast algorithms using the LSTM designed in the last three years. This strategy is helpful in choosing the latest algorithms with a thorough and in-depth description rather than providing a general description of the deep learning algorithms as performed in the existing surveys [15], [16]. The algorithms are categorized based on use of the LSTM in combination with other techniques for processing data input towards the final output forecast. Every algorithm is described with its operational strategy, merits and limitations. A comparative analysis of the categorized algorithms is also provided. These strategies are helpful in selection of algorithms and their classes for traffic forecast in diverse conditions. In addition, the applications of these classes of algorithms to traffic prediction in various networks over the latest decade is graphically depicted. Besides, applications of the LSTM in other fields involving a forecast are also given. Finally, challenges in LSTM-based traffic forecast are identified and techniques are highlighted for their future investigation.

In essence, the primary goals of this research include reviewing and categorizing the LSTM-based traffic forecast algorithms, analyzing their operational strategies, merits, limitations and highlighting the future directions. The contributions of this survey are:

- The LSTM-based traffic forecast algorithms developed in the last three years are incorporated. This allows an in-depth and thorough description of the algorithms rather than their marginal description with other deep learning algorithms as performed in some existing surveys.
- The selected algorithms are classified into various categories based on the techniques combined with the LSTM and a comparative analysis is performed. Every algorithm's key operation, merits and limitations are described too. These strategies are helpful in the choice of the appropriate algorithms for a specific type of road traffic management. Besides, the applications of the LSTM to other fields involving a forecast are given. Moreover, the applications of each class of algorithms in traffic prediction for various networks for one decade: 2013 to 2023, are graphically depicted in various networks.
- The challenges that prevail in the LSTM-based algorithms for short-term traffic forecast are addressed and techniques are highlighted for future investigation.

The rest of this paper is organized as follows: Section II describes the organization of the manuscript. Section III deals with the background of traffic forecast using LSTM and Section IV highlights the origin of the LSTM model. The architecture of the LSTM is addressed in Section V while Section VI classifies the considered algorithms, performs a comparative analysis and depicts their applications in various networks for the latest decade. The challenges associated with the short-term road traffic prediction are given in Section VII while Section VIII concludes the paper with a description of the future research strategies.

II. ORGANIZATION OF THE MANUSCRIPT

Figure 1 depicts the flow and organization of the manuscript. The introduction part discusses the importance of the LSTM, its applications, unique role in short-term traffic prediction and the background for short-term traffic prediction followed by the description of the basic architecture of the LSTM, basic concepts and operation. The short-term traffic prediction algorithms involving LSTM are then chosen, compared and classified into various categories; depending upon whether these algorithms use the single LSTM model or combine it with other algorithms. Finally, the challenges associated with the short-term forecast of traffic using the LSTM are highlighted, conclusions are drawn and future strategies are identified.

III. BACKGROUND OF SHORT-TERM TRAFFIC PREDICTION

The short-term traffic prediction helps in defining traffic routes for autonomous as well as human driven vehicles. It allows the decision to know traffic conditions such as number of vehicles passing a certain traffic route in a specific time, routes existence among various locations, defining trajectories, identifying warning and rescue signals of traffic, adapting secure and pollution free routes with minimal obstructions and knowing the overall future traffic conditions on the traffic routes to be followed, [17] to mention a few. In the context of modern intelligent transportation, knowing these parameters allows the selection of the optimal, safe, secure and travel-friendly routes by the drivers and auto-driving cars. This not only reduces the transportation cost but ensures overall efficient traffic management and resources utilization with minimized traffic fatality rate.

The road traffic varies with the locations of roads and their attributes, the number of vehicles, the speed of vehicles, interconnection of roads and the time of travel, among the many. Various external factors such as weather conditions, pollution and cultural events or festivals, also affect the flow of traffic. This shows that traffic patterns inherently vary in space and time or have spatio-temporal variations or characteristics that change in a dynamic fashion. In addition, when a vehicle is at a certain position, its journey towards a predicted position is dependent on a diverse set of parameters, as described above. This indicates that correlations exist among the spatio-temporal traffic patterns.

The short-term traffic prediction addresses the challenges associated with the intelligent transportation. They include,

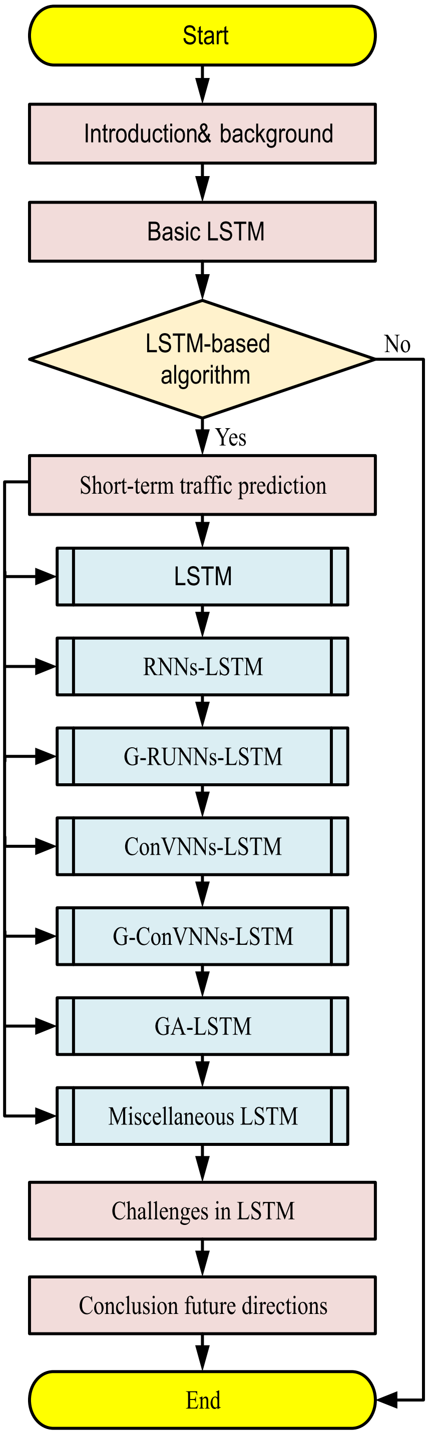

FIGURE 1. Flow chart of manuscript's organization.

for instance, processing of spatio-temporal traffic patterns for predicting traffic parameters: travel time, roads occupancy, road pollution, weather conditions during travel and identification of the best routes. Moreover, it provides guidelines for future traffic conditions that could be followed in anticipation to ensure overall efficient traffic management, resources utilization and reduced emergencies.

IV. THE ORIGIN OF LSTM: RECURRENT NEURAL NETWORKS (RNNs)

The LSTM model is a special case of the recurrent neural network (RNNs) [18]. Therefore, a brief description of the RNNs is given, which is then linked with the LSTM. Figure 2 depicts the basic architecture of RNNs [19]. It consists of an input matrix x, a hidden state (memory) matrix h and an output matrix o. At a specific time step t, these matrices, along with other parameters, are related by

$$h_t = Ux_t + Wh_{t-1} + b_i, \tag{1}$$

$$o_t = \phi_h(Vh_t + b_o), \tag{2}$$

$$\hat{y}_t = \phi_o(o_t), \tag{3}$$

where the weighting matrices are U, W and V, the respective biasing parameters at input and output are b_i and b_o and the respective hidden states at time steps t and $t-1$ are h_t and h_{t-1}. The parameters ϕ_h and ϕ_o are the respective activation functions of hidden and output stages and o_t and $\hat{y}_t$ are the respective output and predicted output at time step t. In training the RNNs, the error (or loss function) is computed and back-propagated to adjust the weighting and biasing matrices involved [20]. For certain activation functions (such as sigmoid and tanh), the derivative values of the error are smaller and their multiplication during back-propagation makes the overall gradient negligibly small, a phenomenon called vanishing-gradient problem. This challenges the learning of RNNs. In contrast, the derivative values get higher during back-propagation due to computation of high weight values (unlike the activation function values) that explodes the gradient. The LSTM and gated recurrent unit-LSTM (GRU-LSTM) are two of the techniques used for solving the vanishing gradient problem [21], [22] in RNNs.

V. THE BASIC LSTM MODEL

The LSTM model was developed by Sepp Hochreiter and Jurgen Schmidhuber in 1997 [23] and since then it has been used for many applications including traffic forecast. As depicted in Figure 3, the LSTM model belongs to the group of deep learning algorithms of machine learning, which further traces back to artificial intelligence. Figure 4 shows the basic architecture and working mechanism of the LSTM [23]. Its architecture is made up of three gates: forget, input and output gates. It has current and previous cell states (long memory) and the hidden states (short memory). The gates involved in the architecture of the LSTM are described one by one.

A. THE FORGET GATE

In this gate, at a specific time step t, an input data stream x_t and the previous hidden state h_{t-1} are processed by the sigmoid function σ to produce values in the 0 and 1 range in the form of the vector f_t. These values are then element-wise multiplied with the previous cell state c_{t-1} to decide whether or not to preserve the previous states. A 0 value corresponds to forgetting the previous cell state; as some new critical information is fed to the system while a 1 value means preserving

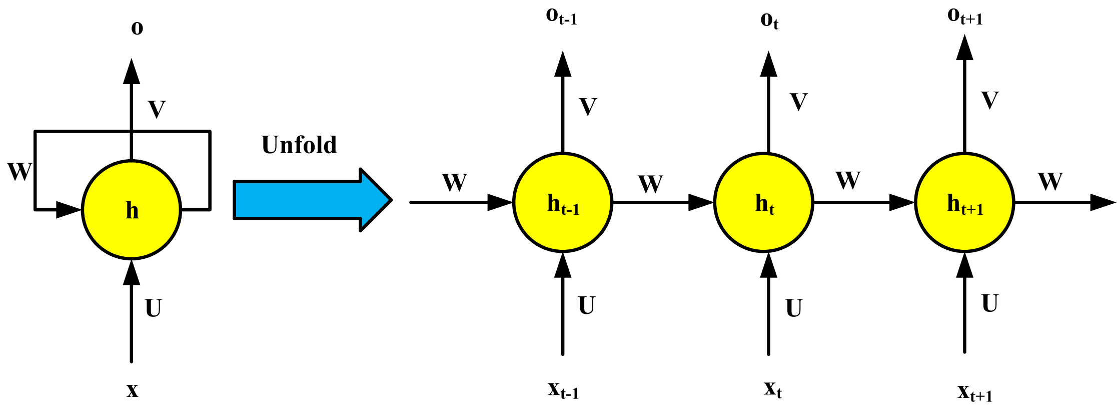

FIGURE 2. **The basic architecture of the recurrent neural networks (RNNs).**

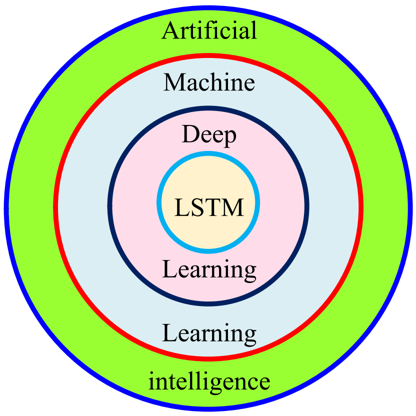

FIGURE 3. **The LSTM is one of the deep learning algorithms and belongs to the class of machine learning, which further originates from artificial intelligence.**

the previous cell state value. Mathematically, the operation performed by the forget gate is written as

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), \tag{4}$$

where W_f and b_f are the respective weighting and bias matrices.

B. THE INPUT GATE

This gate consists of the tanh and the second sigmoid functions whose output values are element-wise multiplied, which updates the cell state. The tanh function uses input data and the previous hidden state (short-term memory) and produces a vector $\tilde{C}_t$ having values in the -1 to 1 range. However, it does not contain any knowledge about the significance of the current input data (to pass further or retain). For this purpose, the sigmoid function output vector i_t is element-wise multiplied with $\tilde{C}_t$ and the result is added to the previous cell state to update the current cell state. The output values $[-1,1]$ of the tanh function, after multiplication with the sigmoid function, decide how much significance a current input data stream has to forget or pass in updating the new cell state. The mathematical description of the various vectors at the input gate is given by

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \tag{5}$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + bc), \tag{6}$$

where W_i and W_c are the respective weighting matrices at input gate of the sigmoid and tanh functions (the subscript c represents the cell state updating process). The parameters b_i and b_c represent the biasing factors corresponding to W_i and W_c, respectively.

C. THE OUTPUT GATE

This gate uses three vectors: C_t, x_t and h_{t-1} and produces the present hidden state h_t corresponding to the following mathematical relationships

$$o_t = \sigma(W_o[h_{t-1}, x_t]) + bo, \tag{7}$$

$$h_t = o_t \tanh(C_t), \tag{8}$$

where o_t is the sigmoid function's output with weighting matrix W_0 and b_o is the biasing factor with respect to W_o. The element-wise multiplication multiplies the corresponding elements of the matrices.

For prediction of short-term traffic data, the LSTM requires large datasets in accordance with the usual requirement by the deep learning models. To predict the short-term traffic, first the datasets are processed before input to the LSTM. This processing usually involves removing outlier data, computing the missing data, deleting the data from the failed sensors and normalizing the data for better presentation. The processed data are then divided into training and testing sets. The training datasets are input to the LSTM that processes them to extract the traffic information. The forget gate of the LSTM decides whether to allow or block the incoming data to the input gate depending upon their weight values. New data values are multiplied by 1 and are allowed to the input gate otherwise they are forgotten by multiplication with 0. The input gate of the LSTM processes the output of the forget gate that further normalizes the data and passes the data with high weight values to the output gate and updates the current

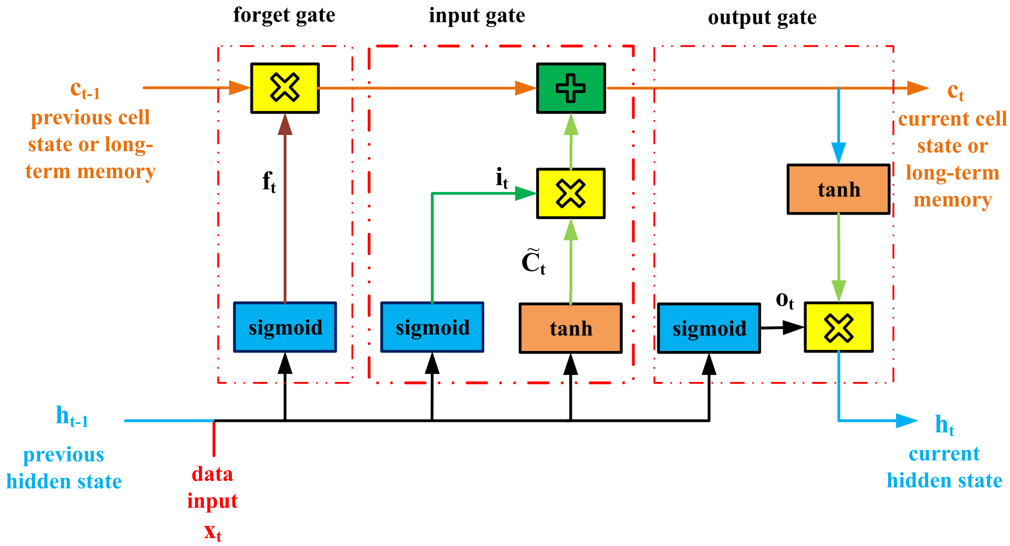

FIGURE 4. The basic architecture of the long short-term memory (LSTM) model.

cell state as well. The cell state is used by the output gate in combination with the input data and previous short-term values of the data to compute the short-term traffic values (output). In the testing phase, the unseen data are input to the LSTM that uses the learned patterns during the training phase to predict the output. Sometimes a data validation process is also used to tune the deep layers parameters of the LSTM before forecasting.

