MF-TCPV: A Machine Learning and Fuzzy Comprehensive Evaluation-Based Framework for Traffic Congestion Prediction and Visualization

LEIXIAO LI, HAO LIN[ID], JIANXIONG WAN[ID], ZHIQIANG MA, AND HUI WANG

College of Data Science and Application, Inner Mongolia University of Technology, Hohhot 010080, China
Inner Mongolia Autonomous Region Engineering and Technology Research Center of Big Data Based Software Service, Hohhot 010080, China

Corresponding authors: Hao Lin (suzukaze_aoba@foxmail.com) and Leixiao Li (llxhappy@126.com)

ABSTRACT A framework for traffic congestion prediction and visualization based on machine learning and Fuzzy Comprehensive Evaluation named MF-TCPV is proposed in this paper. The framework uses DataX and DataV to implement the integration of multi-source heterogeneous traffic data and the visualization of congestion prediction results. A deep prediction model named LSTM-SPRVM based on deep learning algorithms, machine learning algorithms, and Spark parallelization technology for the prediction of traffic congestion features in the future is proposed. In MF-TCPV, traffic congestion is divided into six levels based on Fuzzy Comprehensive Evaluation and traffic congestion features such as average speed, road occupancy rate, and traffic flow density. MF-TCPV is validated based on the real data of Whitemud Drive in Canada. The experimental results demonstrate that MF-TCPV is capable of predicting the traffic congestion accurately and displaying prediction results visually. LSTM-SPRVM is better than other existing deep learning models in terms of prediction accuracy, and MF-TCPV can intuitively visualize the prediction results of traffic congestion.

INDEX TERMS Intelligent traffic systems, traffic congestion prediction, machine learning, fuzzy comprehensive evaluation, visualization, spark.

I. INTRODUCTION

Intelligent Traffic Systems (ITS) is an integrated system that combines advanced science and technology such as electronic information technology, data communication technology, sensor technology, control theory, operational research, and artificial intelligence to improve the transportation industry. Applications and researches of ITS involve highway, railway, civil aviation, water carriage, and other modes of transportation [1]. Traffic congestion prediction, also call as traffic flow state prediction, is an essential part of ITS. Untimely and unreasonable traffic congestion prediction will cause huge economic losses to the society, the significant increase of exhaust gas pollution, and

The associate editor coordinating the review of this manuscript and approving it for publication was Nabil Benamar[ID].

the reduction of citizens' living standards [2], [3]. Therefore, congestion prediction is of great value to traffic management.

The treatment of traffic congestion should focus on prevention, that is, can predict the changing trend of the traffic state in a short time based on the current information of the traffic flow and to provide early warning of possible congestion. Existing researches on traffic congestion mostly focused on the prediction of traffic congestion features such as traffic volume [4] and average vehicle speed [5] or the determination of the current congestion [6], [7]. A framework for traffic congestion prediction and visualization that can consider most traffic congestion features and predict traffic congestion in the future is needed.

In this paper, we consider the prediction of traffic congestion features and the determination of the current congestion to build a framework for traffic congestion prediction and

visualization, called Machine Learning and Fuzzy Comprehensive Evaluation based Framework for Prediction and Visualization of Traffic Congestion (or simply MF-TCPV), which can evaluate short-term traffic congestion in the future by using the prediction results of traffic congestion features. We utilize some of Alibaba's excellent components to implement data integration and data visualization in MF-TCPV. To predict traffic congestion features, we develop a deep prediction model combining deep learning and machine learning. We expand our previous researches to train the deep model in parallel based on parametric parallelism to improve training efficiency [8], [9]. The main contributions of this paper are summarized as follows:

- A novel framework named MF-TCPV is proposed for the prediction and the visualization of traffic congestion, which can be divided into three layers——raw data layer, data processing layer, and data presentation layer.
- In the data processing layer, a deep prediction model based on deep learning and machine learning is proposed for congestion evaluation parameter prediction, called long short-term memory networks combined relevance vector machine based on spark parallelization (or simply LSTM-SPRVM). Fuzzy Comprehensive Evaluation is used to divide the traffic congestion into six levels by prediction results of LSTM-SPRVM. The weight of each traffic congestion feature is determined by the entropy method.
- In the data presentation layer, we set up a cache database and design the visualization of traffic congestion prediction by DataV.
- We validate the performance of MF-TCPV based on the real data of Whitemud Drive in Canada.

The remainder of this paper is organized as follows: the related works are discussed in Section 2. Section 3 presents the MF-TCPV framework. Section 4 describes the experimental results. Finally, Section 5 draws the conclusion and future research directions.

II. RELATED WORK

A. PREDICTION OF TRAFFIC CONGESTION

The existing frameworks for predicting traffic congestion have significant drawbacks in terms of their ability to maximize the use of traffic flow data and evaluate traffic congestion in the future. Some frameworks which cannot predict traffic congestion in the future directly use clustering and classification algorithms to identify the current traffic congestion situation [6], [7], [10]. Ruiqi *et al.* [6] proposed a congestion recognition method based on a multi-class support vector machine (SVM). Firstly, the congestion was divided into three kinds: "traffic," "congestion," and "traffic paralysis" by using pattern recognition technology, and the road traffic situation was divided into "traffic" and "congestion" by using SVM, based on this, the quadratic discriminant of "congestion" and "traffic paralysis" were carried out to "congestion" state. Zhang *et al.* [7] collected traffic flow

velocity, traffic flow density, and traffic volume to judge the traffic congestion level by grey relational membership degree rank clustering algorithm. Zhu *et al.* [10] extracted feature parameters from video data by gray level co-occurrence histogram and histogram of oriented gradient. Then they used SVM to classify traffic congestion levels. These frameworks and methods are simple in process but cannot predict traffic congestion in the future.

Most researchers used traffic congestion features obtained from the original traffic flow data to evaluate congestion. Traffic congestion features can be divided into three categories: traffic congestion features based on the V/C ratio [11], traffic congestion features based on the time ratio [12], and traffic congestion features based on economics [13]. Traffic congestion features based on the V/C ratio include traffic flow, average speed, road occupancy rate, traffic flow density, etc. Due to its operability and collectability, traffic congestion features based on the V/C ratio were widely used. In previous researches, the determination of traffic congestion is based on a specific feature such as vehicle speed and occupancy rate [14], [15]. Islam *et al.* [14] proposed a smart traffic control system based on the measurement of traffic density using the real-time video processing technique, which can detect the traffic congestion. Yang *et al.* [15] proposed a framework for the prediction of traffic congestion. In the framework, backpropagation neural network (BPNN) was used to predict traffic volume in the future. Road occupancy rate and traffic flow density were calculated by prediction results. Finally, traffic congestion was evaluated according to road occupancy rate and traffic flow density in the future.

In addition to traffic congestion features, there are other types of data used by some frameworks to describe congestion. These data include real-time monitoring of video data [16], numerous GPS data [17], [18], and road network map data containing traffic flow information [19]. These data are difficult to obtain, resulting in the lack of practicality of the framework which using these data. Furthermore, the existing frameworks' design of data integration and visualization is not detailed enough, which will also affect the practicality of frameworks. Besides, there are some frameworks based on predicting a single traffic congestion feature [20], [21]. Table 1 shows the comparison of MF-TCPV with existing frameworks or methods.

Therefore, there is a need to develop a framework as a benchmark to:

(1) use most of traffic congestion features based on the V/C ratio.

(2) avoid using hard-to-obtain data.

(3) have the ability for short-term prediction.

(4) have data integration and visualization functions.

Our proposed framework named MF-TCPV addresses these challenges of existing frameworks in this research work. In MF-TCPV, we use LSTM-SPRVM to predict traffic volume and vehicle speed. Traffic congestion features such as average speed, road occupancy rate, and traffic flow density are calculated based on the prediction result to describe

TABLE 1. Comparison of MF-TCPV with exiting frameworks or method.

Framework or method	Feature				Other feature	Machine learning-based prediction	Deep learning-based prediction	Ability for short-term prediction	Component	
	Traffic volume	Average speed	Road occupancy rate	Traffic flow density					Data integration	Visualization
[6]	√					√				
[7]	√		√	√		√				√
[10]					√	√				
[14]				√					√	
[15]	√		√	√			√	√		√
[16]					√	√	√			
[17]					√	√			√	√
[18]	√		√		√			√		
[19]					√		√	√		√
[20]	√					√		√		
[21]	√						√	√		
MF-TCPV	√	√	√	√		√	√	√	√	√

congestion by Fuzzy Comprehensive Evaluation. The calculation amount of Fuzzy Comprehensive Evaluation is small and no need to train a complex classification model. Besides, Fuzzy Comprehensive Evaluation combines with the predicting of traffic congestion features can predict future congestion conditions in the future. MF-TCPV is more practical than other classification algorithms and clustering algorithms in the application scenario of traffic congestion prediction.

