Research Article

A Deep Cycle Limit Learning Machine Method for Urban Expressway Traffic Incident Detection

YunFeng Fang [iD],[1,2] **Qingfang Yang** [iD],[1,2] **Lili Zheng** [iD],[1,2] **Xiangyu Zhou** [iD],[1,2] **and Bo Peng** [iD][1,2]

[1]*Department of Traffic Information and Control Engineering, Jinlin University, Changchun, China*
[2]*Jilin Engineering Research Center for ITS, Changchun 130022, China*

Correspondence should be addressed to Lili Zheng; lilizheng@jlu.edu.cn

In Beijing, Shanghai, Hangzhou, and other cities in China, traffic congestion caused by traffic incidents also accounts for 50% to 75% of the total traffic congestion on expressways. Therefore, it is of great significance to study an accurate and timely automatic traffic incident detection algorithm for ensuring the operation efficiency of expressways and improving the level of road safety. At present, many effective automatic event detection algorithms have been proposed, but the existing algorithms usually take the original traffic flow parameters as input variables, ignoring the construction of feature variable sets and the screening of important feature variables. This paper presents an automatic event detection algorithm based on deep cycle limit learning machine. The traffic flow, speed, and occupancy of downstream urban expressway are extracted as input values of the deep-loop neural network. The initial connection weights and output thresholds of the deep-loop neural network are optimized by using the improved particle swarm optimization (PSO) algorithm for global search. The higher classification accuracy of the extreme learning machine is trained, and the generalization performance of the extreme learning machine is improved. In addition, the extreme learning machine is used as a learning unit for unsupervised learning layer by layer. Finally, the microwave detector data of Tangqiao viaduct in Hangzhou are used to verify the experiment and compared with LSTM, CNN, gradient-enhanced regression tree, SVM, BPNN, and other methods. The results show that the algorithm can transfer low-level features layer by layer to form a more complete feature representation, retaining more original input information. It can save expensive computing resources and reduce the complexity of the model. Moreover, the detection accuracy of the algorithm is high, the detection rate is higher than 98%, and the false alarm rate is lower than 3%. It is better than LSTM, CNN, gradient-enhanced regression tree, and other algorithms. It is suitable for urban expressway traffic incident detection.

1. Introduction

China's road traffic situation is extremely grim. With the rapid development of urban road traffic, the numbers of car ownerships and motor vehicle drivers show a trend of rapid growth, which led road traffic incidents to have maintained a high base and high accident rate for many years. More than 50% of urban road traffic congestion in Shanghai, China, is caused by expressway traffic incidents. The congestion caused by traffic accidents on expressways will reduce the capacity of expressways, affect the operational efficiency of expressways, and seriously cause traffic accidents, threatening people's lives and property safety. Therefore, it is necessary to study the detection algorithm of expressway

traffic incidents, improve the management level of expressway traffic incidents, detect and deal with traffic incidents in time, and reduce the impact of traffic incidents on urban expressway traffic.

In the aspect of traffic incident detection, many experts and scholars have done some research studies, and some achievements have been applied. In [1], Texas Transportation Institute (TTI) has developed a standard normal deviation algorithm. This algorithm establishes a standard deviation value. If traffic volume is detected to exceed this value, the system draws a conclusion that traffic incidents occur. The standard normal deviation algorithm is applied to Houston Bay Highway. In [2], based on Thom's catastrophe theory, the Department of Civil Engineering, McMaster

University, Canada, developed the McMaster algorithm. The basis of this algorithm is that when the traffic condition changes from congestion state to noncongestion state, the change of occupancy and flow rate is not obvious, but the change of speed is obvious. In [3], Hsiao et al. put forward the scheme of using fuzzy logic to solve traffic incident detection. Fuzzy logic method cannot accurately determine whether there are traffic incidents or not but can only give the probability of traffic incidents. In [4], Abdulhai and Ritchie introduced probabilistic neural network into traffic incident detection and carried out simulation analysis with real data (I-880 and I-35W databases). The results show that the model can significantly improve the detection rate and meet the performance requirements of traffic incident detection. In [5], Yuan and Cheu tried to use support vector machines (SVM) to detect traffic incidents. Three SVM models were developed and tested with I-880 traffic database. The results show that the detection performance is as good as MLF. In the context of large data and intelligent transportation, the above algorithms are generally cumbersome in computation, poor in generalization, low in detection efficiency, low in model accuracy, and not suitable for large sample data.