A number of traffic features; such as travel time, traffic flow and traffic volume (density), can be normalized and input to the LSTM that processes them and learns about the traffic flow patterns using the forget, input and output gates operation, as described above. This capability of the LSTM makes it superior than the traditional algorithms in forecasting traffic in diverse traffic scenarios.

VI. TRAFFIC FLOW PREDICTION USING LSTM AND ITS VARIANTS

This section classifies the algorithms and provides a description of each class. The algorithms are classified into various categories based on the use of the LSTM alone or combining one or more techniques with the LSTM for obtaining and processing the traffic features. These techniques include recurrent neural networks, gated recurrent neural networks, convolution operation, graph convolution operation and graph attention. In all the classes of the algorithms, the LSTM is used for the traffic prediction while the techniques combined with the LSTM are useful in capturing the traffic features and making the prediction process meaningful, effective and accurate. In addition, only the LSTM-based algorithms developed in the last three years are considered. This strategy provides a thorough and in-depth description of the LSTM-based techniques for traffic prediction rather than providing a general description of the deep learning techniques.

A. TRAFFIC PREDICTION USING LSTM

This section describes the algorithms using only the LSMT as the algorithm for traffic forecast. The approach designed in [24] uses the LSTM for data prediction that evaluates data input to the network and, depending upon the effectiveness of information, data are either discarded or retained in the network. The data are obtained for various city zones from the concerned department and aggregated into a five minutes interval in linear datasets. The LSTM algorithm is then applied to forecast the traffic flow and the error is estimated. A multivariate LSTM algorithm is proposed in [25] that considers a number of parameters such as flow of traffic, speed and occupancy, which provide better performance over the univariate schemes. An LSTM-based approach is used for speed forecast of vehicles in [26]. The datasets are obtained from a traffic simulator dealing with real world data. The model is evaluated for both univariate and multivariate features involved in traffic speed prediction. The authors in [27] develop an LSTM model to cope with the stochastic and time-varying nature of the traffic flow patterns, which involves dynamic traffic patterns forecasting than static patterns and conditions. The authors in [28] argued that sometimes a simple model also produces promising results with reduced complexity of computation. Therefore, an LSTM network for traffic forecasting combines the traffic features in time and space for traffic forecast. The authors in [29] argued that the conventional deep learning is one of the recent approaches for traffic prediction. However, due to its stochastic nature, error in data prediction follows, especially the data distribution imbalance over-fitting problem, in which the predicted model diverges from the true prediction behavior and involves high error. To cope with it, an algorithm is developed that uses a network that learns online by itself and maps the data. It learns and equalizes the statistics of data flow. Moreover, the uncertainty in the stochastic

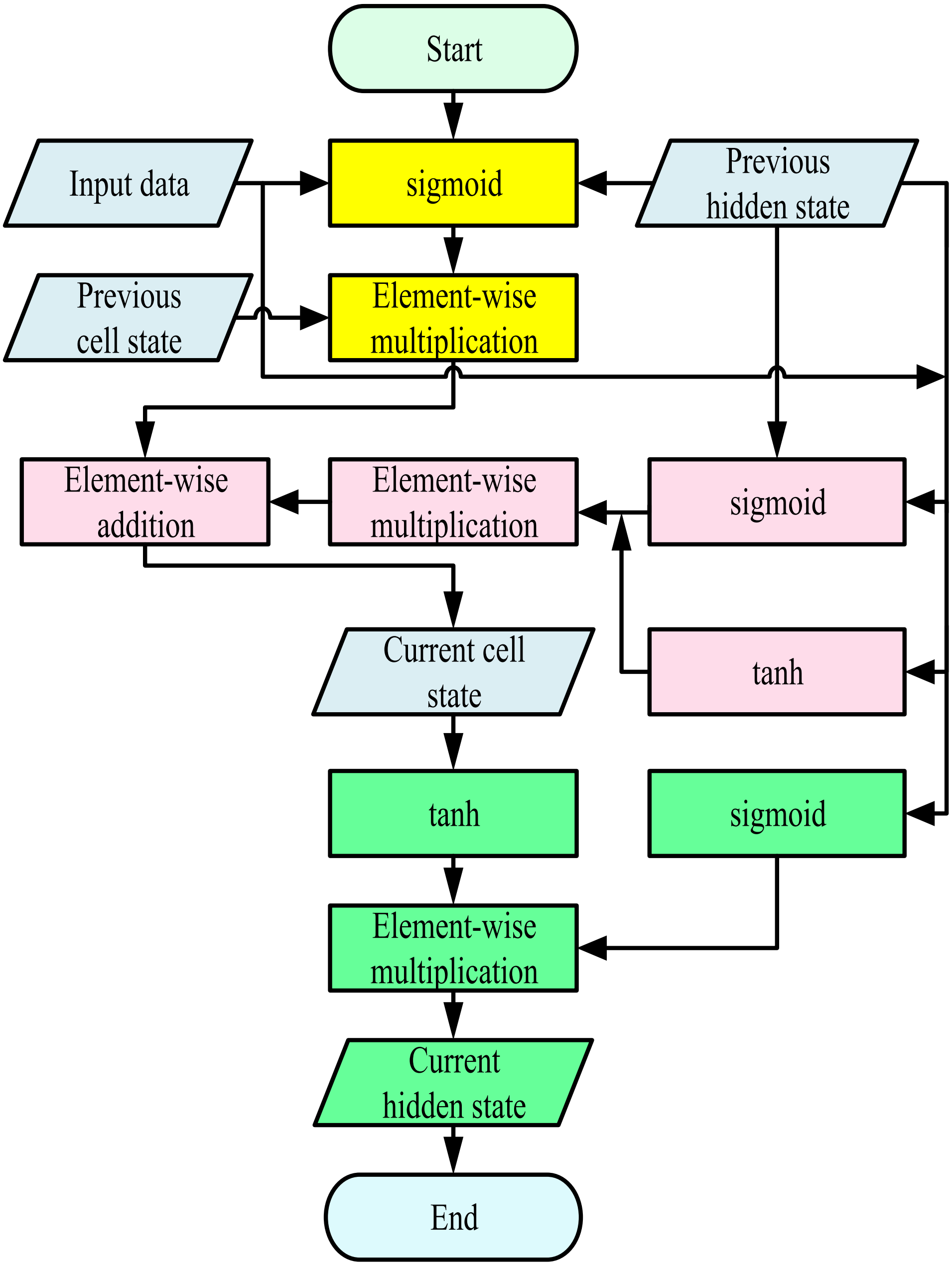

FIGURE 5. Flow chart of LSTM's operation. The yellow, pink and green blocks indicate the forget, input and output gates, respectively. The forget gate allows important data to pass to the input gate that normalizes the data and further passes it to the output gate for further processing to produce the output.

nature of data is reduced by the deep bidirectional LSTM (DB-LSTM). Finally, data forecasting is made by the LSTM. The sequences in which data flow at various time slots are identified and assigned different weights. Then a bidirectional LSTM (B-LSTM) mechanism is proposed that identifies periodic features in data flow on daily and weekly basis from both directions: previous data and the data obtained after processing the previous information.

The concept in [30] first macroscopically analyzes the real traffic lanes to obtain traffic data, which are then used for testing and training the LSTM that obtains better results than GRU and the stacked auto-encoders. The authors in [31] recorded the data of vehicles at an exit station on an expressway in Shanghai, China. The collected data are pre-processed, split, aggregated and normalized. Then the LSTM algorithm is applied for learning the data features in combination with traffic flow at the corresponding exit station. It is argued in [32] that traffic flow prediction based on the use of GPS data contains data coming from certain limited devices and has high cost as well. To overcome this problem, a travel mode classification method using cellular data is given that obtains travel mode characteristics, converts the

TABLE 1. Traffic flow prediction schemes using LSTM.

Reference	Main Algorithm	Achievement	Limitations	Year
[24]	Collects data, discards the irrelevant information and sizes them to five minutes intervals for prediction	Data prediction within a reasonable error range	Does not evaluate the non-linear dataset models	2020
[25]	Traffic forecast is obtained by a multivariate algorithm that involves traffic flow, vehicles speed and occupancy	Achieves better prediction results than univariate algorithms	Involvement of a large number of traffic parameters is linked with computational complexity	2020
[26]	Uses the LSTM algorithm for vehicles' speed prediction involving univariate and multivariate features	Provides a comparison of the univariate and multivariate features	Multivariate features better in prediction than univariate features but are more complex	2021
[27]	Addresses the stochastic and time-varying nature of traffic flow to forecast it with the LSTM model	Easy implementation due to the simple LSTM model	Application to diverse traffic networks is challenging	2020
[28]	Uses a simple LSTM network for traffic forecasting	Easy to implement and less complex than the algorithms involving multiple steps	Struggles with traffic forecast for diverse and complex traffic conditions	2021
[29]	Uses delay based deep learning with LSTM and bidirectional LSTM	Overcomes the false prediction (error) in forecasting data	Computational complexity issue as it combines multiple steps	2020
[30]	Obtains real traffic datasets and applies LSTM	Obtains better results than GRU and stacked auto-encoders, ease of implementation	Less capable of features extraction in more diverse traffic conditions	2021
[31]	Collects, pre-processes, splits and normalizes data collected at an exit station of vehicles in Shanghai and the features are then learned by the LSTM to predict traffic flow	Improvement in accuracy	Accuracy based on multiple processing steps	2021
[32]	Uses travel mode classification based on cellular data, obtains travel mode characteristics, converts the time information in the datasets into time characteristics and inputs them together with traffic flow to an LSTM for prediction	Improved accuracy of prediction	Requires consistency in input data for effective data prediction	2021
[33]	Combines the noise pollution with time-series data to predict traffic flow	Effective for urban roads and industrial areas where noise pollution is of serious concern	Does not incorporate the spatial traffic flow features	2021
[34]	Uses big data approaches of Spark and Kafka with the LSTM to process high streams of data for traffic flow prediction	Effective for processing traffic data of complex traffic networks	Requires huge resources due to processing high streams of data	2020
[35]	Combines rainfall and traffic data to forecast the traffic flow	Effective for roads with severe weather conditions	Restricted to snowfall areas	2020
[36]	Uses periodic events or context of locations (population based) for identification of anomalous traffic and forecasts the resulting traffic	Good in identifying the anomalies in traffic flow with periodic and population based characteristics of locations	Applicable to specific periodic events	2021
[37]	Predicts driving intention and trajectory at intersections	Effective for self-driving vehicles and congested networks with intersections	Requires advance knowledge of trajectories network and their statistical connections and effects on one another	2021
[38]	Considers the interactive autonomous vehicles that share their cell and hidden states and then evaluates the output states from which the surrounding trajectories are predicted	Effective for finding the travelling trajectories in advance	Requires complex requirements of memory and processing for large traffic networks	2019
[39]	Proposes a time-series forecasting based on few-shot algorithm that does not require historical trends of data with large dataset	Traffic prediction with few data input values	The LSTM algorithm used is accurate mostly when the large amount of data are used to train it	2022
[40]	Obtains traffic patterns and analyzes them at the previous and later states for future prediction with the LSTM providing the most accurate prediction among the five compared techniques	High accuracy	Further validation of results required for complex traffic conditions	2022

time information in the datasets into the time characteristics. Then it inputs them with traffic flow data into the LSTM for improved prediction accuracy. The authors in [33] combined the noise pollution with traffic data as input to the LSTM to forecast traffic at major roads of Mardid, Spain. Previously, this approach has been combined with air pollution for traffic prediction. The authors in [34] used the big data frameworks of Spark and Kafka for traffic forecast in real time using the LSTM with an effective prediction accuracy. It is mentioned in [35] that environmental parameters such as snowfall, rainfall, fog and wind affect the eyesight of the drivers, the density and speed of vehicles. Therefore, the effect of road traffic due to rainfall is studied using data from roadside installed sensors. The data are processed by the LSTM and recurrent neural networks with the LSTM providing the best prediction accuracy. The LSTM model is applied in [36] to note an increase or decrease in the traffic flow based on an event (periodic events such as cultural or political events) and context (tourist towns or mountain towns) and, therefore, identify the anomalies in flow of traffic. The sensors placed at entry and exit points of highways provide the data. The authors in [37] predict driving intention and trajectory prediction at intersections. A large number of trajectories, their clustering and paths are statistically analyzed using the trajectories model of intersection priority. This is combined with the fitted probabilistic density model that approximates the distribution of a trajectory of interest and then uses it as a standard to evaluate the credibility. The LSTM then forecasts the intention and is also used for trajectory forecasting at early stage. The structural-LSTM proposed in [38] predicts the trajectories of surrounding roads for autonomous vehicles. It treats every interacting vehicle as a single LSTM cell and shares its cell state and hidden state with its spatial neighbors through a radial connection and evaluates its own and the neighbor output states. These output states are then used for the prediction of surrounding trajectories. The authors argue in [39] that the traditional time-series forecasting algorithms use large values of historical data to forecast the output. Therefore, these algorithms struggle in performance when a huge amount of data is not available. To overcome this problem, an algorithm is proposed based on the time-series concept of few-shots that uses the Siamese twin network approach to compute the time-series difference between data pairs rather than using the established trends of data. The data types not observed during the training phase of the LSTM are also learned. The LSTM is compared with the forecast algorithms that use the concepts of back-propagation, classifying and regressing the trees, the geographically closest neighbor and the regression based on a support vector in [40]. The LSTM provides the best accuracy among the compared techniques over a number of traffic conditions. Table 1 summarizes the key aspects of the algorithms using LSTM as the major forecast algorithm. Figure 5 shows the flow chart of the working of the LSTM. The forget gate allows specific important data to the input gate that applies sigmoid and tanh functions and element-wise addition and multiplication. The data from the input gate are further processed by the output gate to produce the final output.

B. TRAFFIC PREDICTION USING RECURRENT NEURAL NETWORKS LSTM (RNNs-LSTM)

This section describes the algorithms combining the RNNs and LSTM (RNNs-LSTM) for effective prediction of the short-term traffic. The authors in [41] argued that the traditional forecasting algorithms use continuous data input from history to perform the prediction decisions that do not accurately calculate the time delay. Therefore, an algorithm is proposed that uses three memory blocks that work together in a multiplicative manner to best calculate dime delay. Unlike the algorithms that consider only traffic conditions and use single data processing mechanism, weather conditions are combined with LSTM and GRU modules to design an algorithm that uses a hybrid neural network with a queue in [42]. Optimization of the results is accomplished to accurately predict the traffic conditions. The authors in [43] combined the tensor analysis with RNNs and developed three algorithms: tensor-based-LSTM, tensor-base GRU and tensor-based vanilla RNNs. These algorithms use tensors as high order input and output parameters. The weights are compressed and the traffic flow is predicted. The authors in [44] combined federated learning with RNN-LSTM, the former performed the model interaction and data privacy protection while the latter performed traffic forecast. The authors compared the three structures in [45]: LSTM, GRU and stacked-auto-encoders and concluded that the LSTM error is minimal for specific road traffic conditions. Table 2 summarizes the key aspects of RNNs-LSTM algorithms (see [41], [42], [43], [44], [45]).