B. PREDICTION OF TRAFFIC CONGESTION FEATURES
The prediction problem of traffic congestion features belongs to the time series prediction problem [22]. The prediction accuracy of traffic congestion features determines the availability of Fuzzy Comprehensive Evaluation results. At present, there are two main types of research about predicting traffic congestion features. The researches of the first type were driven by models. They predicted traffic congestion features by using Auto Regression Moving Average Model (ARMA) [23], Autoregressive Integrated Moving Average Model (ARIMA) [24], Kalman Filtering [25], and other models. These methods usually assume that time series data is generated by a linear process and establish the dependence in series data based on a linear relationship. Therefore, although the calculation is simple, the accuracy is poor.

The researches of the second type, such as SVM [26], Recurrent Neural Network (RNN) [27], [28], RVM [4], and other machine learning and deep learning algorithms were driven by data. These models have excellent generalization and suitable for most data. Where RVM has extremely fast test speed and is widely used to predict traffic congestion features. Xin et al. [29] used Genetic Algorithm (GA) to optimize RVM to prove that heuristic algorithms can improve the capability of RVM. Li et al. [30] analyzed the correlation between various traffic flow features and used LSTM to predict vehicle speed. Shen et al. [31] proposed a hybrid algorithm called Chaos Simulated Annealing Algorithm (CSA) and used it to optimize RVM to predict traffic flow. This method has better prediction accuracy than model-driven methods. But none of them considered the efficiency of heuristic algorithms as parameter optimization algorithms. For common machine learning algorithms such as SVM and RVM, we found that the time consumed by the parameter optimization algorithm accounts for 80%-95% of the overall training time [8], [9]. The low efficiency of parameter optimization will significantly affect the training efficiency of the prediction model.

On the other hand, the deep model that extracts features firstly from the traffic flow data by deep learning algorithms gradually favored by scholars. LSTM-SVR is the most commonly used deep model to fit traffic flow data. Zheng et al. [32] proposed a deep model named LSTM-SVR, which uses the advantage of LSTM in processing time series data to extract features of traffic flow and the high-dimensional expression ability of SVM to predict subway traffic flow. The result shows that LSTM-SVR is superior to LSTM and SVR. Li et al. [33] extracted multi-dimensional traffic flow features through Deep Belief Networks (DBN) to create a new feature dataset. Then input it into SVR to predict short-term traffic flow. The method considers the number of lanes, the time ratio of traffic lights, whether to turn left, and other information to obtains better results. Besides LSTM and DBN, deep learning algorithms such as CNN and GRU had also been applied to feature extraction [34], [35].

According to literature research, it can be found that LSTM and GRU are suitable for extracting single-dimensional and low-dimensional features of traffic flow data. DBN and CNN are suitable for extracting high-dimensional features of traffic flow data. RVM, which is more prominent in machine learning algorithms, is rarely used in deep models. Based on the
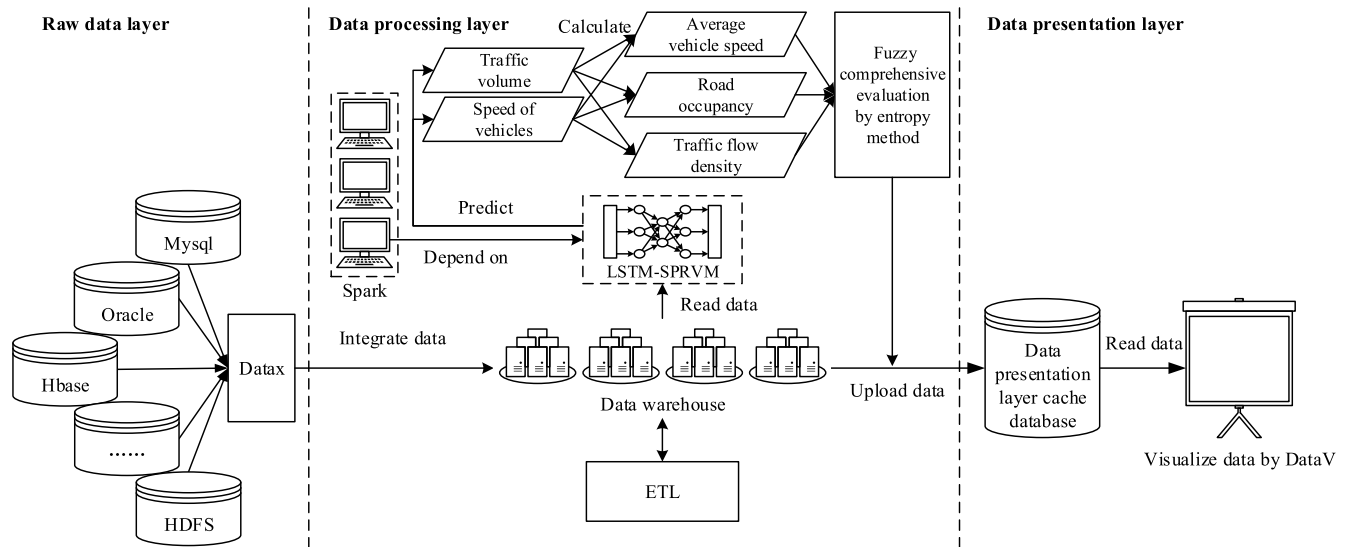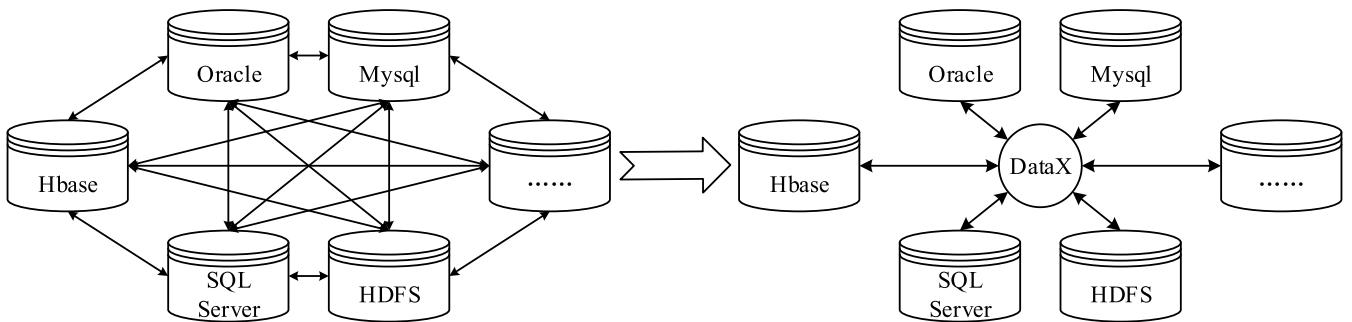
FIGURE 1. Architecture of MF-TCPV.

FIGURE 2. Design concept of DataX.

above literature review, we propose LSTM-SPRVM as a part of MF-TCPV to predict traffic congestion features.

III. MF-TCPV FRAMEWORK

This section presents the detailed description of MF-TCPV. Fig. 1 illustrates the architecture of MF-TCPV.

A. RAW DATA LAYER

Raw traffic data is scattered in different data sources, such as relational databases (MySQL, Oracle), NoSQL data storage (HBase, Hive), and unstructured data storage (HDFS). It is necessary to collect, organize, clean the data from different data sources and load it into a new data source. To begin MF-TCPV, we introduce DataX into the framework. DataX is a tool for offline synchronization and integration of heterogeneous data sources, which design concept is shown in Fig.2 [36].

DataX is widely used within Alibaba and assumes all Alibaba's offline synchronization services of big data. More than 80,000 simultaneous operations are completed every day by DataX, and the transmission volume of daily data

exceeds 300TB. DataX is fully capable of traffic data integration. In MF-TCPV, multi-source traffic data is loaded into a data warehouse by DataX. The advantage of this design is that complex meshed multi-source data links are no longer needed.

B. DATA PROCESSING LAYER

1) LSTM-SPRVM FOR PREDICTING OF TRAFFIC CONGESTION FEATURES

Architecture The principal part of data processing layer is LSTM-SPRVM for the prediction of traffic congestion features. In LSTM-SPRVM, LSTM is chosen for feature extraction of traffic flow data, RVM is selected as the output layer of LSTM-SPRVM, the heuristic algorithm is selected as the parameter optimization algorithm of RVM, and Spark is selected for parallel training of LSTM-SPRVM to improve training efficiency. LSTMs can fit the statistical relationship of time series data through their structure to extract features from original data [37]. Fig.3 shows the feature extraction and prediction process in LSTM-SPRVM.