Extreme learning machine (ELM) is a new learning algorithm for single-layer feedforward neural networks. ELM randomly selects input weights without adjusting them and calculates output weights through Moore-Penrose generalized inverse matrix. ELM has faster training speed and stronger generalization ability and avoids falling into local minima. In order to reduce the time consumed in determining the number of hidden neurons and randomly assigning weights between the extreme learning machines proposed by Huang et al. [6], performance and learning speed can be increased by hundreds of single-hidden-layer feedforward neural networks. The parameters of hidden layer nodes of extreme learning machine are generated randomly according to the continuous probability distribution, and the parameters of hidden layer nodes are independent of training samples. They do not need to be adjusted iteratively in the training process. They have the characteristics of fast training speed and strong generalization ability of the model. Therefore, compared with the traditional neural network method, ELM shows obvious advantages in classification problems and can maintain good learning [7] performance; learning speed can be increased by hundreds or even thousands of times, which is conducive to solving traffic incident detection problems in the context of large data.

Deep Learning, or deep neural network, is the latest research direction in the field of machine learning. It can deeply mine the distribution characteristics of large traffic data and apply them to traffic incident detection, which can greatly improve the accuracy of incident detection. However, with the increase of network layers, their training efficiency is greatly reduced, and the probability of model falling into local optimum is increased. Regarding the difficulty of training, in [8], Springenberg adopts the heterogeneous neural network model to replace the convolution layer and the pool layer in the traditional convolution neural

network model with the cyclic convolution layer of step 2. The batch random gradient descent algorithm is used to train the current complex convolution neural network model to achieve the performance of classification and detection, but the difficulty of training increases. In [9], Zhang et al. proposed a framework of relationship extraction based on RNN. The model of long-distance relationship is built by bidirectional RNN. Experiments on two data sets show that the model based on bidirectional RNN is better than that based on CNN, but the training efficiency is lower.

To solve the above problems, Huang Guangbin, the founder of extreme learning machine, proposed a deep multilayer extreme learning machine algorithm. The multilayer neural network structure enables extracting high-level abstract information from data. At the same time, it can effectively solve the problems of high data dimension, difficult sample labeling, difficult feature construction, and difficult training in the era of big data. The literature in [10] introduces the combination of extreme learning machine and self-encoder for the first time. It is believed that the feature expression ability of ELM-AE can provide a good solution for multilayer feedforward neural network. Moreover, compared with the most advanced deep network, the multilayer network based on ELM can provide better performance.

The document in [11] considers that deep architecture can obtain higher-level feature representation and thus obtain higher-level abstract information. Therefore, a multilayer extreme learning machine model is proposed, which learns the deep representation of data through extreme learning machine according to stack generalization theory. Firstly, this paper will use traffic flow parameters and their combination to construct an initial variable set for traffic incident detection. The importance measure of random forest variables is used to select the characteristic variables for traffic incident detection, and the deep cycle limit learning machine model is used to train the characteristic variables. Finally, the performance of the model is analyzed by the real data of expressway.

2. Data Description and Variable Selection

The acquisition of traffic flow data and traffic incident data is of great significance to the study of traffic incident detection. The traffic flow data obtained in this paper mainly comes from the microwave detector data collected by Hangzhou urban expressway monitoring center in Hangzhou viaduct's section for 5 months (from June 11, 2015, to November 11, 2015). Among them, the sampling [12] interval of microwave traffic detection data is 1 min, and the collected data content is serial number, fixed detector number, date, time, flow, speed, and time occupancy. The traffic incident information source is the expressway incident information released by Hangzhou traffic information network. The study recorded 223 pieces of main traffic incident information of Hangzhou Expressway from June 11 to November 11, 2015, including 107 pieces of effective traffic incident information. Traffic incident information includes serial number, date, time, longitude, and latitude. When analyzing and researching the

event detection algorithm, we need to match the event information with the fixed detector data in space. First, we need to locate the location of the event on the map of Hangzhou and then find the corresponding upstream and downstream fixed detector number and select the corresponding fixed detector data according to the location and time period of the event. Among them, event occurrence is marked as 1 and no event occurrence is marked as 0. Some data formats after preprocessing are shown in Table 1.

The basis of traffic incident detection is the disturbance of normal traffic flow caused by traffic incident. Therefore, before constructing the incident detection algorithm, we must first analyze the traffic flow characteristics in the event state to determine the characteristic parameters of the model. Based on the theory of vehicle flow fluctuation, this paper analyzes the impact of traffic events on the [13] characteristics of traffic flow, based on which the event detection feature variable set is constructed, and the important feature parameters are selected.