C. TRAFFIC PREDICTION USING GATED RECURRENT/RECURSIVE UNIT NEURAL NETWORKS LSTM (G-RUNNs-LSTM)

This section discusses the LSTM algorithms combined with the gated recurrent/recursive unit. The GRU, as mentioned above, is one of the techniques designed to cope with the RNNs inherent issue in which the gradient vanishes and the learning of the networks becomes negligible. The gated recursive unit is a generalization of the gated recurrent unit (suitable for hierarchical patterns than sequences) [46]. A brief description of the GRU is first given to provide an insight to the way it overcomes the vanishing gradient problem. The Figure 6 shows the basic architecture of the GRU having the reset and forget gates. It does not include the cell states in output computation rather uses only the hidden states due to which it is faster than LSTM.

1) THE RESET GATE

The reset gate controls the flow of information between the previous hidden state (memory) h_{t-1} and the present input x_t. It decides the extent by which past information is allowed to pass or forget and is governed by the following mathematical model [47], [48]:

$$r_t = \sigma(W_r[h_{t-1}, x_t] + b_r), \tag{9}$$

where r_t is the memory rate, W_r and b_r are the respective weighting matrix and biasing factor.

2) THE UPDATE GATE

The update gate forgets or passes the information from the past for processing with the current information. This controls the vanishing gradient problem in that the information processing from the past is controlled. The mathematical formulation of the current hidden state h_t (output) is computed using the update gate and other operations as [47]:

$$z_t = \sigma(W_u[h_{t-1}, x_t] + b_u), \tag{10}$$

where W_u and b_u are the respective weighting and biasing matrices and z_t is the sigmoid functions' output. The other mathematical operations involved in the output computation are:

$$h_t = \tilde{h}_t z_t + h_{t-1}(1 - z_t), \tag{11}$$

$$\tilde{h}_t = tanh(W_o[r_t h_{t-1}, x_t] + b_o), \tag{12}$$

where W_o and b_o are the respective weighting and biasing vectors and $\tilde{h}_t$ is the output of the activation function corresponding to the update gate. The rest of this section follows the description of the G-RUNNs-LSTM algorithms.

The authors in [49] first used the LSTM to obtain the characteristics of traffic in space and time followed by a bidirectional GRU (Bi-GRU) having positive as well as negative feedback mechanisms. The shortcomings of the common optimizer are overcome by the rectified adaptive algorithm for further accuracy. The cosine, scientific and model learning control the algorithm's convergence. The authors in [50] first reduced the error in the traffic actual and sampled data using a piece-wise and constant co-efficient scheme. Then a gated recursive unit LSTM is applied to forecast the flow of trucks. The RNNs are considered for the traffic flow modeling in [51] with the LSTM model is used to improve the former and solve its problem of vanishing gradient. An attention mechanism is further introduced to ensure engineering of traffic features and improve the prediction accuracy. Table 2 summarizes the key aspects of G-RUNNs-LSTM algorithms ([49], [50], [51]). Figure 7 shows the flow chart of the operation of the G-RUNNs-LSTM. The yellow and pink blocks show the respective reset and update gates of the GRU whose output is fed to the LSTM for traffic forecast.

D. TRAFFIC PREDICTION USING CONVOLUTIONAL NEURAL NETWORKS LSTM (ConVNNs-LSTM)

This section first describes the basic architecture of the ConVNNs-LSTM followed by their combination with the LSTM for forecasting short-term traffic. The ConVNNs are effective in combining the spatio-temporal features in predicting traffic behavior rather than considering a single feature [52]. They are also efficient for large size networks. Figure 8 represents the fundamental architecture of the convolutional neural network [53]. The convolutional layer processes the input with a certain filter to extract information (or

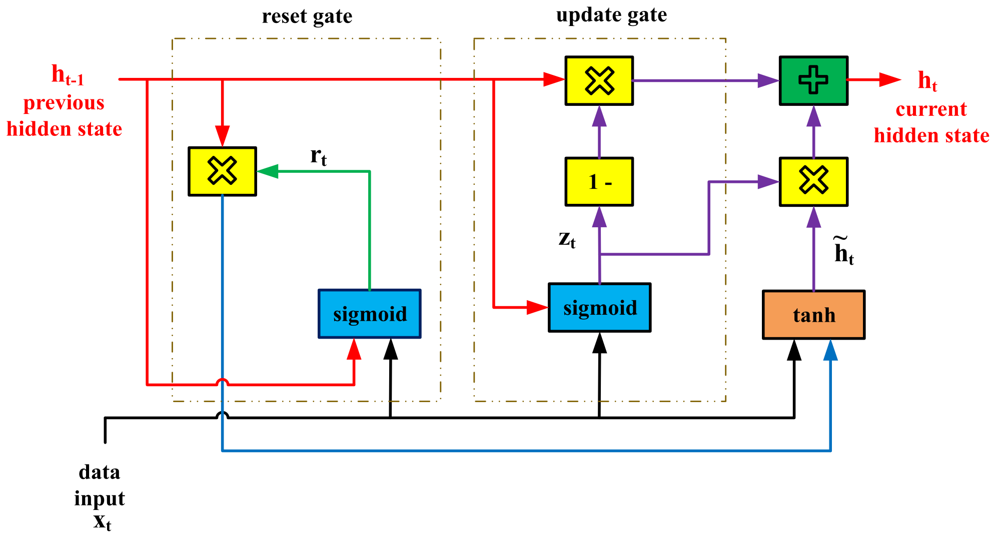

FIGURE 6. The basic architecture of the gated recurrent unit (GRU).

features map). The mathematical operation of the convolutional layer is governed by [54]:

$$O(i,j) = S * F = \sum_{k=1}^{L}\sum_{l=1}^{L} S(i+k-1, j+1-l)F(k,l), \quad (13)$$

where S is the input signal (for example a 2D image) having m rows and n columns, F is the filter or kernel of size $L \times L$, the symbol $*$ shows convolution operation, $O(i,j)$ is the element of the output matrix at the i^{th} row and j^{th} column, assuming a single channel for the convolution. The values of i and j vary from $m - L + 1$ to $n - L + 1$, respectively. All the elements of the output matrix are summed to generate the features map M_f (for the assumed single channel) as:

$$M_f = \sum O_f(i,j). \quad (14)$$

The pooling layer acts on the signal and sub-samples it to reduce its dimension. For the case of 2D image as the input signal, the pooling operation is mathematically written as:

$$O(i,j) = max_{i \le l \le i+R-1, j \le k \le j+R-1} S(l,k), \quad (15)$$

where $O(i,j)$ is the element of the output matrix at the i^{th} row and j^{th} column, R is the size of the filter used by the pooling, which extracts the maximum values of the features (hence max pooling, where average pooling is also used). The fully connected layer connects all the features maps obtained in the previous (convolution and pooling) layers to construct a bigger features map (a one-dimensional array or vector), which is modeled as:

$$O_{fc} = f(W[S]) + b), \quad (16)$$

where S is the input signal vector that represents the features collected by the prior layers, W is the weight vector, b is the bias vector, f is the activation function and O_{fc} is the output of the fully connected layer.

The design in [55] used ConVNNs in combination with the LSTM model to extract the prominent aspects of traffic data in time and space, which are then used for traffic forecast. An algorithm is proposed in [56] for multi-lanes traffic that combines multiple features of data traffic, considers the routing among the lanes, traffic flow in both directions and predicts the traffic in a recursive manner. In [57], it is argued that majority of the deep learning approaches for traffic forecast do not explore data features patterns of intra and inter-day nor they correlate the parameters with traffic flow such as weather conditions. Therefore, an approach is adopted in which the ConVNNs capture the inter and intra-day patterns of traffic, which are fed to the LSTM to acquire traffic features in time and perform the traffic forecast. In [58], a hybrid traffic-flow forecasting model (HTFM) is proposed in which the spatio-temporal characteristics of the vehicles data flow on a large scale are addressed. The time characteristics are learned by the LSTM mechanism while the space characteristics are obtained by the ConVNNs. These two models are then combined and the obtained characteristics are analyzed for data prediction. The temporal and spatial features are correlated by the maximum information coefficient in [59] to forecast the traffic. The authors in [60] design an algorithm that used ConVNNs to derive the historical data traffic features while features exhibiting periodicity are acquired by the LSTM. The wavelet-based transformation decomposes frequency components of the data for obtaining the useful information. The authors in [61] argued that existing traffic flow prediction models usually consider roads in a specific region. Therefore, a road-network is considered based on high-order spatio-temporal road conditions and a generative adversarial nets (TrafficGAN) algorithm is developed. The spatio-temporal features of the traffic flow for days and weeks are analyzed in [62] to forecast traffic speed. The authors in [63] state that most approaches use the hourly data to predict the traffic for

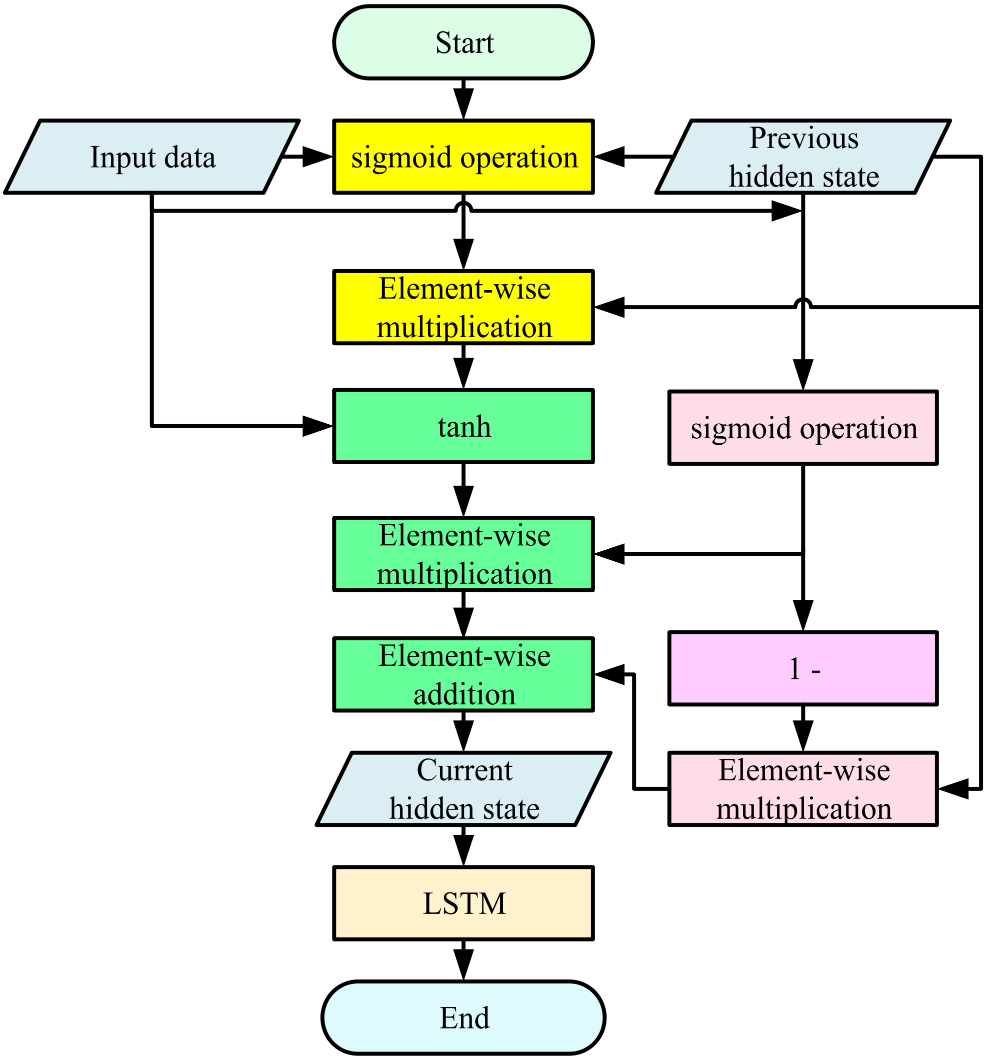

FIGURE 7. Flow chart of G-RUNNs-LSTM. The yellow and pink blocks show the reset and update gates of the GRU, respectively, whose output is fed to the LSTM for forecast.

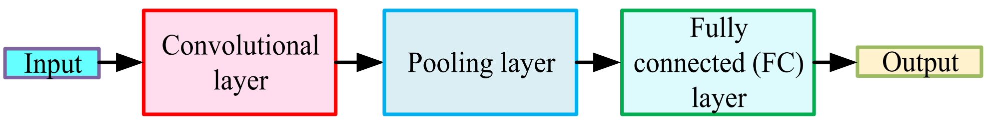

FIGURE 8. The basic architecture of the convolutional neural networks (ConVNNs).

the upcoming hours and that is not effective in forecasting daily and weekly traffic. Therefore, a design is utilized that uses multiple convolutional-LSTM networks to obtain traffic features for days and weeks. An LSTM network is also used for predicting the output in a sequential manner. Every input sequence is assigned a weight in response to its role in output prediction. A snapshot of the traffic map is taken in [64] from an online source and then a combination of ConVNNs, LSTM and transpose ConVNNs (TCNNs) is made to obtain the spatio-temporal characteristics of the traffic and use them for predicting traffic congestion. The measured speed of a vehicle is used in [65] to forecast the driving speed and weather conditions for six hours in mountains. The weather conditions are taken from weather stations. The driving speed of the vehicles and the weather conditions forecast are input to the proposed algorithm and processed by the multi-scale hybrid convolutional LSTM. Moreover, model training is performed by the established data trends and is further applied it to real-time forecast. The effect of the implementation of a number of decomposition algorithms are studied to analyze the performance of neural networks [66]. A decomposition algorithm decomposes traffic flow parameters into various

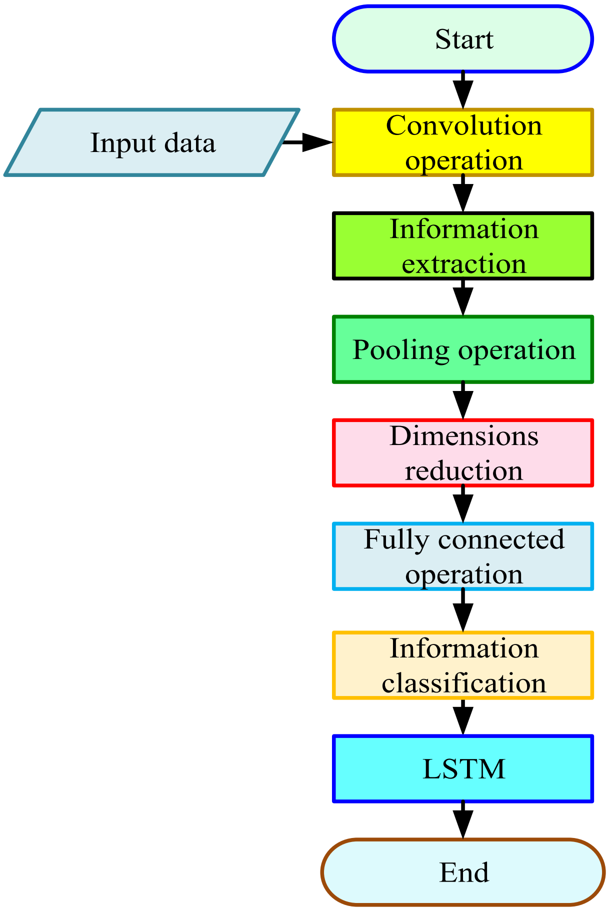

FIGURE 9. Flow chart of ConVNNs-LSTM showing the convolution operation to extract features followed by pooling and fully connected layers for dimension reduction and classification, respectively. The features are fed to the LSTM for forecast.

components that are further clustered into volatile, periodic and residual components. These components were then subjected to a bidirectional LSTM to predict traffic flow. The authors used the convolutional LSTM (ConV-LSTM) in [67], in which the time features are extracted by the LSTM and the spatial features are obtained by the convolutional part. The combined features are further processed to obtain the traffic prediction. The convolutional part is operated to prepare the input data that are then processed by the LSTM layer instead of the pooled layer of ConVNNs. The historical, seasonal, weather and pollution features of data traffic are combined for high speed traffic in [68] that designs a convolutional bidirectional deep LSTM model for accurate prediction of traffic in complex traffic networks. These features provide traffic prediction over a diverse set of possible traffic conditions. A function for the measurement of the spatial traffic features is modeled in [69] that quantifies the correlation among the traffic flow parameters. It is followed by the learning of the spatial features by the convolutional LSTM (ConV-LSTM). Rules for the ensemble diverse convolutional LSTM are established to robust the prediction algorithm.