In LSTM-SPRVM, raw data is transformed into multiple "windows" and input them into LSTMs for feature extraction
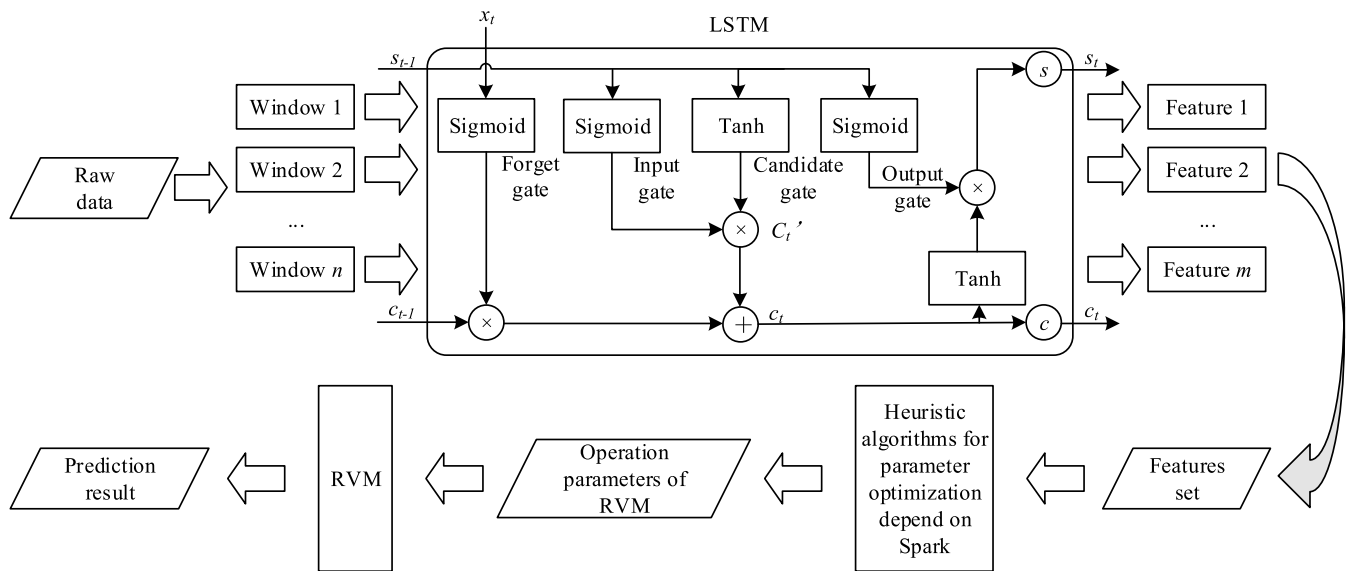
FIGURE 3. Feature extraction and prediction process in LSTM-SPRVM.

to generate a new feature dataset. According to the experimental data and results, we set *n* to 10. The LSTM layers and RVM of our deep prediction model as shown in Fig. 4.

To extract traffic data features, a matrix with traffic data is input into a single LSTM layer. Due to the sparsity of traffic data, optimizers with the adaptive learning rate will perform better. So, we choose NADAM as the optimizer of LSTM, which uses Nesterov to update the gradient-based on ADAM [38]. When a neural network is trained with a small data set, it is easy to cause overfitting. To prevent overfitting, the performance of the neural network can be improved by adding a dropout layer in the neural network [39]. In LSTM-SPRVM, 5% of neurons will be disconnected in the dropout layer. Output the extracted features through a Dense layer. Finally, the new feature set is input into RVM for prediction.

In past research, traffic flow data was mostly processed as time series data with intervals of 10-minutes, 5-minutes, or 2-minutes. For the time series prediction problem, the shorter the time interval, the more practical application value it has and the nonlinearity of the data has [40], [41]. Meanwhile, for traffic flow data, short intervals will increase the sparsity of the data. The greater the sparsity of data and the greater the difficulty of prediction it has. Moreover, too short intervals will result in decision-makers not having time to make reasonable decisions in a short decision-making period (e.g., Observation-Orientation-Decision-Action, OODA). After careful consideration, we process traffic flow data with a 5-minute interval. The method of predicting traffic speed is the same as the method of predicting traffic flow.

Parameter optimization Single kernel functions have poor capability when the samples are unevenly distributed in the high-dimensional space [42]. So, we combine single kernel functions linearly to a construct combined kernel function as
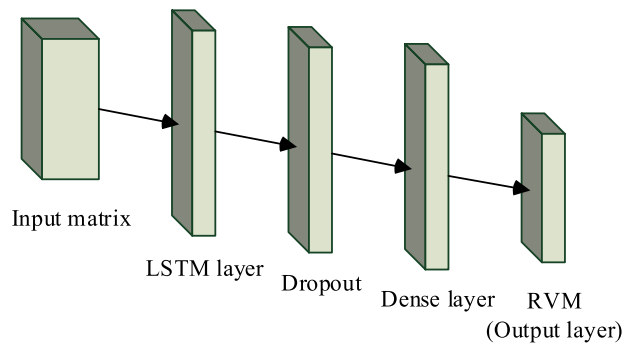

FIGURE 4. Architecture of LSTM layers and RVM layers.

shown in (1).

$$f(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right)\lambda + (1 - \lambda)\left[\gamma(xx_i + 1)^d + c\right] \tag{1}$$

There is no need to prove the availability of the combined kernel functions, because kernel functions of RVM do not need to meet Mercer's theorem. The combined kernel function makes RVM get local learning ability of Gaussian kernel and generalization of Polynomial kernel [43]. Where

σ is the width of the combined kernel function. d determines the distribution of the data in high-dimensional space. λ is weight coefficient and meets $0 \le \lambda \le 1$.

σ, d, λ in (1) are also hyperparameters which need to be optimized. The purpose of the parameter optimization algorithm is to find a $\{\sigma_{best}, \lambda_{best}, d_{best}\}$. Input the $\{\sigma_{best}, \lambda_{best}, d_{best}\}$ as parameters of RVM to calculate the accuracy $p_{Accuracy}$ of the prediction model. The mathematical

model of the parameter optimization algorithm is as follows

$$\max \text{Accuracy} = \text{RVM}(P)$$

$$s.t. \begin{cases} P = \{\sigma_{best}, \lambda_{best}, d_{best}\} \\ -2^{\sigma_{\min}} \leq \sigma_{best} \leq 2^{\sigma_{\max}} \\ d_{\min} \leq d_{best} \leq d_{\max} \\ 0 \leq \lambda_{best} \leq 1. \end{cases} \quad (2)$$

where $-2^{\sigma_{\min}}$ and $2^{\sigma_{\max}}$ mean variation range of σ_{best}. Generally, $\sigma_{\min} = \sigma_{\max} = 3$. $d_{\min}$ and $d_{\max}$ mean variation range of d_{best}. Generally, $d_{\min} = 0$ and $d_{\max} = \infty$.

GA is one of the most commonly used parameter optimization algorithms of RVM [44]–[46]. Moreover, GA's optimization performance is generally better than common algorithms [47], [48]. Therefore, we use GA to solve the mathematical model which is shown in (2).

Spark and MapReduce Currently, the most commonly used parallel computing frameworks are MapReduce and Spark of Apache. When executing tasks, MapReduce uses a multi-process model, while Spark uses a multi-thread model. Moreover, Spark is based on memory computing, which is more suitable for machine learning algorithms, iterative algorithms, and stream computing [49]. Since the heuristic algorithms are equivalent to iterative optimization algorithms, we choose Spark for parallelization.

Parallel design Add parameter optimization algorithms to RVM will double the training time of RVM. For reducing the training time of RVM, we design a parallel training method of RVM based on parametric parallelism. The time-consuming analysis of GA can be seen in our previously published paper [9]. It shows that fitness calculation consumes the most time and accounts for more than 90% of the total time. The overall parallelization process based on parametric parallelism is shown in Fig. 5.

The parallelization method designed in this paper mainly relies on Spark's unique data format——Resilient Distributed Datasets (RDD). In traditional parallel methods based on data parallelism, the training data is created as RDD [50]–[52]. In the parallelization method based on parametric parallelism proposed in this paper, the initial population of GA is created as RDD. The parallelization method based on parametric parallelism is suitable for prediction problems on small and medium-sized data, such as time series prediction. The pseudocode of GA combined with the parallelization method designed in this paper is shown in Procedure 1.

Procedure 1 evaluates the initial population by map function and then updates the population according to the evaluation results. Finally, Procedure 1 finds the index of the best individual that meets the conditions for terminating the iteration and decodes the individual into $\{\sigma_{\text{best}}, \lambda_{\text{best}}, d_{\text{best}}\}$.

2) CALCULATION OF TRAFFIC CONGESTION FEATURES

This section is responsible for using prediction results of LSTM-SPRVM to calculation average speed, road

---

Procedure 1 GA Combined with Parallelization Method

Input: *conf*: Initialization parameters of Spark; *lenchrom*: Chromosome length; *bound*: Individual bound; *sizepop*: Population size; *k*: Number of data partitions; *pcross*: Probability of crossover; *pmutation*: Probability of mutation

Output: *resultParameters*: Optimal parameters of RVM

1. $sc \leftarrow$ sparkConf(*conf*)
2. **For** *i* **IN** *sizepop* **DO**
3.   *chrom*[*i*] $\leftarrow$ Code (*lenchrom*, *bound*)
4. **END FOR**
5. *populationRdd* $\leftarrow$ *sc*.parallelize (*chrom*, *k*)
6. *FitnessRdd* $\leftarrow$ *populationRdd*.map (getFitness (), *k*)
7. *fitness* $\leftarrow$ *FitnessRdd*.collect ()
8. *times* $\leftarrow$ 0
9. **WHILE** (*minFitness* $\geq$ *bestFitness* **OR** *times* $\geq$ *maxTimes*) **DO**
10.   *newChrom* $\leftarrow$ SelectCrossMutation (*chrom*, *fitness*, *sizepop*, *pcross*, *pmutation*)
11.   *populationRdd* $\leftarrow$ *sc*.parallelize (*newChrom*, *k*)
12.   *FitnessRdd* $\leftarrow$ *populationRdd*.map (getFitness (), *k*)
13.   *fitness* $\leftarrow$ *FitnessRdd*.collect ()
14.   *times* $\leftarrow$ *times* ++
15. **END WHILE**
16. *bestFitness* $\leftarrow$ min (*fitness*)
17. **FOR** *k* **IN** *sizepop* **DO**
18.   **IF** (*fitness*[*k*] == *bestFitness*) **DO**
19.   *bestIndex* $\leftarrow$ *k*
20. **END FOR**
21. *resultParameters* $\leftarrow$ Decode (*newChrom* [*bestIndex*])

---

occupancy, and traffic flow density in the future based. The specific calculation method is shown below.