In order to analyze the traffic flow characteristics in the event state, 50 groups of 107 groups of traffic event data are randomly selected for cross validation to analyze the impact of traffic events on the traffic flow characteristics, so as to eliminate the interference of time of the day, detector location, event type and location, downstream signal of the off-ramp, and other factors.

Through a large amount of cross validation detector data and event data, it can be found that the impact of traffic events on traffic flow will directly lead to the change of traffic flow parameters (such as flow, speed, density, and occupancy) at the event points and upstream and downstream sections. Therefore, the significant change of traffic flow parameters during the event occurrence period is the basic basis for the design of automatic traffic event detection algorithm. Among them, the change trend of upstream and downstream traffic flow parameters of the event location is shown in Figures 1 and 2, respectively. In Figures 1 and 2, Q is the traffic flow, O is the occupancy rate, and v is the speed.

In Figure 1, when a traffic accident occurs in a section, the flow and speed acquired by the upstream detector at the location of the traffic accident decrease sharply, and the occupancy rate increases sharply.

In Figure 2, when a traffic accident occurs in a section, the flow acquired by the downstream detector decreases, the speed increases, and the occupancy rate decreases. Therefore, the combination of different traffic parameters and the combination of upper and lower detectors also show strong sensitivity to the occurrence of traffic incidents.

In this paper, a complete set of initial variables is constructed based on the measured, predicted, and combined values of the traffic flow parameters in the upstream and downstream areas. The set of initial variables consists of seven parts: (1) the basic traffic parameters actually obtained by the upstream detector; (2) the basic traffic parameters actually obtained by the downstream detector; (3) the combination ratio of the actual traffic parameters of the upstream detector; (4) the group of the actual traffic parameters of the downstream detector; (5) the ratio of the measured traffic flow parameters and the predicted parameters of the upstream detector; (6) the ratio of the measured traffic flow parameters and the predicted parameters of the downstream detector; (7) the ratio of the measured traffic flow parameters and the predicted parameters of the downstream detector. The ratio of the measured traffic flow parameters of the adjacent detector is shown in Table 2. The predicted values of the traffic flow parameters are obtained by moving average method, and the fifth data is predicted by using the first four adjacent data.

There are 18 initial variable sets in Table 2, which comprehensively cover event characteristics. But, in practical application, there are too many variables and the information needed to be processed is redundant. It increases the difficulty of modeling. Therefore, feature variables need to be screened to reduce the complexity of model construction. Random forest is an effective way to reduce data dimension [14] and improve the accuracy of data classification. It is widely used to measure the importance of variables. It is also suitable for solving the problem of screening important variables.

The Bootstrap random sampling technique and the node random splitting technique are used to extract the new sample set from the training set and establish the decision number model. When random forests were sampled by Bootstrap, about 36.8% of the "out of bag data" were generated at a time (Out of Bag, OOB). Using OOB as a test set to evaluate the predictive performance of RF is called OOB estimation. OOB estimation is unbiased when the number of trees in RF is large enough.

For random forests that have been generated, we assume that the total number of OOB samples is N_{OOB}. When OOB is used as a test set to evaluate the performance of FE classification for stochastic forests, the correct number of samples tested is k_{OOB}. Then the formula of classification accuracy Acc_{OOB} is as follows:

$$\text{Acc}_{\text{OOB}} = \frac{k_{\text{OOB}}}{N_{\text{OOB}}}. \tag{1}$$

Feature importance measurement is an important feature of RF and can be used as a feature selection tool for high-dimensional data. Mean Decrease in Accuracy (MDA) is an important index to measure the importance of feature. Suppose the Bootstrap sample is $B_1, B_2, \ldots, B_n$ (n is the number of training samples), and the feature is $X_1, X_2, \ldots, X_m$ (m is the feature dimension). The steps of calculating the feature importance measure are as follows:

Step 1. Set $i = 1$, create a decision tree T_i on the training sample, and mark the out-of-pocket data as OOB_i.

Step 2. Choose OOB_i as the test set, apply T_i to classify it, and mark the correct number of predictions as $\text{Acc}_{\text{OOB}_i}$.

Step 3. For each feature X in OOB_i, add artificial noise to the data set and record it as $\text{OOB}_i^{(1)}$. Apply T_i to classify it, and mark the correct number of predictions as $\text{Acc}_{\text{OOB}_i^{(1)}}$.

Step 4. For $i = 2, 3, \ldots, n$, repeat steps 1 to 3.

TABLE 1: Some data used for traffic incident detection.