The optimization of the weights of the ensemble elements is performed and the variations in the traffic are tracked. A regression analysis model of vectors in [70] first evaluated the correlations among the variables of traffic. Then a convolution LSTM is used to predict traffic speed with multiple features for a single location first and then for multiple locations later. It is also revealed that forecast accuracy varies with the spatial variations in parameters of traffic. The authors in [71] argued that the existing deep learning models for forecasting short-term traffic have the accuracy constraints beyond a certain threshold due to the presence of noise, insufficient extraction of features and aliasing in the components of the data signals. Therefore, an algorithm is proposed that decomposes the traffic modalities by the variational mode decomposition and redistributes them by the self-attention approach, which has improved accuracy. The authors in [72] argued that the ConVNNs cannot retain the information in time after the first layer. In addition, the retaining of short-term spatio-temporal information by the 3D ConVNNs is not effective for retaining the long-term information. The LSTM can extract the long-term information, it cannot extract the spatial information extraction single-handedly. Therefore, a framework is given that makes use of the 3D ConVNNs for extracting the short-term correlations in space and time and the ConV-LSTM for extracting the corresponding long-term correlations. Table 2 summarizes the key aspects of ConVNNs-LSTM algorithms ([55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70], [71], [72]). Figure 9 shows the flow chart of the the ConVNNs-LSTM in which the convolution operation extracts features, the pooling layer reduces the dimensions of the features and the fully connected layer classifies the features that are then fed to the LSTM for forecast.

E. TRAFFIC PREDICTION USING GRAPH CONVOLUTIONAL NEURAL NETWORKS LSTM (G-ConVNNs-LSTM)

This section describes the algorithms using the graph convolutional neural networks with the LSTM (G-ConVNNs-LSTM). The G-ConVNNs part functions similar to ConVNNs but it uses graph structures as data input [73], where a graph $G = (V, E)$ represents the flow of data over the vertices set $V = \{v_1, v_2, \ldots, v_n\}$ and the set of edges $E \subseteq V \times V$ with p and q being the total number of respective vertices and edges [74]. A feature adjacency matrix A of dimension $p \times p$ also represents the graph with $A_{i,j} = 1$ if an edge exists from vertex v_i to vertex v_j else $A_{i,j} = 0$. The output H_{l+1} of the l^{th} layer of the G-ConVNNs is modeled as [75]:

$$H^{l+1} = \sigma(\hat{D}^{-1/2}\hat{A}\hat{D}^{-1/2}H^{l}Wl), \tag{17}$$

where σ is the activation function, $\hat{A} = A + I_N$, I_N is the identity matrix and $\hat{D}_{ii} = \sum_j \hat{A}_{ij}$ is the diagonal node degree matrix of $\hat{A}$ and represents the number of edges for a node i. The W^l is the trainable weight matrix of the l^{th} layer and H^l is an $N \times F$ features matrix with N representing the number of nodes and F expressing the features of each node in the l^{th} layer.

According to spectral graph theory, an undirected graph is represented by the normalized Laplacian $\hat{L}$ as [76]:

$$\hat{L} = I_N - \hat{D}^{-1/2}\hat{A}\hat{D}^{-1/2} = U\Lambda U^T, \quad (18)$$

where U is an orthogonal matrix that consists of the eigenvectors of $\hat{L}$ and Λ is a diagonal matrix consisting of the eigenvalues of $\hat{L}$. The graph convolution operation is defined, in the spectral domain as [76]:

$$x *_g F = U(U^T x \odot U^T F) = U\Lambda U^T x, \quad (19)$$

where $*_g$ represents the graph convolution operation between the input signal x and the kernel (or filter) F, $U^T x$ is the Fourier transform of x and $\odot$ shows the Hadamard product. The convolution used by the G-ConvNNs may also be applied in different domains such as using spatial domain or wavelet transform (among the many) [77].

The authors in [78] argued that knowing the traffic features in space and time is necessary in predicting behavior of the traffic. Temporal characteristics are significantly explored by most of the algorithms while spatial characteristics are not significantly explored due to their variations and complexity. In this direction, an algorithm is designed that combines the various features of vehicles such as the distance, distance covered in a unit time and the angle. These features are combined in multi-weighted adjacency matrices and the similarities in the features are learned. Temporal dependencies are learned by inputting the output of the algorithm to an LSTM model. A graph convolutional network and LSTM algorithms are combined in [79] to forecast traffic data. The graph convolution connects the vehicles in the form of graph structures and the correlation and dependencies are determined among the graphs, which are then passed to the LSTM that learns and predicts the traffic trends. A traffic graph convolution LSTM (TGC-LSTM) algorithm is proposed in [80]. The network is considered as a graph and interaction among the roadways is considered in the network. The graph is described in terms of its physical topology and various convolutional weights are assigned to the different physical features and a loss function is computed.

In [81], an attentive traffic flow machine (ATFM) algorithm is proposed that consists of two convolutional LSTM parts. The first part takes traffic input and processes it to generate traffic flow features and the second part processes them to obtain dynamic weighted spatio-temporal characteristics. Moreover, the short and long term traffic predictions are made by interacting with the sequential and periodic data and the external parameters that influence them. The authors in [82] proposed a framework that involves dynamic and non-linear data with high complexity. The network is considered as a graph and the spatio-temporal characteristics are obtained and predicted by ConVNNs. The authors in [83] argued that the existing traffic prediction algorithms mainly focus on time correlations of traffic data and ignore the correlations in space as well as their mutual correlations, which affects the prediction by not providing a through correlational analysis of data features. These issues are addressed by designing an algorithm that reduces data dimension by principal component analysis. The graph convolution network algorithm learns network's topology and the LSTM correlates the data features in time. The authors in [84] applied spatial densities to the LSTM and evaluated the missing values of the densities using interpolation. They concluded that the spatial densities are better than the speed input in predicting traffic congestion and travel time. The interpolation process inserts extra computational time to determine the missing values. An attention based graph Bi-LSTM (AGBN) algorithm is developed in [85]. It uses graph convolution networks for obtaining spatial features and Bi-LSTM temporal features. Table 2 summarizes the key aspects of the G-ConVNNs-LSTM algorithms ([78], [79], [80], [81], [82], [83], [84], [85]).

F. TRAFFIC PREDICTION USING GRAPH-ATTENTION-BASED LSTM (GA-LSTM)

The GA-LSTM algorithms use the graph data structures at the input and an attention layer that adaptively assigns weights to the data depending upon the data they process. The graph attention layer captures prominent features from the data for future ease of use [86] and uses a set of features $h = \{\vec{h}_1, \vec{h}_2, \ldots, \vec{h}_N\}$ of N nodes with each node having features F to produce a set of new features $h' = \{\vec{h'_1}, \vec{h'_2}, \ldots, \vec{h'_N}\}$ with each node having features F' [87]. For any two nodes i and j in the graph, the features of these nodes are linearly transformed by a weight matrix W of dimension $F \times F'$ and then a self attention mechanism in the form of a parameterized weight vector a is applied to obtain the attention co-efficients e_{ij} as [87]:

$$e_{ij} = a(W\vec{h}_i, W\vec{h}_j), \quad (20)$$

which is computed for every node j in the first-link neighbors set N_i of node i and then normalized for every neighbor j of i by the softmax function as [87]:

$$\alpha_{ij} = softmax_j(ij) = \frac{exp(e_{ij})}{\sum_{k\in N_i} exp(e_{ik})}, \quad (21)$$

where α_{ij} is the normalized attention co-efficient, which is further processed by the attention mechanism using leakyReLu function as [87]:

$$\alpha_{ij} = \frac{exp(leakyReLU(\vec{a}^T(Wh_{conc})))}{\sum_{k\in N_i} exp(leakyReLU(\vec{a}^T(Wh_{conc})))}, \quad (22)$$

where $\vec{a}^T$ is the transposition of the parameterized weight vector $\vec{a}$ of dimension $2F'$, $Wh_{conc} = W\vec{h}_i \parallel W\vec{h}_j$ and $\parallel$ is the symbol of concatenation (conc) operation that transfers the learned features from one layer to another. Figure 10 shows the operation of graph-attention while Figure 11 shows the flow chart of GA-LSTM.

A GA-LSTM algorithm is proposed in [88] in which the traffic network is an un-weighted directed graph and the

TABLE 2. Traffic flow prediction using RNNs-LSTM [41], [42], [43], [44], [45], G-RUNNs-LSTM [49], [50], [51], ConVNNS-LSTM [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70], [71], [72] and G-ConVNNs-LSTM [78], [79], [80], [81], [82], [83], [84], [85].

Reference	Main Algorithm	Achievement	Limitations	Year
[41]	Dynamically uses three multiplicative memory blocks to predict traffic	Does not require data history as a continuous input	Error sensitivity due to the multiplicative nature of the involved memory blocks	2022
[42]	Combines weather conditions with GRU and LSTM followed by optimization process	Highly effective for traffic prediction in rapidly changing weather conditions	Involves several computational steps	2021
[43]	Combines tensor analysis with RNNs to predict traffic flow for multi-features traffic parameters	Accuracy of prediction for complex traffic networks	Features extraction is challenging in complex traffic and requires sophisticated techniques	2021
[44]	Combines federated learning with RNN-LSTM for traffic data prediction with data privacy	Effective for traffic conditions requiring security	Extra computational tasks to perform	2022
[45]	Compares three RNN units: LSTM, GRU and stacked auto-encoders	Advocates the low prediction error of LSTM in certain traffic conditions	Variation in results of the algorithm is dependent on the considered specific traffic conditions	2021
[49]	Obtains spatial and temporal features with LSTM and GRU and then uses a bi-directional GRU to complete the prediction process along with controlling the convergence rate of the algorithm to a local maximum by the rectified adaptive algorithm, cosine, scientific and model learning	High accuracy due to various steps for features obtaining and controlling the algorithm convergence	Sensitivity of controlling the algorithm convergence	2011
[50]	Uses a data expansion phase to reduce the errors in actual and sampled data from GPS before prediction	Pre and post data expansion and prediction strategies for prediction accuracy	Complexity of processing in switching between post and pre data processing phases	2020
[51]	Considers RNNs for traffic flow modeling and uses LSTM to deal with the vanishing gradient issue in data prediction supported by a learning mechanism	Improved prediction accuracy and ease of training LSTM with an attention mechanism	RNNs tend to lose information weight for high data streams	2021
[55]	Data features are identified in time and space and periodic data sequences are assigned different weights followed by processing previous and the newly generated data values for forecasting	Ease of traffic flow forecasting based on periodicity of data flow	Processing delay due to computation of previous and newly computed data values (bidirectional)	2020
[56]	Predicts traffic for multi-lanes by considering routing among the vehicles in the lanes, the bidirectional traffic and recursive data processing	Promising forecasting for multi-lane highways	Data congestion/interference due to multi-lanes traffic and routing	2021
[57]	Uses inter and intra-day traffic patterns with extracted features and their correlation with the weather conditions	90 % prediction accuracy	Acquisition of extra intra-days information is required	2020
[58]	Obtains the features of a large scale traffic flow in space and time and analyzes them to perform data traffic prediction	Accurately predicts data flow patterns due to their time and space characteristics on a large scale	Processing large scale data increases resources utilization and delays the real time implementation necessary for congested highways	2021
[59]	First obtains the space and time features of the traffic and then uses the maximum information coefficients to correlate the extracted features and predict parameters	Conveniently detects and predicts traffic fluctuations	Lacks in time-efficiency for dense traffic conditions with rapid fluctuations	2020
[60]	Correlates the traffic transfer patterns among roads and their spatio-temporal characteristics for traffic flow prediction	Suitable for traffic transfer among the roads and its prediction	Requires extra processing for connecting traffic transfer among the roads	2020
[61]	Predicts traffic for a network of roads using the generative adversarial net based on convolutional and LSTM networks	Effective to predict traffic for a varied set of road conditions	Requires high data processing due to involvement of the network of roads	2019
[62]	Combines the ConVNNs and LSTM to predict traffic speed based on daily and weekly periodicity of data	Promising performance in traffic conditions with periodicity in exhibited characteristics	Compromised performance in fluctuating traffic conditions	2020
[63]	Involves daily and weekly traffic data than hourly data to input to the convolutional LSTM networks and predict the traffic speed	Traffic speed prediction based on daily and weekly data	Multiple LSTM networks and other processing steps are involved	2021
[64]	Combines three sub-algorithms: ConVNNs, LSTM and transpose ConVNNs to predict traffic for data in the form of images of traffic map obtained from a web-server with open source working mechansim	Effective in obtaining accurate data traffic patterns due to processing of the data by three different methods	Has high data processing delay	2020
[65]	Uses multi-scale hybrid convolutional LSTM to predict mountain traffic and weather conditions by taking weather conditions information from weather stations	Provides real-time traffic forecast during mountain driving	Dependency on data from mountain stations is challenging in some areas with connectivity issues	2020
[66]	Studies the effect of various decomposition algorithms on traffic prediction of neural networks	A thorough analysis of the effect of decomposition algorithms on neural networks	Combining multiple steps delays the output forecast, especially in real-time traffic	2021
[67]	The concepts of ConVNNs and LSTM are combined	Better accuracy than LSTM, two-layer LSTM and bidirectional LSTM	Takes extra steps than traditional LSTM	2020
[68]	Combines various conditions to predict traffic over a diverse set of possible traffic conditions using a convolutional bi-directional LSTM	Suitable for diverse traffic conditions	High complexity of computation	2022
[69]	A function measures the correlation among the traffic spatial parameters and processes by the convolutional LSTM for learning and prediction followed by establishing rules for the ensemble parameters and optimizing their weights in response to the variations in the road network flow	Improved accuracy	Requires high processing time due to the involved features and processes	2021
[70]	First uses the vector auto-regression to correlate traffic variables and then the convolutional LSTM is applied for prediction at a single and multiple locations	Improved accuracy due to finding the relationship among traffic variables before the training and learning processes	Generalized prediction applied to single locations	2022
[71]	Combines ConVNNs-LSTM with variational mode decomposition having a mechanism of self-attention to decompose and weigh the traffic parameters predicted by the later algorithm	Effective in accurate traffic prediction	Complexity of operation	2022
[72]	The short and long-term traffic features are obtained by the respective ConVNNs and ConV-LSTM methods	Combines the advantages of short and long-term features at the same time	Time-consuming features extraction makes it inefficient for real-time forecast	2022
[78]	Combines various features of traffic data, assigns them weights and then similarities are learned to classify and identify them	Accurate in traffic flow due to combination of various features	High delay in computational steps	2020
[79]	Considers road vehicles as nodes in a graph to obtain the data and processes them by deep learning algorithm to obtain forecast results	Data prediction efficiency due to reduction in data size	Requires constant information of geographical positions of vehicles as this information changes with the mobility of the vehicles	2020
[80]	A network is considered as a graph and the convolutional weights are assigned to the data features	Effective in capturing the specific traffic flow patterns	Physical characteristics of the network increase computational complexity	2019
[81]	Inputs traffic data to two convolutional LSTM modules to get weighted spatio-temporal characteristics and combines the periodic and sequential data and their external influencers to predict traffic flow patterns	Accurate in dealing with data prediction involving periodic and sequential datasets	Struggles in diverse data traffic networks where periodic data sequences are difficult to obtain	2020
[82]	Considers the network as a graph and uses the bidirectional LSTM and the ConVNNs to forecast traffic	Effective in highly complex, nonlinear and dynamic traffic conditions	Computationally complex steps involved in the forecast output	2021
[83]	Designs an algorithm that uses principal component analysis for reducing data dimension, graph convolutional network for learning network's topology and LSTM for obtaining features in time	Improved accuracy due to combination of various sub-algorithms for varied tasks in prediction	Inefficient for real-time traffic analysis due to involvement of multiple algorithms	2021
[84]	Applies interpolation with spatial densities as input to the GCN-LSTM to predict traffic congestion and travel time	Effective for congested traffic networks	Requires use of high storage for data processing of congested traffic networks	2022
[85]	Obtains spatial traffic data by GCN and temporal dependencies by Bi-LSTM with attention mechanism	Effectively correlates spatial and temporal dependencies	Features extraction is cumbersome in complex traffic networks	2020