Average speed Average speed refers to the average distance traveled by all vehicles on the road in a unit time. Since one road may correspond to multiple lanes, the calculation method is as follows

$$\overline{v_i} = \begin{cases} \dfrac{1}{N} \displaystyle\sum_{i=1}^{N} v_i, & \text{no lane with zero speed} \\ \max(v_i), & \text{lane with zero speed.} \end{cases} \quad (3)$$

where v_i means average speed corresponding to the ith lane; N means the total number of lanes on the road. In general, the higher the average speed, the smoother the road. The lower the average speed, the more blocked the road.

Road occupancy Road occupancy means the ratio of the actual traffic flow to the maximum capacity in a specific section. It reflects the actual load capacity of the road.
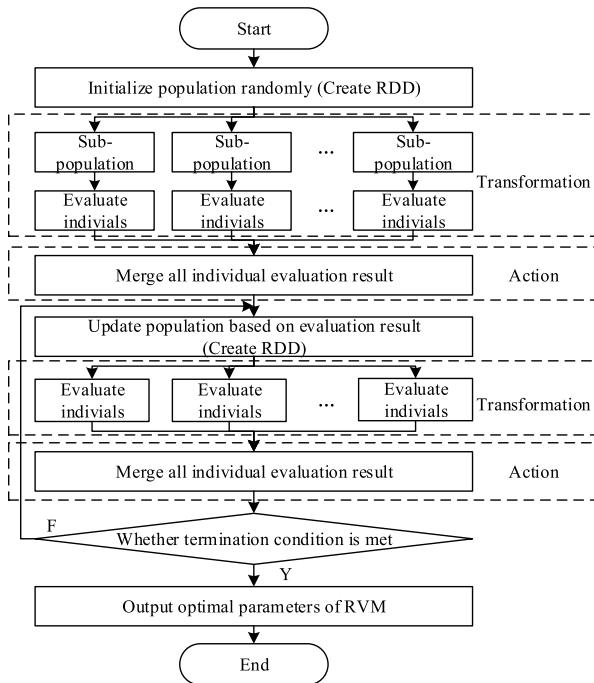
FIGURE 5. Parallelization process of RVM using parametric parallelism.

The calculation method is as follows

$$S = \frac{V}{C}. \tag{4}$$

where V means current traffic volume; C means the maximum capacity.

Traffic flow density Traffic flow density means the number of vehicles in particular length of road within a unit time. The calculation method is as follows

$$D = \frac{f}{v}. \tag{5}$$

where f means traffic volume; v means average speed.

3) FUZZY COMPREHENSIVE EVALUATION

This section is responsible for evaluating the traffic congestion in the future by entropy method and Fuzzy Comprehensive Evaluation method. According to related research such as the "Highway Capacity Manual," we divide traffic congestion into six levels: especially unblocked, unblocked, light congestion, moderate congestion, heavy congestion, and lock up in MF-TCPV.

Suppose the prediction results of LSTM-SPRVM are $R = (R_1, R_2, \cdots, R_i)^T$. We have selected three traffic congestion features (i.e., average speed, road occupancy, and traffic flow density), so $i = 3$. The evaluation matrix in Fuzzy Comprehensive Evaluation is calculated by

$$R = \begin{pmatrix} R_1 \\ R_2 \\ R_3 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{16} \\ r_{21} & r_{22} & \cdots & r_{26} \\ r_{31} & r_{32} & \cdots & r_{36} \end{pmatrix}. \tag{6}$$

For each factor i, the membership degree r_{ij} of the jth level is obtained through a membership function. We choose the

trapezoidal membership function in this paper. The trapezoidal membership function is shown in Fig. 6.
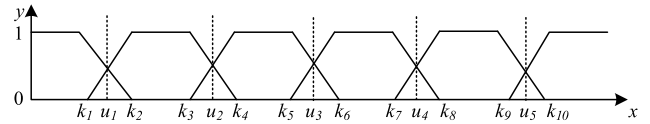

FIGURE 6. Trapezoidal membership function.

Where $\{u_1, u_2, \cdots, u_5\}$ is the threshold range of each factor indicator. $\{k_1, k_2, \cdots, k_{10}\}$ is the proximity threshold of each factor indicator. In the Y-axis, "1" means that the element belongs to this level and "0" means not. For positive factors such as road occupancy and traffic flow density, the status of the function is from "especially unblocked" to "lock up" (i.e., The lower road occupancy or traffic flow density, the smoother the road). For negative factors such as average speed, the status of the function is from "lock up" to "especially unblocked" (i.e., The greater speed, the smoother the road). According to the membership function, the calculation formula of r_{ij} in (6) is as follows

$$r_{i1} = \begin{cases} 0 & R_1 \le k_1 \\ \dfrac{k_2 - R_i}{k_2 - k_1} & k_1 < R_1 < k_2 \\ 1 & R_1 \ge k_2, \end{cases} \tag{7}$$

$$r_{ij} = \begin{cases} 0 & x \le k_i \text{ or } x \ge k_{i+3} \\ \dfrac{R_i - k_i}{k_{i+1} - k_i} & k_i < x < k_{i+1} \\ 1 & k_{i+1} \le x \le k_{i+2} \\ \dfrac{k_{i+3} - x}{k_{i+3} - k_{i+2}} & k_{i+2} < x < k_{i+3}, \end{cases} \quad (j = 2, 3, 4, 5), \tag{8}$$

$$r_{i6} = \begin{cases} 0 & x \le k_9 \\ \dfrac{R_i - k_9}{k_{10} - k_9} & k_9 < x < k_{10} \\ 1 & x \ge k_{10}. \end{cases} \tag{9}$$

Use Entropy method to calculate the weight of average speed, road occupancy, and traffic flow density $W = \{w_1, w_2, w_3\}$. Due to the significant difference in traffic flow between peak hours and regular hours, we calculate the weight in peak hours and regular hours, respectively. They marked as W_{peak} and W_{regular}. Calculate the Fuzzy Comprehensive Evaluation matrix by a composite operator, which calculation method is as follow

$$B = W \circ R = (b_1, b_2, \cdots, b_6). \tag{10}$$

We use $M(\cdot, \oplus)$ as the fuzzy composition operator, which can make full use of information in R. Based on the maximum membership principle, if $b_1 = \max(b_1, b_2, \ldots, b_6)$, the level of traffic congestion at that moment will be especially unblocked.

C. DATA PRESENTATION LAYER

1) DATA PRESENTATION LAYER CACHE DATABASE

In MF-TCPV, evaluation results involved in section 3.2.3 are uploaded to a data presentation layer cache database, which

only saves the data that needs to be displayed. The advantage of the design is that it can increase data security in the framework and the attack cost of potential attackers. If the data presentation layer is attacked, the raw data will not be leaked, and the attacker can only obtain the data of the data presentation layer cache database.

Compared with the amount of the raw data, the data amount that needs to be displayed is much smaller. Therefore, the size of the data presentation layer cache database does not need to be too large.

2) VISUALIZE DATA BY DataV

In MF-TCPV, we choose DataV to visualize data. DataV is a Software-as-a-Service (SaaS) visual deployment tool that enables rapid construction and cross-platform publishing of interactive visualization on large high-resolution displays (LHDs), which consists of four parts, including data importing, visual components, editor toolchain, and application publishing [53].

Datav's development objects are mainly visualization based on LHDs. This is an advantage that other visualization components on the market, such as D3.js, Dygraphs.js, and ECharts do not have.

Firstly, load the data presentation layer cache database into the data importing center of DataV, and then add visual components for visual mapping. Control components and design visual layout through editor toolchain. Finally, use the one-click SaaS publishing function provided by DataV to publish the visualization design to Wide Area Network.

IV. PERFORMANCE EVALUATION AND RESULT ANALYSIS
A. INTEGRATION OF ALL COMPONENTS IN MF-TCPV

We implement MF-TCPV based on two servers with 6-core CPUs, 64G RAM, 2T Disks, and CentOS 7.8. The specific information about the software we used is shown in table 2.