dependencies that are present among the nodes in the graph in space and time are extracted. The attention mechanism is used for modeling the non-Euclidean data while the LSTM models the data in time sequences. A recurrent attention unit (RAU) concept integrating LSTM with an attention mechanism is given in [89] that learns about the inner data features to be used for traffic forecast. The authors in [90] considered the traffic of a region as a directed graph that is un-weighted and

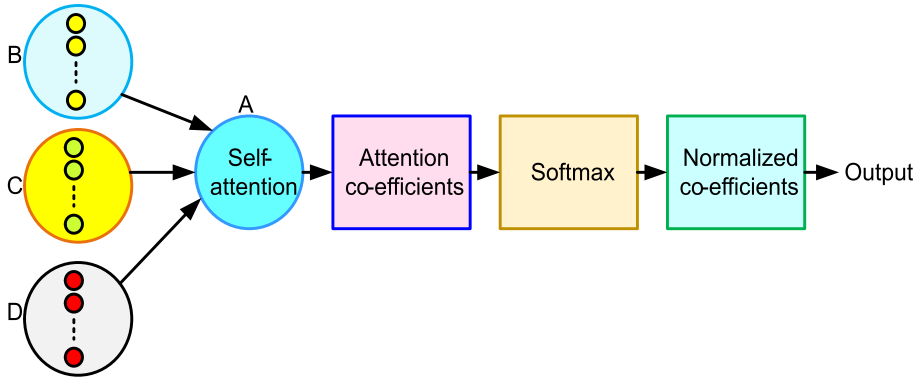

FIGURE 10. In graph-attention, a node (A) applies self-attentions to its neighbors'(B, C and D) features to compute attention co-efficients normalized by the softmax fuction.

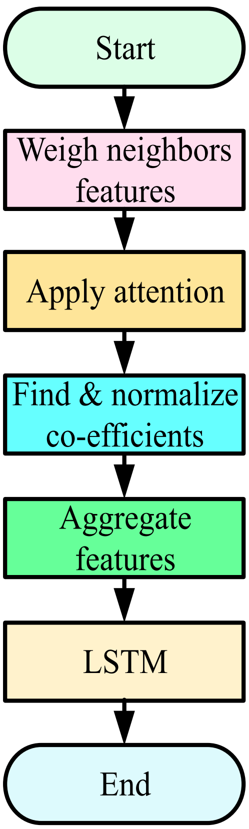

FIGURE 11. Flow chart of GA-LSTM. Nodes obtain features of neighbors by graph attention mechanism that involves obtaining neighbors features, applying self-attention to determine attention co-efficients and normalizing them to be fed to the LSTM for forecast.

acquires the data features in time and space among the nodes of the graph. It models the non-Euclidean data structures with graph-attention mechanism and the time-series modeling by the LSTM with an effective prediction accuracy. An attention mechanism based LSTM is proposed in [91] in which the past values from the long sequences of the data traffic are used to learn more features for the current values. This allows better prediction accuracy than the simple LSTM. Table 3 summarizes the key aspects of the GA-LSTM algorithms (see [61], [89], [90], [91]).

G. TRAFFIC PREDICTION USING MISCELLANEOUS VARIANTS OF LSTM

This section describes the short-term road traffic forecasting algorithms involving the LSTM and one or more other algorithms that are not distinguishable enough to be placed in the classes described above. An algorithm utilizing ensemble learning of the LSTM with no negativity constraint theory and using the integration of weights and the optimization of extremal population methods is proposed in [92]. It first produces a cluster of LSTM network and its forecasting is made with respect to different time tags and then the weight co-efficients of the cluster elements for prediction are specified using no negativity constraint and the population extremal. The authors in [93] proposed the stack auto-encoder LSTM (SAE-LSTM). The SAE part acquires the the features in space and the LSTM acquires the time features, which are then combined to forecast the traffic. The vehicle trajectory is predicted in [94] by using an LSTM with an encoder-decoder model that captures the features in time while the features in space among vehicles are acquired by a graph learning representation mechanism. A mechanism that aggregates the attention is also used to know about the attention paid by the drivers to other vehicles in the neighborhood. An edge-native LSTM with encode-decoder approach is used for traffic data prediction in [95] that takes multivariate data input from the micro-boxes of a selected network.

The authors in [96] combined performance evaluation of LSTM, LSTM encoder-decoder, ConVNNs-LSTM and ConV-LSTM. It is found that the ConV-LSTM is the best in acquiring the data features in time and space. A linear LSTM scheme is proposed in [97] that forecasts the loss function by the mean square error (MSE) to measure the prediction accuracy and then uses the Adam algorithm to measure, comprehend and learn the extracted features and optimize them for prediction (MSE-Adam). A dilated LSTM based algorithm is proposed in [98] for dealing with the problem of exploding and vanishing of gradient in the RNNs. The data are obtained from vehicles over a wide network and preprocessed and organized in time-series of five minute resolution. Then they are subjected to the LSTM networks; that consist of four hidden layers with multi-steps dilation

TABLE 3. Traffic flow prediction using GA-LSTM [61], [89], [90], [91] and miscellaneous variants of LSTM [92], [93], [94], [95], [96], [97], [98], [99], [100], [101], [102], [103], [104], [105], [106], [107], [108], [109], [110], [111], [112], [113].

Reference	Main Algorithm	Achievement	Limitations	Year
[88]	Models the non-Euclidean data with an attention mechanism and time-series with LSTM	Effective for networks with data variations	Complex data processing	2020
[89]	Combines LSTM with an attention mechanism for acquiring data features	Simple architecture	Challenging accuracy in diverse traffic	2021
[90]	In un-weighted directed graph, the non-Euclidean and time-series data are obtained by the respective attention mechanism and LSTM	Effective for traffic with stochastic features	Has high computational complexity	2020
[91]	An attention mechanism obtains current data information for LSTM using past values	Improves the accuracy of LSTM	Is more complex than LSTM	2020
[92]	Ensembles the LSTM cluster data with no negativity constraint with weights integration and extremal population optimization	High prediction accuracy	Making LSTM clusters is challenging	2019
[93]	Combines stacked auto-encoder with LSTM	Improved accuracy	High computational complexity	2022
[94]	Combines an encoder-decoder based LSTM with graph representation learning mechanism and attention aggregator to acquire the features in time and space for vehicles trajectory forecast	Effective for autonomous vehicles driving	Network congestion and diversity affect the output of the prediction	2021
[95]	Uses the encoder-decoder LSTM for edge devices to predict network data traffic	Effective for congested networks	Data burden on the edge devices	2021
[96]	Compares LSTM and its variants for traffic forecast	Provides a base for the use of LSTM and its variants for traffic flow	Comparison lacks the diverse traffic conditions for concluding the best algorithm	2021
[97]	Implements a linear LSTM algorithm based on mean square error and Adam optimization	Less complex due to linearity	Requires validation for non-linear features	2020
[98]	Uses a dilated LSTM model consisting of four layers with multi-step dilation	Makes the RNNs stable by overcoming their explosion and gradient vanishing issues	Requires additional resources	2020
[99]	Further improves the time-series analysis of the data of LSTM by adding dynamic time wrapping so they match more with the historical trends	Suitable in applications requiring high accuracy of data traffic prediction	Vehicles connectivity is the major issue, especially in rural areas	2020
[100]	Considers traffic of the adjacent paths in addition to a single chosen path for effective traffic prediction	Effective for congested areas with multiple traffic paths	Traffic from adjacent paths increases the interference and noise in the traffic data	2021
[101]	Integrates the multi-layer perception with LSTM to forecast the time that will impact the traffic data	Provides a real-time traffic prediction and avoids the delay in data prediction and its actual happening	The complexity of the network increases for diverse traffic conditions	2021
[102]	Extends the memory of LSTM with data divided into long and short-term traffic flow patterns	Efficient method for modeling and predicting large-scale diverse traffic conditions	Requires additional memory and other resources	2020
[103]	The imputation of missing traffic data is performed by the method of tensor completion	Effective to forecast traffic with randomly missing data values	Intensive computations due to the use of various sub-algorithms	2020
[104]	Uses a boosting model to obtain spatio-temporal characteristics by a bagging-based random forest method and extracted by the LSTM	Effective for self-driving vehicles with lane-level map navigation	Requires further validation for dynamic and diverse lane traffic conditions	2022
[105]	Optimizes the window size for data features using the grid search method	Improves the prediction accuracy	Window size requires constant updates in varying traffic in diverse traffic conditions	2021
[106]	The traffic data are stored in the LSTM and the time series are obtained then a Kalman filter dynamically adjusts the weights to predict traffic	Adaptation in Kalman filter reduces the prediction error	Has more computational complexity than LSTM	2021
[107]	Optimizes the learning rate of algorithm, batch size and neurons count using optimization by the particle swarm method for LSTM	Improved performance over the traditional algorithms	Needs further evaluation of performance in complex traffic networks	2011
[108]	Transforms traffic features into a new feature and then obtains a matrix followed by the LSTM model to track traffic states changes and identify the temporal dependencies	Effective in predicting the traffic patterns	Prediction involves multiple steps than simple prediction algorithms	2021
[109]	Models the LSTM, GRU and ConVNNs-LSTM to address the spatiotemporal dependencies of complex road traffic conditions	Justifies that complex models are not always better than simple models for traffic prediction	The simple prediction algorithms do not work always in varied traffic conditions	2021
[110]	Uses a stacked LSTM algorithm with real time data from Google Map	Real time time prediction	Complexity due to multiple sub-algorithms	2021
[111]	Preprocessed data features are input to LSTM optimized hidden layer	High prediction accuracy	Requires operation of other algorithms in addition to LSTM	2020
[112]	Periodic features are chosen at input	Improved prediction accuracy	Diverse features selection is cumbersome	2021
[113]	Extracts traffic flow features in space and time and combines them with individual traffic flow of a lane and aggregate traffic flow	Improved prediction accuracy	Requires constant updates of individual lane and aggregate traffic conditions	2021

to predict the traffic. The authors in [99] fine tune the time-series analysis of the already existing algorithms using global position system (GPS) data by introducing dynamic time wrapping (DTW) algorithm to the LSTM to design the D-LSTM system. It introduces the attenuation effect in the data and further enhances the accuracy of the already existing data processed by deep learning algorithms so that they match more with the historical trends. It also excludes the effect of special treatment of data on holidays. The authors in [100] evaluated the spatio-temporal dependencies not only for a single traffic path but for the adjacent paths as well. This results in effective traffic flow prediction than the case when traffic over a single trajectory is considered.

The authors in [101] argued that the most of the traffic prediction algorithms do not consider the time that is left until the prediction results become prevailing in the real-time data traffic. To overcome this problem, a method is designed that considers traffic incident data and traffic flow data to forecast the remaining time for which the traffic is impacted. The framework in [102] addressed that the LSTM has limited memory and it does not significantly remember the traffic data from the past. Therefore, a memory time-series network having additional memory is designed. The data are classified into long and short-term data, the former represents the overall traffic patterns while the latter models the recent traffic trends. The imputation of missing data is addressed in [103] to improve data traffic accuracy by imputing the missing data using the tensor completion method. Various tensor methods are used to impute the missing data. Four methods are used to predict the traffic: support vector regression, the nearest neighbor technique, gradient boost regression tree and LSTM. The data missing may occur in time, space or element-wise in a random manner. The results showed that improving the accuracy of missing data imputation increases the forecast accuracy. The authors in [104] argued that the traditional forecast algorithms do not fit well in traffic forecasting for automatic driving vehicles when they use the lane map for navigation. The bias in the model is reduced by the selection of the features while its variance is reduced by the random forest algorithm based on bagging. The temporal trends in traffic data for a lane is obtained by the LSTM that are also optimized [105]. The method of grid search is used to select the window size for the spatial and temporal characteristics so as to achieve accuracy. The time steps for the features at the input define size of the temporal window while the size of geographical neighborhood is termed as the optimal size of the window. The window size is optimized for the features in time and space and forecast is made by the LSTM. A Kalman-LSTM model is presented in [106]. First the LSTM model stores the pre-order data followed by Kalman filter that dynamically adjusts the data to forecast the traffic. The LSTM is combined with the particle swarm optimization in [107] that considers the speed of vehicles and the traffic flow to and from a checkpoint. The rate with which

TABLE 4. Comparative analysis of the classified categories of algorithms.

Algorithms' Class	Functional Strategy	Merits	Limitations	Data Nature
LSTM	Designed to solve the vanishing gradient problem in RNNs. It has a forget gate to allow only meaningful information and processes them with an input gate that scales (according to the information content) and normalizes the data to determine long-term output and forwards them to be further processed by an output gate to obtain the short-term output	Minimizes the loss of information weight of data streams (vanishing gradient problem) by having short and long-term memories for retaining the desired information, effective for sequential (time-series) data, has short and long-term internal memories helpful for retaining desired information	For long data sequences, the vanishing (or exploding) gradient problem becomes significant and the output decisions of the gating mechanism becomes compromised. Also, the recurrent nature of the data limits the parallel data processing and affects the speed	Temporal
RNNs-LSTM	This class uses the basic structural architecture of the RNNs in which the data streams at a certain time step are multiplied with their weights, added with weighted memory states and processed collectively to perform activation functions and produce the output	Possesses internal short-term memory and takes into account the time order of the input data sequences making it ideal for information with temporal dependencies	The weight of information reduces with the increasing length of time sequence causing the well known vanishing gradient problem (or gradient explosion as well)	Temporal
G-RUNNs-LSTM	Designed to solve the vanishing gradient problem of the RNNs. Makes the architectural change in the LSTM by structuring its forget and input gates into a single update gate	Reduces the complex circuitry for short-term memory requirement and, therefore, reduces the computational and processing complexity	Does not possess the long-term memory useful for information retention, especially in long data sequences	Temporal
ConVNNs-LSTM	It uses the convolution operation to extract the spatial features that are reduced in size and classified. The LSTM is used for temporal features and performs the overall prediction	Effective for a diverse traffic network with dynamic spatio-temporal characteristics	High computational complexity due to a sequence of processing layers that further increases with increasing datasets, especially required for training. The convolution operation specifically requires intensive computation	Spatio-temporal
G-ConVNNs-LSTM	This class of algorithms considers the traffic network as a graph that obtains the normalized traffic features by performing convolution in spatial or frequency domain while the LSTM obtains the temporal features, which are then processed collectively to perform the prediction	Convenience of modeling for networks inherently existing in graphical form such as traffic networks with irregular data structures unlike the structured data (images, text, etc.)	The graph structure dynamically changes that makes the training of a deep learning algorithm (like LSTM) a challenging task. Also, establishing graph structure in complex networks is cumbersome because the nature of the data varies among the nodes and each node has different characteristics that are cumbersome to handle. Besides, this class of algorithms performs aggregation on data that considers that all the features are equally important that increases the redundancy in data features and reduces the information content that, in turn, challenges the features extraction during the training phase of the algorithms. Moreover, the spectral domain convolution requires the network to be homogeneous that is less likely in most of the real world cases	Spatio-temporal
GA-LSTM	It considers the road network as a graph and the nodes receive the weighted features of the neighbors according to their importance to produce corresponding transformed features that are further processed by the LSTM to predict the output	Reduces the complexity of processing by focusing on the important features that, in turn, enhances the accuracy as well	Slow convergence of the network during training as the attention has to be learned besides the parameters of the network	Spatio-temporal

the algorithm learns, the batch size and neurons count are optimized for the LSTM by the particle swarm algorithm. The model predicts the traffic in a more accurate manner than some traditional algorithms. The authors in [108] argued that the existing traffic prediction schemes suffer from the issues of compromised accuracy and adaptability, difficulty in real-time applications and the challenging nature of extracting spatial and temporal features. Therefore, a neural network based on Chebeshev graph is designed. The traffic data features are first transformed to new features in the form of a matrix using a fully connected layer. It is followed by the LSTM model that learns about traffic state changes for the sake of capturing temporal dependencies. To address the complex road traffic conditions and spatio-temporal dependencies of data, the LSTM, GRU, and ConVNNs-LSTM models are compared in [109] with the conclusion that the GRU is the best in handling the complex traffic conditions and that a complex model such as ConVNNs-LSTM is not always better than the less complex models such as LSTM and GRU. The authors in [110] proposed a stacked LSTM algorithm. They considered real time and historical data from Google Map using three different urban road conditions. The predicted speed of the algorithm is then mapped into the predicted flow of traffic using a correlation algorithm. A dynamic optimization based LSTM is proposed in [111] that performed traffic data pre-processing and classified them into normal and outlier data. The training of the LSTM is performed only by the normal data to further perform traffic prediction. A chaotic particle swarm optimization (CPSO) algorithm dynamically optimizes the hidden layer of the LSTM so as to achieve robust performance. The feature selection is combined with the LSTM model in [112] while features extraction in space and time is performed in [113] for forecasting with a bi-directional LSTM. Table 3 summarizes the key aspects of the miscellaneous variants of the LSTM algorithm (see [92], [93], [94], [95], [96], [97], [98], [99], [100], [101], [102], [103], [104], [105], [106], [107], [108], [109], [110], [111], [112], [113]). These algorithms opt various diverse and varied strategies and mechanisms to process the data before input to the LSTM. In addition, they also combine one or more strategies with the LSTM. Based on the description of the classified algorithms, Table 4 provides a comparative analysis. Every algorithm class has its own operational strategy, merits and demerits that make it applicable to various traffic conditions and other applications as well. The limitations of these algorithms are helpful in further improving them. Applying a specific class of algorithms to the traffic conditions depends upon the types of the traffic networks. In addition, the nature of the data these classes of algorithms generally process is described for their use in various fields. The complexity of algorithms processing spatio-temporal data is generally higher than algorithms using temporal data. Figure 12 shows the number of publications made per year from 2013 to 2023 for traffic prediction. These publications include the traffic prediction in various applications, such as railway, aerial, pedestrian, vehicular and computer networks, to mention a few. Some of the values; especially the smaller values, have been normalized for the sake of better graphical

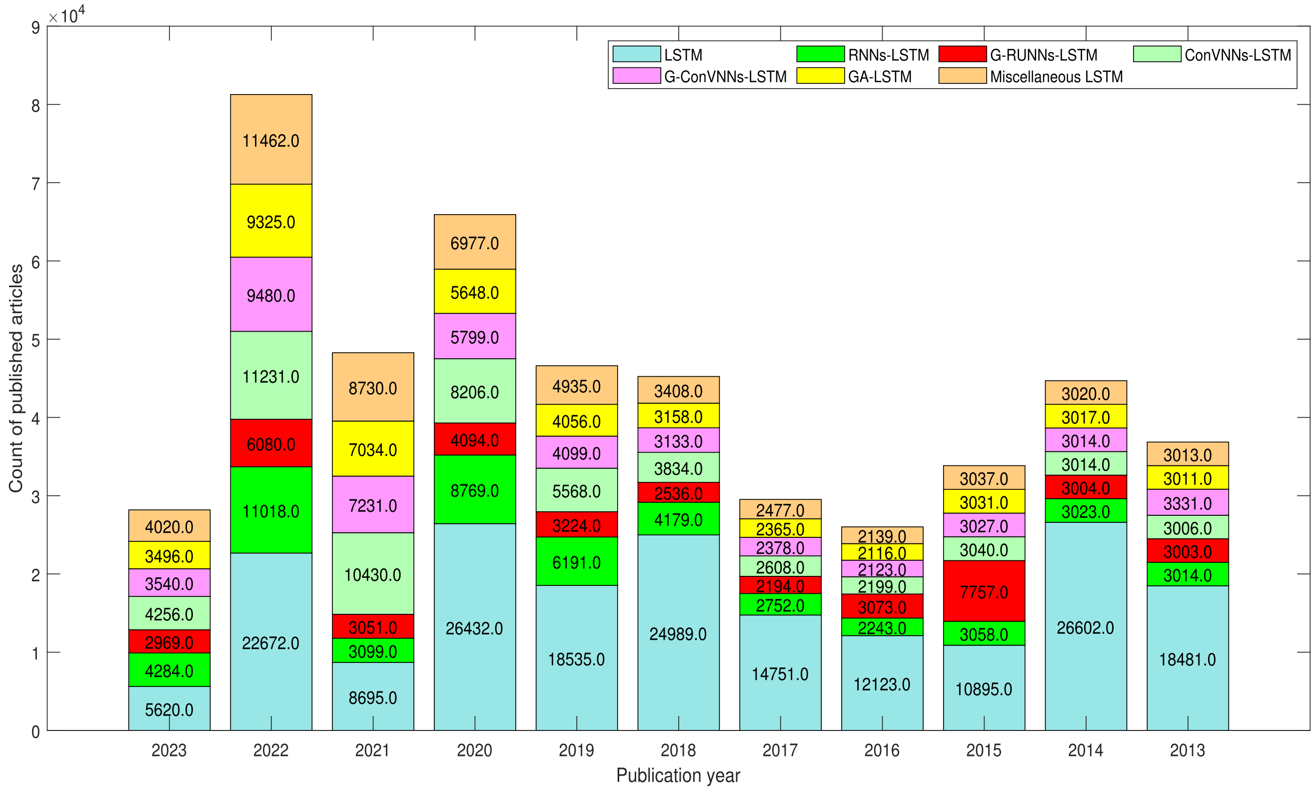

FIGURE 12. The publication count per year involving LSTM for traffic (aerial, vehicular, pedestrian, vehicular and computer network traffic, to mention a few) prediction from 2013 to 2023. As the graph depicts, the LSTM has been one of the widely used algorithms for traffic prediction in the mentioned decade. Data taken from: https://app.dimensions.ai.

visibility and presentation. It also shows the the LSTM is one of the utmost important algorithms of deep learning that has been extensively used for traffic forecast in various applications.

VII. CHALLENGES IN SHORT-TERM TRAFFIC PREDICTION USING LSTM

There are some challenges inherently associated with the LSTM and the short-term road traffic prediction algorithms, which are described in the following lines.

- The LSTM model requires a significant amount of input data for training the network before performing prediction. This not only challenges the availability of data but also enhances the computational complexity of the algorithm. Moreover, when the datasets are large, the training process becomes cumbersome and time-consuming.
- The past information retention capability of the LSTM model degrades for large datasets, which requires more sophisticated strategies to be addressed. This also causes the vanishing gradient problem to become significant with heavy data size. This is because the gradient values are multiplied and the smaller values of the gradient with heavy data load significantly reduce the overall gradient.
- The accuracy of prediction of the LSTM hampers with highly nonlinear traffic variables and noise. Therefore, information extraction from the data in such scenarios is challenging.
- The application of the LSTM is limited to datasets that exist in the form of sequences (time-series). It poorly performs in dealing with non-sequential datsets, such as tasks prediction and classification in online learning.
- In applications involving the use of multiple features of data for prediction, optimal information extraction from the stochastic and random data is challenging.
- The designed traffic prediction strategies are generally valid for a small network with specific conditions and parameters (such as dependency on geographic locations, weather conditions, to mention a few). Designing prediction strategies applicable to a diverse set of traffic networks takes into account multiple design and prediction strategies.

VIII. CONCLUSION AND FUTURE WORK

This paper addressed a survey of the latest and state-of-the-art short-term road traffic algorithms using the LSTM deep learning model. The algorithms developed in the last three years are studied and analyzed to provide an in-depth and thorough description rather than a marginal description of deep learning algorithms. The algorithms dealing with the short-term traffic prediction in intelligent transportation are useful in forecasting travel time, traffic volume, clean, safe and secure roads. They are also helpful to predict the optimal travelling trajectories for the vehicles, especially for the driverless vehicles during travelling. They are useful to estimate accident and fatality rate of the road conditions and allow the traffic management authorities to adopt robust, efficient, secure and travel-friendly policies for traffic management. They are also effective to forecast the adverse weather conditions (snow, rain, wind, flood, among the many) and provide an alarm before they actually take place to cause road emergencies. The challenges associated with the LSTM-based short-term traffic forecast are addressed and strategies are mentioned for their solutions in future investigation.

As future work, the following research strategies are promising to address the challenges associated with the LSTM-based short-term traffic prediction [114].

- The use of an attention mechanism will ease the features extraction process by focusing on the important features of the input data. This will reduce the computational complexity to a significant extent.
- The application of noise removal techniques to the input data such as filtering the image data and leveraging the data numbers to practically existing values would improve the accuracy of prediction.
- The use of transformer architecture allows parallel processing of time-series data that can be applied to reduce the training time and expedite the output prediction. Moreover, they can also be used to capture dependencies among the elements of sequences, even for long sequences. This can be helpful in dealing with the vanishing gradient problem.
- The investigation in the use of transfer learning techniques will be helpful in learning the parameters from the input data and then applying it again on the similar network or the same network in different time slots. This will reduce the computational time. This is particularly helpful in diverse traffic networks with complex data.
- The use of the concept of cloud computing could reduce the computational delay and ensure efficient utilization of the resources, especially in training the LSTM and other deep learning algorithms, as the cloud resources are used for these purposes.

REFERENCES

[1] G.-L. Huang, A. Zaslavsky, S. W. Loke, A. Abkenar, A. Medvedev, and A. Hassani, "Context-aware machine learning for intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 1, pp. 17–36, Jan. 2023.

[2] A. Abdelraouf, M. Abdel-Aty, and N. Mahmoud, "Sequence-to-sequence recurrent graph convolutional networks for traffic estimation and prediction using connected probe vehicle data," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 1, pp. 1395–1405, Jan. 2023.

[3] C. Kim, Y. Yoon, S. Kim, M. J. Yoo, and K. Yi, "Trajectory planning and control of autonomous vehicles for static vehicle avoidance in dynamic traffic environments," *IEEE Access*, vol. 11, pp. 5772–5788, 2023.

[4] P. Wang, H. Yu, C. Liu, Y. Wang, and R. Re, "Real-time trajectory prediction method for intelligent connected vehicles in urban intersection scenarios," *Sensors*, vol. 23, no. 6, pp. 1–20, Mar. 2023.

[5] W. Zhuang and Y. Cao, "Short-term traffic flow prediction based on a K-nearest neighbor and bidirectional long short-term memory model," *Sensors*, vol. 13, no. 4, pp. 1–15, Feb. 2023.

[6] R. Shi and L. Du, "Multi-section traffic flow prediction based on MLR-LSTM neural network," *Sensors*, vol. 22, no. 19, pp. 1–20, Sep. 2022.

[7] S. Zhou, C. Wei, C. Song, W. Chang, and L. Yang, "A hybrid deep learning model for short-term traffic flow prediction considering spatiotemporal features," *Sustainability*, vol. 14, no. 16, pp. 1–14. Aug. 2022.

[8] S. Mohammed, S. H. Krishna, P. K. Mudalkar, N. Verma, P. Karthikeyan, and A. S. Yadav, "Stock market price prediction using machine learning," in *Proc. 5th Int. Conf. Smart Syst. Inventive Technol. (ICSSIT)*, Tirunelveli, India, Jan. 2023, pp. 823–828.

[9] E. A. Devi, S. Gopi, U. Padmavathi, S. R. Arumugam, S. P. Premnath, and D. Muralitharan, "Plant disease classification using CNN-LSTM techniques," in *Proc. 5th Int. Conf. Smart Syst. Inventive Technol. (ICSSIT)*, Tirunelveli, India, Jan. 2023, pp. 1225–1229.

[10] S. T. Himi, N. T. Monalisa, M. Whaiduzzaman, A. Barros, and M. S. Uddin, "MedAi: A smartwatch-based application framework for the prediction of common diseases using machine learning," *IEEE Access*, vol. 11, pp. 12342–12359, 2023.