TABLE 2. Specific information of software we used

Name	Specific information
DataX	3.0
Data warehouse	MySQL-5.6.22
Spark	Spark-2.1.1-bin-hadoop2.7, pyspark-2.3.2
Data presentation layer cache database	Alibaba RDSMySQL-8.0
DataV	Enterprise Edition

DataX and Spark were deployed on Server 1. Data warehouse was deployed on Server 2. In LSTM-SPRVM, the activation function is ReLU; Batch size is 32; Epoch is 100; population size is 20; maximum number of iterations is 10; mutation rate is 0.2; crossover rate is 0.6. We use the MSE of RVM as the fitness function of GA, which calculation formula is as follows

$$fitness = \sum_{i=1}^{n} \left(s_i - s_i' \right) \Big/ N. \tag{11}$$

where N is the population size; s_i means the actual value; s_i' means the predictive value of LSTM-SPRVM. Before predicting, we normalize experimental data based on extremum to compress the data into [0,1] by

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}. \tag{12}$$

B. EXPERIMENTAL RESULTS ANALYSIS

The MF-TCPV framework used traffic flow data from Whitemud Drive motorway in Canada collected by Intelligent Transportation Research Center at University of Alberta to verify its performance [54]. The collection frequency of the data is 20 seconds. These raw data were stored in MySQL. We use the traffic volume data obtained from 6 August 2015 to 27 August 2015 as a training set and the traffic volume data obtained from 0:00 to 24:00 on 28 August 2015 as a testing set. The information about the traffic sections corresponding to the experimental data is shown in table 3.

TABLE 3. Specific information of experiment data.

Number	Road name	Orientation	Number of lanes	Monitoring point number
Sections 1	Whitemud Drive	Eastbound	Three lanes	ID 1027
Sections 2	Whitemud Drive	Westbound	Three lanes	ID 1036
Sections 3	Whitemud Drive ramp	Westbound	Two lanes	ID 1042

Because the data in station ID 1042 was collected from the Whitemud Drive ramp, so it has high randomness and fit difficultly.

1) FEATURE EXTRACTION

We should justify the choice of LSTM. In this experiment, we design a comparison between LSTM and other types of RNNs, such as GRU and Bi-LSTM. In the process of feature extraction, the change trends of their loss are shown in Fig. 7.

As shown in Fig. 7, for performance, LSTM is better than GRU and slightly worse than Bi-LSTM. We further analyze their efficiency, which results are shown in the in table 4.

TABLE 4. Efficiency of RNNs using feature extraction.

Algorithm	Epochs	Batch size	Time spent in each step
GRU	100	32	28 us
LSTM	100	32	33 us
Bi-LSTM	100	32	48 us

As table4 shows, for efficiency, Bi-LSTM is significantly lower than LSTM. After comprehensive consideration, we choose LSTM for feature extraction.

2) KERNEL FUNCTION PERFORMANCE

In MF-TCPV, we provide six kernel functions for LSTM-SPRVM. The purpose of the kernel function performance

TABLE 5. Experimental results of kernel function performance based on accuracy.

NO.	kernel function	without LSTM			with LSTM		
		ID 1027	ID 1036	ID 1042	ID 1027	ID 1036	ID 1042
1	$\exp\left(-\dfrac{\|x-x_i\|}{2\sigma^2}\right)$	0.2151	0.3368	Overfitting	0.3085	0.3555	0.6808
2	$\exp\left(-\dfrac{\|x-x_i\|^2}{2\sigma^2}\right)$	0.8520	0.8355	0.7452	0.8538	0.8385	0.7569
3	$x^T x_i$	0.8529	0.8310	0.7472	0.8541	0.8358	0.7576
4	$\gamma\left(xx_i+1\right)^d+c$	0.8514	0.8367	0.7544	0.8499	0.8397	0.7548
5	$\exp\left(-\dfrac{\|x-x_i\|}{2\sigma^2}\right)\lambda+(1-\lambda)\left[\gamma\left(xx_i+1\right)^d+c\right]$	0.8529	0.8368	0.7554	0.8538	0.8484	**0.7595**
6	$\exp\left(-\dfrac{\|x-x_i\|^2}{2\sigma^2}\right)\lambda+(1-\lambda)\left[\gamma\left(xx_i+1\right)^d+c\right]$	**0.8535**	**0.8387**	**0.7555**	**0.8543**	**0.8496**	0.7579
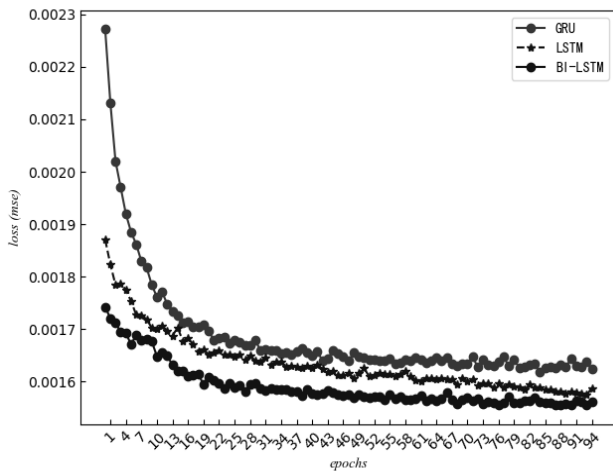
FIGURE 7. Loss change trends of RNNs using feature extraction (Because of the convenience of drawing, we did not draw the trends of the first five epochs).

experiment is to find the kernel function that is most suitable for the experiment data.

In terms of evaluation indicators, we use accuracy (i.e., 1-MAPE) to verify kernel function performance. Besides, the use of LSTM is treated as a variable factor in the experiment. The relevant experimental results are shown in table 5.

As table 5 shows, LSTM for feature extraction can improve the accuracy of RVM. In the data of ID 1027, compared with the RVM without LSTM, the accuracy of the RVM with LSTM increased by 0.75% on average (Except No.1 kernel function). In the data of ID 1036 and ID 1042, the values are 0.79% and 0.77%.

According to the results, we choose $\exp\left(-\frac{\|x-x_i\|^2}{2\sigma^2}\right)\lambda +$ $(1-\lambda)\left[\gamma\left(xx_i+1\right)^d+c\right]$ for MF-TCPV.

3) TRAFFIC VOLUME AND TRAFFIC SPEED PREDICTION
Different prediction models based on machine learning and deep learning, namely the proposed LSTM,

SPGAPSO-CKRVM, CNN-LSTM, CNN-GRU, CNN-Bi-LSTM, LSTM-SVR, LSTM-SPRVM are applied herein. Where SPGAPSO-CKRVM is the model we proposed before, that is, using parallelized GA and PSO to optimize the combined kernel RVM [9]. LSTM-SVR represents a model use SVR to predict traffic flow after features extraction by LSTM [32]. LSTM-SVR is the most commonly used deep model in recent years and the closest model to LSTM-SPRVM. Table 6 lists the MSE, RMSE, and MAPE of the different prediction models.

As shown in table 6, LSTM has the worst performance. Other deep models based on RNN, such as CNN-LSTM, CNN-GRU, and CNN-Bi-LSTM are significantly better than LSTM. LSTM-SVR is slightly better than the above models in particular data sets. In all data sets, LSTM-SPRVM is always superior to other comparable models. Compared with LSTM-SVR, the RMSE of LSTM-SPRVM reduces by 2.14%, and the MAPE reduces by 4.31%. Figure 8, Fig. 9, and Fig. 10 show the prediction results of LSTM-SPRVM.
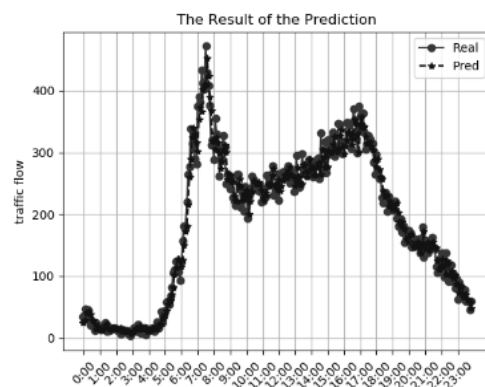

FIGURE 8. Traffic flow prediction results in ID 1027 section by LSTM-SPRVM =.

We also use LSTM-SPRVM to predict the traffic speed data in three sections. In this experiment, we use the traffic speed data obtained from 6 August 2015 to 27 August 2015 as a

TABLE 6. Comparison of traffic flow predicted by different algorithms and models.