[11] J. Hu, L. Weng, Z. Gao, and B. Yang, "State of health estimation and remaining useful life prediction of electric vehicles based on real-world driving and charging data," *IEEE Trans. Veh. Technol.*, vol. 72, no. 1, pp. 382–394, Jan. 2023.

[12] M. Singh and R. Dubey, "Deep learning model based CO_2 emissions prediction using vehicle telematics sensors data," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 1, pp. 768–777, Jan. 2023.

[13] X. Chen, W. Chen, V. Dinavahi, Y. Liu, and J. Feng, "Short-term load forecasting and associated weather variables prediction using ResNet-LSTM based deep learning," *IEEE Access*, vol. 11, pp. 5393–5405, 2023.

[14] Y. Wei, M.-M. Zhao, A. Liu, and M.-J. Zhao, "Channel tracking and prediction for IRS-aided wireless communications," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 563–579, Jan. 2023.

[15] Y. Hou, X. Zheng, C. Han, W. Wei, R. Scherer, and D. Połap, "Deep learning methods in short-term traffic prediction: A survey," *Inf. Technol. Control*, vol. 51, no. 1, pp. 139–157, Mar. 2022.

[16] A. Chen, J. Law, and M. Aibin, "A survey on traffic prediction techniques using artificial intelligence for communication networks," *Telecom*, vol. 2, no. 4, pp. 518–535, Dec. 2021.

[17] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 3–19, Jun. 2014.

[18] J. Zhu, Z. Yang, M. Mourshed, Y. Guo, Y. Zhou, Y. Chang, Y. Wei, and S. Feng, "Electric vehicles charging load forecasting: A comparative study of deep learning approaches," *Energies*, vol. 12, pp. 1–19, Jul. 2019.

[19] A. A. E. Donkol, A. G. Hafez, A. I. Hussein, and M. M. Mabrook, "Optimization of intrusion detection using likely point PSO and enhanced LSTM-RNN hybrid technique in communication networks," *IEEE Access*, vol. 11, pp. 9469–9482, 2023.

[20] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2017.

[21] L. Feng, "Predicting output responses of nonlinear dynamical systems with parametrized inputs using LSTM," *IEEE J. Multiscale Multiphys. Comput. Techn.*, vol. 8, pp. 97–107, 2023.

[22] E. Ahmadzadeh, H. Kim, O. Jeong, N. Kim, and I. Moon, "A deep bidirectional LSTM-GRU network model for automated ciphertext classification," *IEEE Access*, vol. 10, pp. 3228–3237, 2022.

[23] T. Padmapritha, "Prediction of blood glucose level by using an LSTM based recurrent neural networks," in *Proc. IEEE Int. Conf. Clean Energy Energy Efficient Electron. Circuit Sustain. Develop. (INCCES)*, Krishnankoli, India, Dec. 2019, pp. 1–4.

[24] P. Poonia and V. K. Jain, "Short-term traffic flow prediction: Using LSTM," in *Proc. Int. Conf. Emerg. Trends Commun., Control Comput. (ICONC)*, Lakshmangarh, India, Feb. 2020, pp. 1–4.

[25] B. Praveen Kumar and K. Hariharan, "Multivariate time series traffic forecast with long short term memory based deep learning model," in *Proc. Int. Conf. Power, Instrum., Control Comput. (PICC)*, Thrissur, India, Dec. 2020, pp. 1–5.

[26] Y. N. Malek, M. Najib, M. Bakhouya, and M. Essaaidi, "Multivariate deep learning approach for electric vehicle speed forecasting," *Big Data Mining Analytics*, vol. 4, no. 1, pp. 56–64, Mar. 2021.

[27] J. Zheng and M. Huang, "Traffic flow forecast through time series analysis based on deep learning," *IEEE Access*, vol. 8, pp. 82562–82570, 2020.

[28] H. Bouchemoukha, M. Nadjib, and Z. A. Lahoulou, "Is classical LSTM more efficient than modern GCN approaches in the context of traffic forecasting?" in *Proc. Int. Conf. Recent Adv. Math. Informat. (ICRAMI)*, Tebessa, Algeria, Sep. 2021, pp. 1–6.

[29] Q. Zhaowei, L. Haitao, L. Zhihui, and Z. Tao, "Short-term traffic flow forecasting method with M-B-LSTM hybrid network," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 1, pp. 225–235, Jan. 2022.

[30] Q. Chu, G. Li, R. Zhou, and Z. Ping, "Traffic flow prediction model based on LSTM with Finnish dataset," in *Proc. 6th Int. Conf. Intell. Comput. Signal Process. (ICSP)*, Xi'an, China, Apr. 2021, pp. 389–392.

[31] H. Ruan, B. Wu, B. Li, Z. Chen, and W. Yun, "Expressway exit station short-term traffic flow prediction with split traffic flows according originating entry stations," *IEEE Access*, vol. 9, pp. 86285–86299, 2021.

[32] L. Yao, J. Bao, F. Ding, N. Zhang, and E. Tong, "Research on traffic flow forecast based on cellular signaling data," in *Proc. IEEE Int. Conf. Smart Internet Things (SmartIoT)*, Jeju, South Korea, Aug. 2021, pp. 193–199.

[33] F. M. Awan, R. Minerva, and N. Crespi, "Using noise pollution data for traffic prediction in smart cities: Experiments based on LSTM recurrent neural networks," *IEEE Sensors J.*, vol. 21, no. 18, pp. 20722–20729, Sep. 2021.

[34] S. Anveshrithaa and K. Lavanya, "Real-time vehicle traffic analysis using long short term memory networks in apache spark," in *Proc. Int. Conf. Emerg. Trends Inf. Technol. Eng. (ic-ETITE)*, Vellore, India, Feb. 2020, pp. 1–5.

[35] A. Nigam and S. Srivastava, "Macroscopic traffic stream variable prediction with weather impact using recurrent learning approach," in *Proc. IEEE-HYDCON*, Hyderabad, India, Sep. 2020, pp. 1–6.

[36] S. M. Snineh, N. E. A. Amrani, M. Youssfi, O. Bouattane, and A. Daaif, "Detection of traffic anomaly in highways by using recurrent neural network," in *Proc. 5th Int. Conf. Intell. Comput. Data Sci. (ICDS)*, Fez, Morocco, Oct. 2021, pp. 1–6.

[37] T. Zhang, W. Song, M. Fu, Y. Yang, and M. Wang, "Vehicle motion prediction at intersections based on the turning intention and prior trajectories model," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 10, pp. 1657–1666, Oct. 2021.

[38] L. Hou, L. Xin, S. E. Li, B. Cheng, and W. Wang, "Interactive trajectory prediction of surrounding road users for autonomous driving using structural-LSTM network," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 11, pp. 4615–4625, Nov. 2020.

[39] V. Tran and A. Panangadan, "Few-shot time-series forecasting with application for vehicular traffic flow," in *Proc. IEEE 23rd Int. Conf. Inf. Reuse Integr. Data Sci. (IRI)*, San Diego, CA, USA, Aug. 2022, pp. 20–26.

[40] Y. Gao, C. Zhou, J. Rong, Y. Wang, and S. Liu, "Short-term traffic speed forecasting using a deep learning method based on multitemporal traffic flow volume," *IEEE Access*, vol. 10, pp. 82384–82395, 2022.

[41] B. R. Krishna, M. H. Reddy, P. S. Vaishnavi, and S. V. Reddy, "Traffic flow forecast using time series analysis based on machine learning," in *Proc. 6th Int. Conf. Comput. Methodol. Commun. (ICCMC)*, Erode, India, Mar. 2022, pp. 943–947.

[42] F. Miao, L. Tao, J. Xue, and X. Zhang, "A queue hybrid neural network with weather weighted factor for traffic flow prediction," in *Proc. IEEE 24th Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD)*, Dalian, China, May 2021, pp. 788–793.

[43] Q. Wu, Z. Jiang, K. Hong, H. Liu, L. T. Yang, and J. Ding, "Tensor-based recurrent neural network and multi-modal prediction with its applications in traffic network management," *IEEE Trans. Netw. Service Manage.*, vol. 18, no. 1, pp. 780–792, Mar. 2021.

[44] J. Deng and G. Shen, "Federated learning-based privacy-preserving traffic flow prediction scheme for VANETs," in *Proc. 4th Int. Conf. Commun., Inf. Syst. Comput. Eng. (CISCE)*, Shenzen, China, May 2022, pp. 374–378.

[45] Z. Li, C. Li, X. Cui, and Z. Zhang, "Short-term traffic flow prediction based on recurrent neural network," in *Proc. Int. Conf. Comput. Commun. Artif. Intell. (CCAI)*, Guangzhou, China, May 2021, pp. 81–85.

[46] D. Wu and M. Chi, "Long short-term memory with quadratic connections in recursive neural networks for representing compositional semantics," *IEEE Access*, vol. 5, pp. 16077–16083, 2017.

[47] A. Lawi, H. Mesra, and S. Amir, "Implementation of long short-term memory and gated recurrent units on grouped time-series data to predict stock prices accurately," *J. Big Data*, vol. 9, no. 1, p. 89, Jul. 2022.

[48] G. Shen, Q. Tan, H. Zhang, P. Zeng, and J. Xu, "Deep learning with gated recurrent unit networks for financial sequence predictions," *Proc. Comput. Sci.*, vol. 131, pp. 895–903, May 2018.

[49] W. Shu, K. Cai, and N. N. Xiong, "A short-term traffic flow prediction model based on an improved gate recurrent unit neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 16654–16665, Sep. 2022.

[50] S. Wang, J. Zhao, C. Shao, C. Dong, and C. Yin, "Truck traffic flow prediction based on LSTM and GRU methods with sampled GPS data," *IEEE Access*, vol. 8, pp. 208158–208169, 2020.

[51] H. Hu, Z. Lin, Q. Hu, and Y. Zhang, "Attention mechanism with spatial–temporal joint model for traffic flow speed prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 16612–16621, Sep. 2022.

[52] M. F. Adnan, N. Ahmed, I. Ishraque, M. S. A. Amin, and M. S. Hasan, "Traffic congestion prediction using deep convolutional neural networks: A color-coding approach," in *Proc. Int. Conf. Eng. Emerg. Technol. (ICEET)*, Kuala Lumpur, Malaysia, Oct. 2022, pp. 1–5.

[53] X. Ma, Z. Dai, Z. He, and Y. Wang, "Learning traffic as images: A deep convolution neural network for large-scale transportation network speed prediction," *Senors*, vol. 17, no. 4, pp. 1–16, Apr. 2017.

[54] J. Teuwen and N. Moriakov, "Convolutional neural networks," in *Handbook of Medical Image Computing and Computer Assisted Intervention* (The Elsevier and MICCAI Society), S. K. Zhou, D. Rueckert, and G. Fichtingerm, Eds. Cambridge, MA, USA: Academic, 2020, pp. 481–501.

[55] H. Zheng, F. Lin, X. Feng, and Y. Chen, "A hybrid deep learning model with attention-based conv-LSTM networks for short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 11, pp. 6910–6920, Nov. 2021.

[56] Y. Ma, Z. Zhang, and A. Ihler, "Multi-lane short-term traffic forecasting with convolutional LSTM network," *IEEE Access*, vol. 8, pp. 34629–34643, 2020.

[57] D. Ma, X. Song, and P. Li, "Daily traffic flow forecasting through a contextual convolutional recurrent neural network modeling inter- and intra-day traffic patterns," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 2627–2636, May 2021.

[58] N. Bansal, R. S. Bali, K. Jakhar, M. S. Obaidat, N. Kumar, S. Tanwark, and J. J. P. C. Rodrigues, "HTFM: Hybrid traffic-flow forecasting model for intelligent vehicular ad hoc networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Montreal, QC, Canada, Jun. 2021, pp. 1–6.

[59] W. Lu, Y. Rui, and B. Ran, "Lane-level traffic speed forecasting: A novel mixed deep learning model," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 4, pp. 3601–3612, Apr. 2022.

[60] J. Cao, X. Guan, N. Zhang, X. Wang, and H. Wu, "A hybrid deep learning-based traffic forecasting approach integrating adjacency filtering and frequency decomposition," *IEEE Access*, vol. 8, pp. 81735–81746, 2020.

[61] Y. Zhang, S. Wang, B. Chen, J. Cao, and Z. Huang, "TrafficGAN: Network-scale deep traffic prediction with generative adversarial nets," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 1, pp. 219–230, Jan. 2021.

[62] M. Cao, V. O. K. Li, and V. W. S. Chan, "A CNN-LSTM model for traffic speed prediction," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, Antwerp, Belgium, May 2020, pp. 1–5.

[63] A. Abdelraouf, M. Abdel-Aty, and J. Yuan, "Utilizing attention-based multi-encoder–decoder neural networks for freeway traffic speed prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 11960–11969, Aug. 2022.

[64] N. Ranjan, S. Bhandari, H. P. Zhao, H. Kim, and P. Khan, "City-wide traffic congestion prediction based on CNN, LSTM and transpose CNN," *IEEE Access*, vol. 8, pp. 81606–81620, 2020.

[65] J. Skjermo, P. Arnesen, C.-J. Södersten, and E. Dahl, "Predicting driving conditions at mountain crossings using deep learning," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Rhodes, Greece, Sep. 2020, pp. 1–5.

[66] H. Huang, J. Chen, X. Huo, Y. Qiao, and L. Ma, "Effect of multi-scale decomposition on performance of neural networks in short-term traffic flow prediction," *IEEE Access*, vol. 9, pp. 50994–51004, 2021.

[67] X. Chen, X. Xie, and D. Teng, "Short-term traffic flow prediction based on ConvLSTM model," in *Proc. IEEE 5th Inf. Technol. Mechatronics Eng. Conf. (ITOEC)*, Chongqink, China, Jun. 2020, pp. 846–850.

[68] S. Bilotta, E. Collini, P. Nesi, and G. Pantaleo, "Short-term prediction of city traffic flow via convolutional deep learning," *IEEE Access*, vol. 10, pp. 113086–113099, 2022.

[69] Y. Zhang and D. Xin, "A diverse ensemble deep learning method for short-term traffic flow prediction based on spatiotemporal correlations," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 16715–16727, Sep. 2022.

[70] Z. Cheng, J. Lu, H. Zhou, Y. Zhang, and L. Zhang, "Short-term traffic flow prediction: An integrated method of econometrics and hybrid deep learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 5231–5244, Jun. 2022.

[71] J. Li, Z. Zhang, F. Meng, and W. Zhu, "Short-term traffic flow prediction via improved mode decomposition and self-attention mechanism based deep learning approach," *IEEE Sensors J.*, vol. 22, no. 14, pp. 14356–14365, Jul. 2022.

[72] L. He and W. Luo, "3D-ConvLSTMNet: A deep spatio-temporal model for traffic flow prediction," in *Proc. 23rd IEEE Int. Conf. Mobile Data Manage. (MDM)*, Paphos, Cyprus, Jun. 2022, pp. 147–152.

[73] A. Agafonov, "Traffic prediction using graph convolutional neural networks," in *Proc. 10th Int. Conf. Inf. Sci. Technol.*, London, U.K., Sep. 2020, pp. 91–95.