Algorithm or model	ID 1027			ID 1036			ID 1042		
	MSE	RMSE	MAPE	MSE	RMSE	MAPE	MSE	RMSE	MAPE
LSTM [55]	493.13	22.21	0.1435	234.37	15.027	0.2075	82.88	9.10	0.3221
CNN-LSTM [34]	483.18	21.92	0.1438	235.89	14.88	0.1734	82.43	9.08	0.2659
CNN-GRU [56]	487.46	22.07	0.1429	229.62	14.92	0.1729	84.10	9.17	0.2852
CNN-Bi-LSTM [57]	477.35	21.84	0.1396	227.11	15.07	0.1622	81.79	9.04	0.2600
SPGAPSO-CKRVM [9]	392.43	19.81	0.1383	161.1	12.69	0.1589	41.09	6.41	0.2232
LSTM-SVR [32]	365.57	19.12	0.1365	171.08	13.08	0.1592	40.7	6.38	0.2217
LSTM-SPRVM	**359.48**	**18.96**	**0.1349**	**160.02**	**12.65**	**0.1437**	**38.81**	**6.26**	**0.2159**
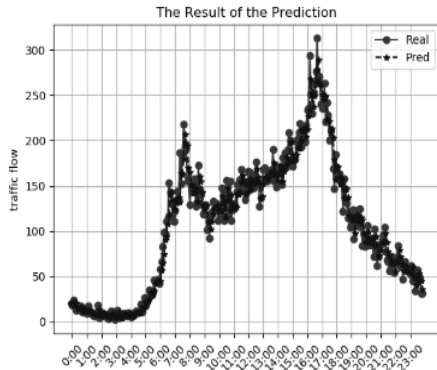
FIGURE 9. Traffic flow prediction results in ID 1036 section by LSTM-SPRVM.
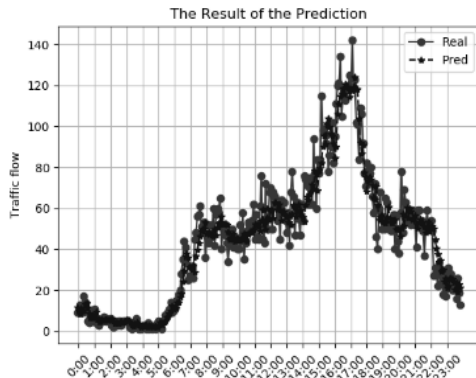
FIGURE 10. Traffic flow prediction results in ID 1042 section by LSTM-SPRVM.

training set and the traffic speed data obtained from 0:00 to 24:00 on 28 August 2015 as a testing set. The speed prediction results are shown in table 7.

As shown in table 7, the comparison results of speed prediction and traffic flow prediction are similar. The result predicted by LSTM-SPRVM is optimal. Compared with LSTM-SVR, MSE of LSTM-SPRVM reduces by an average of 9.37%, and MAPE reduces by 0.6% respectively.

Next, we compare the papers that used similar data (i.e., the traffic flow data from Whitemud Drive) for experiments. The comparison results are shown in table 8.

In the above comparison experiment, for ensuring the results' credibility, we use the training set and the test set are

the same as those used in these papers. As shown in table 8, LSTM-SPRVM is better than other models.

The scalability of the parallelized training method proposed in this paper had been proven in our previous works. LSTM-SPRVM proposed in this paper has good scalability. See [9] for details.

4) TRAFFIC CONGESTION PREDICTION RESULTS USING LSTM-SPRVM

According to "Highway Capacity Manual," the maximum road capacity of the road C in Whitemud Drive is about 180pch per lane. We calculate the road occupancy S and traffic flow density D of ID1027 section in the future by (4) and (5).

$W_{\text{peak}} = [0.32, 0.4546, 0.2253]$ and $W_{\text{regular}} = [0.424, 0.1933, 0.3827]$ are calculated by MATLAB. Finally, the traffic congestion level in ID1027 section is determined according to the maximum membership principle. The results are shown in Fig. 11.
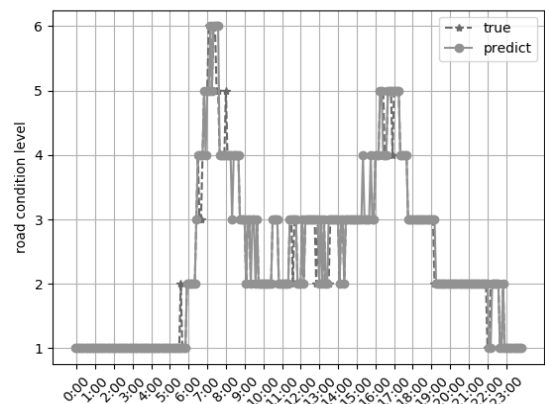

FIGURE 11. Traffic congestion prediction results in ID1027 section by LSTM-SPRVM.

The "1,2,3,4,5,6" on the y-axis in Fig. 11 correspond to especially unblocked, unblocked, light congestion, moderate congestion, heavy congestion, and lock up road conditions, respectively. As shown in Fig. 11, the ID1027 section during the morning peak is more congested than the ID1027 section during the evening peak. There was even lock up road condition in the morning peak. There are 22 moments in error, and the prediction accuracy is 92.36%. Meanwhile, it also

TABLE 7. Comparison of speed predicted by different algorithms and models.

Algorithm	Sections 1			Sections 2			Sections 3		
	MSE	RMSE	MAPE	MSE	RMSE	MAPE	MSE	RMSE	MAPE
LSTM	5.194	2.279	0.0596	5.6122	2.369	0.0625	5.5984	2.3661	0.0629
CNN-LSTM	5.1726	2.274	0.0582	5.3024	2.3027	0.0613	5.5418	2.3541	0.0622
CNN-GRU	5.0687	2.2514	0.0576	5.2964	2.3014	0.0612	5.5469	2.3552	0.0621
GA-CKRVM	4.8118	2.1936	0.0581	5.0895	2.256	0.0603	5.5145	2.3483	0.0597
CNN-Bi-LSTM	5.0794	2.2538	0.0572	5.1734	2.2745	0.0599	5.5286	2.3513	0.0603
SPGAPSO-CKRVM	4.6656	2.1601	0.0570	4.9809	2.2318	0.0586	5.4214	2.3284	0.0589
LSTM-SVR	4.1842	2.0455	0.0497	4.6537	2.1572	0.0497	4.4124	2.1005	0.0484
LSTM-SPRVM	**3.8348**	**1.9582**	**0.0418**	**4.2423**	**2.0596**	**0.0434**	**3.9295**	**1.9822**	**0.0430**

TABLE 8. Comparison results with other papers which tested on Whitemud Drive similarly.

Literature	Sections	Algorithm or model	Prediction object	Test set	Result in the paper	Result by LSTM-SPRVM
[58]	ID 1030	BP neural network (BPNN)	Traffic flow	From 4:00 pm to 12:00 pm on August 27th	R2 =0.5331	**R2 =0.9732**
[58]	ID 1030	Extreme Learning Machine (ELM)	Traffic flow	From 4:00 pm to 12:00 pm on August 27th	R2 =0.6947	**R2 =0.9732**
[41]	ID 1029	Seasonal Autoregressive Integrated Moving Average (SARIMA)	Traffic speed	From 5:00 pm to 7:00 pm on August 28th	MAPE=0.0315	**MAPE=0.0243**

illustrates the rationality of traffic congestion evaluated by Fuzzy Comprehensive Evaluation method.

C. VISUALIZATION OF TRAFFIC CONGESTION PREDICTION

The visualization of traffic congestion prediction results relies on DataV to achieve. The traffic congestion is described by a heat map. We use the data in station ID1008, ID1017, ID1037, ID1034, ID1033, ID1031, ID1029, and ID1019 for traffic congestion prediction. The settings of the training set and test set are the same as the above experiment. The "1,2,3,4,5,6" in heat map correspond to especially unblocked, unblocked, light congestion, moderate congestion, heavy congestion, and lock up road conditions respectively. The visualization result at 17:00 on August 28, 2015 is shown in Fig. 12.
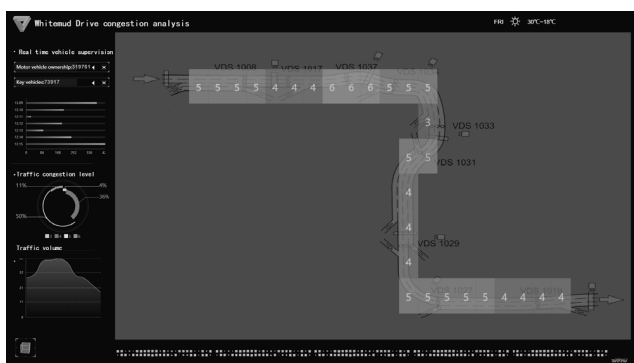
FIGURE 12. Visualization of traffic congestion prediction results in Whitemud Drive.

As shown in Fig. 12, we divide the road into multiple road sections according to the position of the surveillance camera. The visualized results are intuitive and beautiful, which can

show the congestion situation well and conform to the actual condition.

V. CONCLUSION AND FURTHER RESEARCH

To construct a framework for traffic congestion called MF-TCPV including raw data layer, data processing layer, and data presentation layer. A deep model for the prediction of traffic congestion features called LSTM-SPRVM was proposed. The real data of Whitemud Drive in Canada were utilized to verify the framework. Experimental results show that LSTM-SPRVM is superior to other methods in accuracy. The prediction accuracy of future traffic congestion can reach 92.36%, and the visualization results of MF-TCPV are intuitive. Further research will include:

- Consider more factors affecting traffic flow, such as traffic speed and volume on the adjacent roads. We plan to use a graph neural network to solve this problem.
- We used the original GA in this paper. For parameter optimization of RVM, most heuristic algorithms have the problem that the convergence speed is too fast in the early period of iterations, which leads to a decrease of the population diversity in the later period of iterations. We plan to combine RVM with complex heuristic algorithms, such as Adaptive Genetic Algorithm (AGA) to solve the problem.
- We will try to combine the science of cause and effect to explore the causes of congestion [59]. Construct a traffic congestion cause-effect diagram and transform a data-driven model into a cause-effect driven model.