[74] M. Niepert, M. Ahmed, and K. Kutzkov, "Learning convolutional neural networks for graphs," in *Proc. 33rd Int. Conf. Mach. Learn.*, New York, NY, USA, Jun. 2016, pp. 2014–2023.

[75] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Represent.*, Toulon, France, Apr. 2017, pp. 1–14.

[76] H. Zhu and P. Koniusz, "Simple spectral graph convolution," in *Proc. Int. Conf. Learn. Represent.*, Vienna, Austria, May 2021, pp. 1–15.

[77] M. Balcilar, G. Renton, P. Heroux, B. Gauzere, S. Adam, and P. Honeine. *Bridging the Gap Between Spectral and Spatial Domains in Graph Neural Networks*. Accessed: Jul. 22, 2023. [Online]. Available: https://hal.science/hal-02515637/

[78] Y. Shin and Y. Yoon, "Incorporating dynamicity of transportation network with multi-weight traffic graph convolutional network for traffic forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 2082–2092, Mar. 2022.

[79] D. Liu, S. Hui, L. Li, Z. Liu, and Z. Zhang, "A method for short-term traffic forecasting based on GNC-LSTM," in *Proc. Int. Conf. Comput. Vis., Image Deep Learn.*, Chongqing, China, Jul. 2020, pp. 364–368.

[80] Z. Cui, K. Henrickson, R. Ke, and Y. Wang, "Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 11, pp. 4883–4894, Nov. 2020.

[81] L. Liu, J. Zhen, G. Li, G. Zhan, Z. He, B. Du, and L. Lin, "Dynamic spatial–temporal representation learning for traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 11, pp. 7169–7183, Nov. 2021.

[82] S. Wu, "Spatiotemporal dynamic forecasting and analysis of regional traffic flow in urban road networks using deep learning convolutional neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 1607–1615, Feb. 2022.

[83] N. Li, S. Jia, and Q. Li, "Traffic message channel prediction based on graph convolutional network," *IEEE Access*, vol. 9, pp. 135423–135431, 2021.

[84] H. Katayama, S. Yasuda, and T. Fuse, "Traffic density based travel-time prediction with GCN-LSTM," in *Proc. IEEE 25th Int. Conf. Intell. Transp. Syst. (ITSC)*, Macau, China, Oct. 2022, pp. 2908–2913.

[85] H. Zhao, H. Yang, Y. Wang, D. Wang, and R. Su, "Attention based graph bi-LSTM networks for traffic forecasting," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Rhodes, Greece, Sep. 2020, pp. 1–6.

[86] A. Drif and H. Cherifi, "MIGAN: Mutual-interaction graph attention network for collaborative filtering," *Entropy*, vol. 24, no. 8, pp. 1–14, Aug. 2022.

[87] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," in *Proc. Int. Conf. Learn. Represent.*, Vancouver, BC, Canada, May 2018, pp. 1–12.

[88] T. Wu, F. Chen, and Y. Wan, "Graph attention LSTM network: A new model for traffic flow forecasting," in *Proc. 6th Int. Conf. Inf. Sci. Control Eng. (ICISCE)*, Zhengzhou, China, Jul. 2018, pp. 241–245.

[89] Z. Wei, Z. Li, C. Wang, Y. Chen, Q. Miao, Y. Lv, and F.-Y. Wang, "Recurrent attention unit: A simple and effective method for traffic prediction," in *Proc. IEEE Int. Intell. Transp. Syst. Conf. (ITSC)*, Indianapolis, IN, USA, Sep. 2021, pp. 1272–1277.

[90] T. Zhang and G. Guo, "Graph attention LSTM: A spatiotemporal approach for traffic flow forecasting," *IEEE Intell. Transp. Syst. Mag.*, vol. 14, no. 2, pp. 190–196, Mar. 2022.

[91] D. Chen, C. Xiong, and M. Zhong, "Improved LSTM based on attention mechanism for short-term traffic flow prediction," in *Proc. 10th Int. Conf. Inf. Sci. Technol. (ICIST)*, London, U.K., Sep. 2020, pp. 71–76.

[92] F. Zhao, G.-Q. Zeng, and K.-D. Lu, "EnLSTM-WPEO: Short-term traffic flow prediction by ensemble LSTM, NNCT weight integration, and population extremal optimization," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 101–113, Jan. 2020.

[93] Y. Tian, C. Wei, and D. Xu, "Traffic flow prediction based on stack AutoEncoder and long short-term memory network," in *Proc. IEEE 3rd Int. Conf. Autom., Electron. Electr. Eng. (AUTEEE)*, Shenyang, China, Nov. 2020, pp. 385–388.

[94] Z. Gao and Z. Sun, "Modeling spatio-temporal interactions for vehicle trajectory prediction based on graph representation learning," in *Proc. IEEE Int. Intell. Transp. Syst. Conf. (ITSC)*, Indianapolis, IN, USA, Sep. 2021, pp. 1334–1339.

[95] S. Zeb, M. A. Rathore, A. Mahmood, S. A. Hassan, J. Kim, and M. Gidlund, "Edge intelligence in softwarized 6G: Deep learning-enabled network traffic predictions," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Madrid, Spain, Dec. 2021, pp. 1–6.

[96] D. Haputhanthri and A. Wijayasiri, "Short-term traffic forecasting using LSTM-based deep learning models," in *Proc. Moratuwa Eng. Res. Conf. (MERCon)*, Moratuwa, Sri Lanka, Jul. 2021, pp. 602–607.

[97] C. Kang and Z. Zhang, "Application of LSTM in short-term traffic flow prediction," in *Proc. IEEE 5th Int. Conf. Intell. Transp. Eng. (ICITE)*, Beijing, China, Sep. 2020, pp. 98–101.

[98] P. Fafoutellis, E. I. Vlahogianni, and J. D. Ser, "Dilated LSTM networks for short-term traffic forecasting using network-wide vehicle trajectory data," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Rhodes, Greece, Sep. 2020, pp. 1–6.

[99] X. Meng, H. Fu, L. Peng, G. Liu, Y. Yu, Z. Wang, and E. Chen, "D-LSTM: Short-term road traffic speed prediction model based on GPS positioning data," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 2021–2030, Mar. 2022.

[100] M. Karimzadeh, S. M. Schwegler, Z. Zhao, T. Braun, and S. Sargento, "MTL-LSTM: Multi-task learning-based LSTM for urban traffic flow forecasting," in *Proc. Int. Wireless Commun. Mobile Comput. (IWCMC)*, Harbin, China, Jun. 2021, pp. 564–569.

[101] W. Zhu, J. Wu, T. Fu, J. Wang, J. Zhang, and Q. Shangguan, "Dynamic prediction of traffic incident duration on urban expressways: A deep learning approach based on LSTM and MLP," *J. Intell. Connected Vehicles*, vol. 4, no. 2, pp. 80–91, 2021.

[102] S. Zhao, S. Lin, Y. Li, J. Xu, and Y. Wang, "Urban traffic flow forecasting based on memory time-series network," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Rhodes, Greece, Sep. 2020, pp. 1–6.

[103] Q. Li, H. Tan, Y. Wu, L. Ye, and F. Ding, "Traffic flow prediction with missing data imputed by tensor completion methods," *IEEE Access*, vol. 8, pp. 63188–63201, 2020.

[104] D. Zhao and F. Chen, "A hybrid ensemble model for urban lane-level traffic flow prediction," *IEEE J. Radio Freq. Identificat.*, vol. 6, pp. 820–824, Oct. 2022.

[105] Z. Ara and M. Hashemi, "Traffic flow prediction using long short-term memory network and optimized spatial temporal dependencies," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Orlando, FL, USA, Dec. 2021, pp. 1550–1557.

[106] W. Fang, W. Cai, B. Fan, J. Yan, and T. Zhou, "Kalman-LSTM model for short-term traffic flow forecasting," in *Proc. IEEE 5th Adv. Inf. Technol., Electron. Autom. Control Conf. (IAEAC)*, Chongqing, China, vol. 5, Mar. 2021, pp. 1604–1608.

[107] C. Zhou, J. Du, R. Chen, and A.-H.-T. Jieensi, "Urban road checkpoints traffic flow prediction model of hyper-parameter based on PSO optimization LSTM," in *Proc. 2nd Int. Conf. Electron., Commun. Inf. Technol. (CECIT)*, Sanya, China, Dec. 2021, pp. 1262–1268.

[108] B. Yan, G. Wang, J. Yu, X. Jin, and H. Zhang, "Spatial–temporal Chebyshev graph neural network for traffic flow prediction in IoT-based ITS," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 9266–9279, Jun. 2022.

[109] S. A. A. Shah, K. Illanko, and X. Fernando, "Deep learning based traffic flow prediction for autonomous vehicular mobile networks," in *Proc. IEEE 94th Veh. Technol. Conf.*, Norman, OK, USA, Sep. 2021, pp. 1–5.

[110] A. K. Azad and M. S. Islam, ''Traffic flow prediction model using Google map and LSTM deep learning,'' in *Proc. IEEE Int. Conf. Telecommun. Photon. (ICTP)*, Dhaka, Bangladesh, Dec. 2021, pp. 1–5.

[111] Y. Zhang and D. Xin, ''Dynamic optimization long short-term memory model based on data preprocessing for short-term traffic flow prediction,'' *IEEE Access*, vol. 8, pp. 91510–91520, 2020.

[112] J. Liu, F. Zheng, X. Liu, and G. Guo, ''Dynamic traffic flow prediction based on long-short term memory framework with feature organization,'' *IEEE Intell. Transp. Syst. Mag.*, vol. 14, no. 6, pp. 221–236, Nov. 2022.

[113] L. Xing and W. Liu, ''A data fusion powered bi-directional long short term memory model for predicting multi-lane short term traffic flow,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 16810–16819, Sep. 2022.

[114] C.-H. Lin, Y.-C. Lin, Y.-J. Wu, W.-H. Chung, and T.-S. Lee, ''A survey on deep learning-based vehicular communication applications,'' *J. Signal Process. Syst.*, vol. 93, no. 4, pp. 369–388, Aug. 2020.

ANWAR KHAN received the M.Sc. degree in electronics from the Department of Electronics, University of Peshawar, in 2006, the M.Phil. degree in electronics from Quaid-i-Azam University, Islamabad, Pakistan, in 2009, and the Ph.D. degree in electronics in underwater wireless sensor networks from the Department of Electronics, Quaid-i-Azam University, in 2018. He has been with the Department of Electronics, University of Peshawar, since 2009, as a Faculty Member. He has published articles with IEEE, Springer, Elsevier, Taylor & Francis, MDPI, and *KSII Transactions on Internet and Information Systems*. He is also a reviewer with these publishers and one of the members of the Technical Program Committee of several conferences. His current research interests include underwater wireless sensor networks, relayed communications, energy harvesting, deep/machine learning, intelligent transportation, cloud/fog/edge computing, and game-theoretic optimization.

MOSTAFA M. FOUDA (Senior Member, IEEE) received the B.S. (as the valedictorian) and M.S. degrees in electrical engineering from Benha University, Egypt, in 2002 and 2007, respectively, and the Ph.D. degree in information sciences from Tohoku University, Japan, in 2011. He was an Assistant Professor with Tohoku University and a Postdoctoral Research Associate with Tennessee Technological University, TN, USA. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, Idaho State University, ID, USA. He is a Full Professor with Benha University. He has (co)authored more than 160 technical publications. His current research focuses on cybersecurity, communication networks, signal processing, wireless mobile communications, smart healthcare, smart grids, AI, and the IoT. He has guest-edited a number of special issues covering various emerging topics in communications, networking, and health analytics. He is currently serving on the Editorial Board of IEEE Transactions on Vehicular Technology (TVT), IEEE Internet of Things Journal (IoT-J), and IEEE Access. He has received several research grants, including NSF Japan-U.S. Network Opportunity 3 (JUNO3).

DINH-THUAN DO (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in electrical engineering from Vietnam National University-Ho Chi Minh City (VNU-HCM), in 2007 and 2012, respectively. Prior to joining academia, he was a Senior Engineer in the telecommunications industry with VinaPhone Mobile Network (the biggest cellular network provider in Vietnam), from 2003 to 2009. Before joining the University of Mount Union, he was a Research Scientist with the University of Colorado Denver, in 2022, and The University of Texas at Austin, in 2021, and as an Assistant Professor with Asia University, Taiwan, from 2020 to 2021. His publications include over 120 SCIE/SCI-indexed journal articles, five edited books (IET, Springer), and over 50 international conference papers. He was a recipient of the 2015 Golden Globe Award by the Vietnamese Ministry of Science and Technology (top ten outstanding scientists nationwide). He also received the Medal of Creative Young Talents, in 2015. He was named in the top 14 highly cited scientists with Asia University, Taiwan, in 2021 (Stanford's list of top 2% scientists in the world, October 2020, October 2021, and October 2022). He is currently an Associate Editor of IEEE Transactions on Vehicular Technology and *Computer Communications* (Elsevier). He has also served as a Lead Guest Editor/Guest Editor in more than 20 special issues of journals, such as *Physical Communications* (Elsevier) and *Annals of Telecommunications* (Elsevier).

ABDULAZIZ ALMALEH received the master's and Ph.D. degrees in information science from the University of Pittsburgh. He is a renowned academic and researcher in the field of information science, specializing in machine learning, smart cities, and smart infrastructure. He has established himself as an expert in his domain, with a strong understanding of emerging technologies and research methodologies. Throughout his academic journey, he has demonstrated a deep commitment to advancing the field of information science, focusing on the practical applications of machine learning and its implications for the development of smart cities and infrastructure. His research interests encompass a wide variety of subjects, including data analytics, artificial intelligence, the Internet of Things (IoT), and the integration of these technologies into urban planning and development processes. He has an impressive publication record that attests to his dedication and expertise in his chosen field. His numerous works have been published in well-respected journals and conference proceedings, contributing to the collective knowledge, and understanding of machine learning applications in the creation of intelligent urban environments. His research findings have greatly influenced the way in which advanced technologies are applied to urban planning, offering practical solutions to the modern challenges faced by cities around the world.

ATIQ UR RAHMAN received the B.S. degree in computer science from the University of Peshawar, in 2004, the M.S. degree in computer system engineering from the GIK Institute of Engineering Sciences and Technology, Pakistan, in 2008, and the Ph.D. degree in information technology from Universiti Teknologi PETRONAS, Malaysia, in 2013. He was a System Engineer with Pakistan's National Database and Registration Authority, from 2004 to 2006. Later, he was an Assistant Professor, the Head of the Computer Science Department, and the Head of the Quality and Academic Accreditation Unit with Northern Border University, Saudi Arabia, from 2013 to 2019. He is currently an Assistant Professor with the Higher Colleges of Technology, United Arab Emirates. His research focuses on energy-efficient design, routing, deployment, coverage, and quality of service in wireless sensor networks and underwater sensor networks. He is a Distinguished Senior Member of esteemed professional organizations, such as the Universal Association of Computer and Electronics Engineers (UACEE), the Institute of Research Engineers and Doctors (IRED), and the American Society for Engineering Education (ASEE). With a significant number of research publications in renowned journals and conferences, he actively contributes as a reviewer to several journals. He is a technical program committee member at international conferences. In addition, he also participated in the ''Fundamental and Advanced Assessment Workshop'' organized by ABET, USA.