REFERENCES

[1] H. Makino, K. Tamada, K. Sakai, and S. Kamijo, "Solutions for urban traffic issues by ITS technologies," *IATSS Res.*, vol. 42, no. 2, pp. 49–60, Jul. 2018, doi: 10.1016/j.iatssr.2018.05.003.

[2] J. Y. Wang, "Survey of data-centric smart city," *J. Comput Res. Dev.*, vol. 51, no. 2, pp. 239–259, Feb. 2014, doi: 10.7544/issn1000-1239.2014.20131586.

[3] H. Kamani, A. H. Mahvi, M. Seyedsalehi, J. Jaafari, M. Hoseini, G. H. Safari, A. Dalvand, H. Aslani, N. Mirzaei, and S. D. Ashrafi, "Contamination and ecological risk assessment of heavy metals in street dust of tehran, Iran," *Int. J. Environ. Sci. Technol.*, vol. 14, no. 12, pp. 2675–2682, Apr. 2017, doi: 10.1007/s13762-017-1327-x.

[4] Z. Shen, W. Wang, Q. Shen, S. Zhu, H. M. Fardoun, and J. Lou, "A novel learning method for multi-intersections aware traffic flow forecasting," *Neurocomputing*, vol. 398, pp. 477–484, Jul. 2020, doi: 10.1016/j.neucom.2019.04.094.

[5] J. Wang, Q. Gu, J. Wu, G. Liu, and Z. Xiong, "Traffic speed prediction and congestion source exploration: A deep learning method," in *Proc. IEEE 16th Int. Conf. Data Mining (ICDM)*, Barcelona, Spain, Dec. 2016, pp. 499–508, doi: 10.1109/ICDM.2016.0061.

[6] L. Ruiqi, Z. Xian, Z. Luo, and L. Lin, "Research on the intelligent judgment of traffic congestion in intelligent traffic based on pattern recognition technology," *Cluster Comput.*, vol. 22, no. 5, pp. 12581–12588, Mar. 2018, doi: 10.1007/s10586-017-1684-8.

[7] Y. Y. Zhang, "A method for traffic congestion clustering judgment based on grey relational analysis," *Int. J. Geo-Inf.*, vol. 5, no. 5, pp. 71–86, Mar. 2016, doi: 10.3390/ijgi5050071.

[8] H. Lin, L. X. Li, and H. Wang, "Survey on research and application of support vector machines in intelligent transportation system," *J. Comput. Sci. Technol.*, vol. 14, no. 6, pp. 901–917, Jun. 2020, doi: 10.3778/j.issn.1673-9418.2001029.

[9] H. Lin, L. Li, H. Wang, Y. Wang, and Z. Ma, "Traffic flow prediction using SPGAPSO-CKRVM model," *Revue d'Intell. Artif.*, vol. 34, no. 3, pp. 257–265, Jun. 2020, doi: 10.18280/ria.340303.

[10] X. Zhu, "A method of classifying the traffic smooth or congested by support vector machine," *J. Nature Univ.*, vol. 56, no. 2, pp. 278–283, Jan. 2020, doi: 10.13232/j.cnki.jnju.2020.02.014.

[11] P. Liu, "System dynamic model for coordination among stakeholders of regional freight systems," *Comput Int. Man. Syst.*, vol. 22, no. 3, pp. 859–869. May 2016, doi: 10.13196/j.cims.2016.03.030.

[12] X. Kong, J. Yang, and Z. Yang, "Measuring traffic congestion with taxi GPS data and travel time index," in *Proc. CICTP*, Beijing, China, Jul. 2015, pp. 3751–3762, doi: 10.1061/9780784479292.346.

[13] A. Chakrabartty and S. Gupta, "Traffic congestion in the metropolitan city of kolkata," *J. Infrastruct. Develop.*, vol. 6, no. 1, pp. 43–59, Jun. 2014, doi: 10.1177/0974930614543046.

[14] M. R. Islam, N. I. Shahid, D. T. ul Karim, A. A. Mamun, and M. K. Rhaman, "An efficient algorithm for detecting traffic congestion and a framework for smart traffic control system," in *Proc. 18th Int. Conf. Adv. Commun. Technol. (ICACT)*, Pyeongchang, South Korea, Jan. 2016, pp. 802–807, doi: 10.1109/ICACT.2016.7423566.

[15] X. Yang, S. Luo, K. Gao, T. Qiao, and X. Chen, "Application of data science technologies in intelligent prediction of traffic congestion," *J. Adv. Transp.*, vol. 2019, pp. 1–14, Apr. 2019, doi: 10.1155/2019/2915369.

[16] H. Cui, G. Yuan, N. Liu, M. Xu, and H. Song, "Convolutional neural network for recognizing highway traffic congestion," *J. Intell. Transp. Syst.*, vol. 24, no. 3, pp. 279–289, Apr. 2020, doi: 10.1080/15472450.2020.1742121.

[17] A. Gupta, S. Choudhary, and S. Paul, "DTC: A framework to detect traffic congestion by mining versatile GPS data," in *Proc. 1st Int. Conf. Emerg. Trends Appl. Comput. Sci.*, Shillong, India, Sep. 2013, pp. 97–103, doi: 10.1109/ICETACS.2013.6691403.

[18] S. K. Lin, "The congestion road segment prediction based on GPS trajectory data," *J. Nor. Univ.*, vol. 36, no. 11, pp. 1530–1534, Nov. 2015, doi: 10.3969/j.issn.1005-3026.2015.11.003.

[19] G. Zhou and F. Chen, "Urban congestion areas prediction by combining knowledge graph and deep spatio-temporal convolutional neural network," in *Proc. 4th Int. Conf. Electromechan. Control Technol. Transp. (ICECTT)*, Guilin, China, Apr. 2019, pp. 105–108, doi: 10.1109/ICECTT.2019.00031.

[20] L. Zhang, N. R. Alharbe, G. Luo, Z. Yao, and Y. Li, "A hybrid forecasting framework based on support vector regression with a modified genetic algorithm and a random forest for traffic flow prediction," *Tsinghua Sci. Technol.*, vol. 23, no. 4, pp. 479–492, Aug. 2018, doi: 10.26599/TST.2018.9010045.

[21] J. Wang, "DST: A deep urban traffic flow prediction framework based on spatial-temporal features," in *Knowledge Science*. Athens, GA, USA: Springer, 2019, pp. 417–427.

[22] D. Wang, K. Hu, and X. Han, "Research on the application of intelligent algorithm in short-term traffic flow forecast," in *Proc. Int. Conf. Autom., Mech. Control Comput. Eng.*, Jinan, China, 2015, pp. 277–282, doi: 10.2991/amcce-15.2015.50.

[23] C. Zheng and L. Li, "The improvement of the forecasting model of short-term traffic flow based on wavelet and ARMA," in *Proc. 8th ICSCMIS*, Hong Kong, 2011, pp. 1–4.

[24] C. Chen, J. Hu, Q. Meng, and Y. Zhang, "Short-time traffic flow prediction with ARIMA-GARCH model," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Baden-Baden, Germany, Jun. 2011, pp. 607–612.

[25] J. Cao, G. H. Xu, L. K. Hou, and J. H. Wang, "Detection and estimation for the traffic flow based on Kalman filter," *J. B. Inst. Technol.*, vol. 20, no. 5, pp. 271–275, Dec. 2011.

[26] H. Lin, L. X. Li, and H. Wang, "SPGAPSO-SVM: A prediction algorithm of urban bus passenger flow," *J. Chin. Comput. Sys.*, vol. 41, no. 11, pp. 2458–2464, Nov. 2020.

[27] F. Rui, Z. Zuo, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proc. 31st YAC*, Wuhan, China, 2016, pp. 324–328, doi: 10.1109/YAC.2016.7804912.

[28] S. Peng, A. Boukerche, and Y. J. Tao, "SSGRU: A novel hybrid stacked GRU-based traffic, volome prediction, approach in a road network," *Comput. Commun.*, vol. 160, no. 1, pp. 502–511, Jun. 2020, doi: 10.1016/j.comcom.2020.06.028.

[29] J. Xin and F. H. Xiao, "Financial assets price prediction based on relevance vector machine with genetic algorithm," *J. Inf. Technol.*, vol. 7, no. 5, pp. 90–96, Mar. 2012, doi: 10.4156/jcit.vol7.issue5.12.

[30] Y. F. Li, M. N. Chen, and W. Z. Zhao, "Investigating long-term vehicle speed prediction based on BP-LSTM algorithms," *IET Intell. Transp. Syst.*, vol. 13, no. 8, pp. 1281–1290, Apr. 2019, doi: 10.1049/iet-its.2018.5593.

[31] Z. G. Shen, W. L. Wang, Q. Shen, and Z. C. Li, "Hybrid CSA optimization with seasonal RVR in traffic flow forecasting," *KSII Trans. Internet Inf. Syst.*, vol. 11, no. 10, pp. 4887–4907, Oct. 2017, doi: 10.3837/tiis.2017.10.011.

[32] Y. Zheng, M. Xie, and X. Wang, "Research on passenger flow forecast of Hangzhou metro based on LSTM-SVR," in *Proc. 11st AIAM*, Dublin, Ireland, 2019, pp. 273–276, doi: 10.1109/AIAM48774.2019.00061.

[33] F. Li and S. Ge, "Prediction and analysis of road traffic efficiency based on DBN-SVR," in *Proc. 1st IAI*, Shenyang, China, 2019, pp. 1–6, doi: 10.1109/ICIAI.2019.8850814.

[34] Q. Liu, B. Wang, and Y. Zhu, "Short-term traffic speed forecasting based on attention convolutional neural network for arterials," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 11, pp. 999–1016, 2018.

[35] D. Impedovo, V. Dentamaro, G. Pirlo, and L. Sarcinella, "TrafficWave: Generative deep learning architecture for vehicular traffic flow prediction," *Appl. Sci.*, vol. 9, no. 24, pp. 5504–5517, Dec. 2019, doi: 10.3390/app9245504.

[36] Alibaba. (Feb. 2, 2018). *Introduction to Alicloud Open Source Offline Synchronization Tool Datax3.0*. [Online]. Available: https://github.com/alibaba/DataX/blob/master/introduction.md

[37] J. Guo, Z. Xie, Y. Qin, L. Jia, and Y. Wang, "Short-term abnormal passenger flow prediction based on the fusion of SVR and LSTM," *IEEE Access*, vol. 7, pp. 42946–42955, 2019.doi: 10.1109/ACCESS.2019.2907739.

[38] D. Timothy, "Incorporating nesterov momentum into adam," *Nat Hazards*, vol. 3, no. 2, pp. 437–453, Feb. 2016, doi: 10.13140/RG.2.2.21344.43528.

[39] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014, doi: 10.1.1.669.8604.

[40] S. Shu, "A coordination operation method of wind power and energy storage hybrid system for smoothing short-term and long-term wind power fluctuations," *Automat. Electron. Power Syst.*, vol. 39, no. 8, pp. 12–18, Apr. 2015, doi: 10.7500/AEPS20140816003.

[41] Z. Song, Y. Guo, Y. Wu, and J. Ma, "Short-term traffic speed prediction under different data collection time intervals using a SARIMA-SDGM hybrid prediction model," *PLoS ONE*, vol. 14, no. 6, Jun. 2019, Art. no. e0218626, doi: 10.1371/journal.pone.0218626.

[42] A. Azizi, R. Rooki, and N. Mollayi, "Modeling and prediction of wear rate of grinding media in mineral processing industry using multiple kernel support vector machine," *Social Netw. Appl. Sci.*, vol. 2, no. 9, pp. 1–10, Aug. 2020, doi: 10.1007/s42452-020-03212-0.

[43] Y. Xing, X. Ban, and C. Guo, "Probabilistic forecasting of traffic flow using multikernel based extreme learning machine," *Sci. Program.*, vol. 2017, pp. 1–12, Dec. 2017, doi: 10.1155/2017/2073680.

[44] Y. Yang, S. Mohan, M. Jialiang, and Y. Ping, "Corrosion acoustic emission signal recognition based on relevance vector machine," in *Proc. 13th IEEE Int. Conf. Electron. Meas. Instrum. (ICEMI)*, Yangzhou, China, Oct. 2017, pp. 453–458, doi: 10.1109/ICEMI.2017.8265853.

[45] Y. Liu, B. Liu, X. Zhao, and M. Xie, "Development of RVM-based multiple-output soft sensors with serial and parallel stacking strategies," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 6, pp. 2727–2734, Nov. 2019, doi: 10.1109/TCST.2018.2871934.

[46] M. Gopinath and S. Prabu, "An efficient multiangle weight updated haralick and relevance vector machine algorithm for classifying diabetic foot from medical thermal image," *J. Test Eval.*, vol. 47, no. 6, pp. 4077–4095, Nov. 2019, doi: 10.1520/JTE20180503.

[47] Y. Wang, Z. Xie, I. Lou, W. K. Ung, and K. M. Mok, "Algal Bloom prediction by support vector machine and relevance vector machine with genetic algorithm optimization in freshwater reservoirs," *Eng. Comput.*, vol. 34, no. 2, pp. 664–679, Apr. 2017, doi: 10.1108/ec-11-2015-0356.

[48] B. Wang, "Rolling bearings' intelligent fault diagnosis based on RVM optimized with quantum genetic algorithm," *J. Vib. Shock.*, vol. 34, no. 17, pp. 207–212, Sep. 2015, doi: 10.13465/j.cnki.jvs.2015.17.034.

[49] S. Gopalani and R. Arora, "Comparing apache spark and map reduce with performance analysis using K-Means," *Int. J. Comput. Appl.*, vol. 113, no. 1, pp. 8–11, Mar. 2015.

[50] X. Meng, "MLlib: Machine learning in Apache Spark," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 1235–1241, 2016, doi: 10.5555/2946645.2946679.

[51] Y. Fan, Z. Wei, J. Zhang, N. Zhao, Y. Ren, J. Wan, L. Zhou, Z. Shen, J. Wang, and J. Zhang, "Model aggregation method for data parallelism in distributed real-time machine learning of smart sensing equipment," *IEEE Access*, vol. 7, pp. 172065–172073, 2019, doi: 10.1109/ACCESS.2019.2955547.

[52] L. Venkataramana, S. G. Jacob, and R. Ramadoss, "A parallel multilevel feature selection algorithm for improved cancer classification," *J. Parallel Distr. Comput.*, vol. 138, pp. 78–98. Apr. 2020, doi: 10.1016/j.jpdc.2019.12.015.

[53] H. H. Mei, "Datav: Data visualization on large high-resolution displays," *Vis. Inf.*, vol. 4, no. 3, pp. 12–23, Sep. 2020, doi: 10.1016/j.visinf.2020.07.001.

[54] *Introduction to Open Data of Canada's Whitemud Drive Highway*. Accessed: Apr. 29, 2020. [Online]. Available: http://www.openits.cn/openData1/700.jhtml

[55] X. W. Chen, "Research on short-term traffic flow forecasting model based on LSTM," *World Sci. Res. J.*, vol. 6, no. 5, pp. 191–200. May 2020, doi: 10.6911/WSRJ.202005_6(5).0021.

[56] S. D. Du, "A hybrid method for traffic flow forecasting using multimodal deep learning," *Int J. Comput. Int. Syst.*, vol. 13, no. 2, pp. 85–97, Mar. 2020, doi: 10.2991/ijcis.d.200120.001.

[57] Y. Liu, H. Zheng, X. Feng, and Z. Chen, "Short-term traffic flow prediction with Conv-LSTM," in *Proc. 9th WCSP*, Nanjing, China, 2017, pp. 1–6, doi: 10.1109/WCSP.2017.8171119.

[58] W. Feng, H. Chen, and Z. Zhang, "Short-term traffic flow prediction based on wavelet function and extreme learning machine," in *Proc. 8th ICSESS*, Beijing, China, 2017, pp. 531–535, doi: 10.1109/ICSESS.2017.8342971.

[59] J. Pearl and D. Mackenzie, *The Book of Why: The New Science of Cause and Effect*. Basic Books, 2018.

HAO LIN was born in Tianjin, China, in 1995. He received the B.A. degree in engineering from the Tianjin University of Technology and Education, in 2018. He is currently pursuing the M.A. degree in engineering with the Inner Mongolia University of Technology. His research interests include cloud computing, data mining, big data processing, and cyberspace security.

JIANXIONG WAN received the B.Sc. degree in computer science from Shanxi Normal University, Xi'an, China, in 2004, the M.Sc. degree in management science from the Beijing Information Technology Institute, Beijing, China, in 2009, and the Ph.D. degree in computer science from the University of Science and Technology Beijing, China, in 2013. He was a Postdoctoral Research Fellow with Massey University, Palmerston North, New Zealand, from 2016 to 2017, and also a Visiting Researcher with the Nara Institute of Science and Technology, Japan, from 2017 to 2018. He joined the Inner Mongolia University of Technology in 2013, where he is currently a Professor with the School of Data Science and Application. His research interests include dynamic optimization, intelligent control, and performance modeling of distributed systems, with special focuses on cloud scale systems.

ZHIQIANG MA received the B.E. degree in computer application technology from Hohai University, Nanjing, China, in 1995, and the M.E. degree in computer science and technology from the Beijing Information Science and Technology University, Beijing, China, in 2007. He joined the Inner Mongolia University of Technology, in 1995, where he is currently a Professor with the College of Data Science and Application. His research interests include speech recognition and natural language processing, with a special focus on emotional text generation.
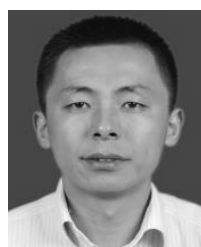
LEIXIAO LI was born in Chengwu, China, in 1978. He received the M.A. degree in engineering from the Inner Mongolia University of Technology, in 2007, and the Ph.D. degree in engineering from Inner Mongolia Agricultural University, in 2019. His research interests include cloud computing, data mining, and big data processing.

HUI WANG was born in 1980. She received the M.S. degree in engineering from Harbin Engineering University, in 2006. She is currently a Lecturer with the College of Data Science and Application, Inner Mongolia University of Technology. Her research interests include cloud computing, data mining, and big data processing.

• • •