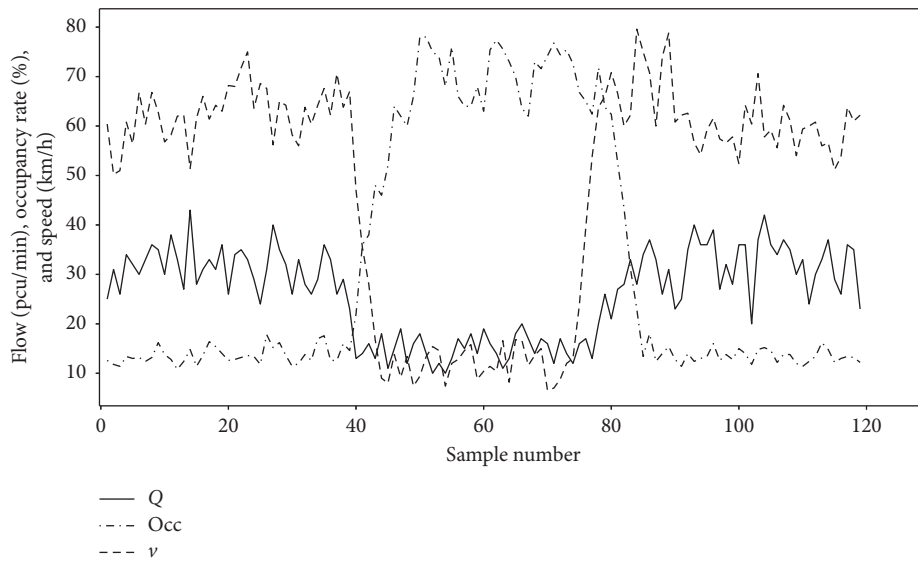|	Microwave detector data						
Time	Upstream section			Downstream section			Label
	Flow (pcu/min)	Speed (km/h)	Occupancy (%)	Flow (pcu/min)	Speed (km/h)	Occupancy (%)	
10:03	40	68.43	24.51	42	67.65	23.7	0
10:04	44	69.54	19.8	48	72.03	20.6	0
10:05	38	76.75	23.6	41	74.12	21.2	0
10:06	25	24.12	75	56	68.24	12.6	1
10:07	22	16.45	72.4	45	73.45	11.3	1
...	...	...	...	...	...	...	...

FIGURE 1: Change in traffic flow parameters for upstream detectors.
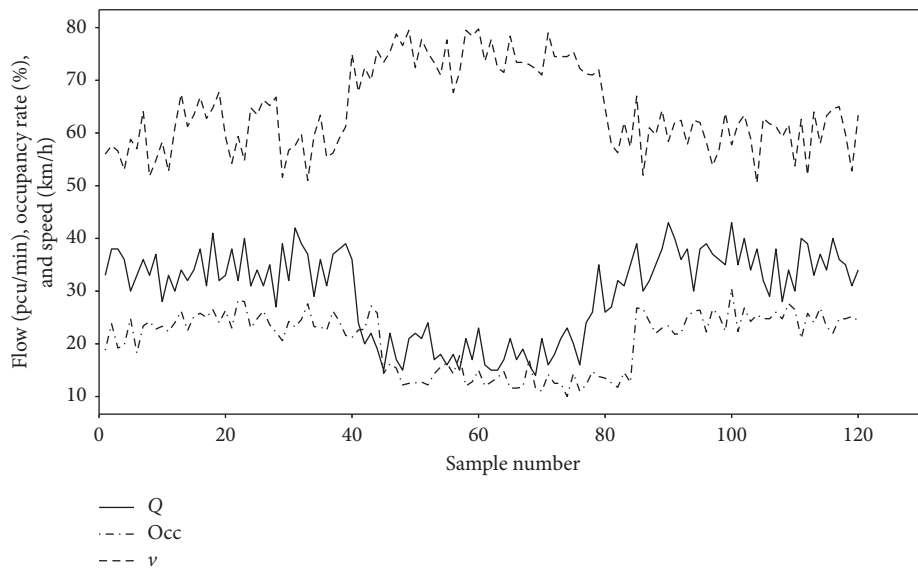

FIGURE 2: Change of traffic flow parameters for downstream detector.

TABLE 2: Initial variables set of traffic incident detection.

Variable serial number	Initial variable
1	Measured flow rate of upstream detector
2	Measured speed of upstream detector
3	Measured occupancy of upstream detector
4	Measured flow rate of downstream detector
5	Measured speed of downstream detector
6	Measured occupancy of downstream detectors
7	Ratio of occupancy to flow measured at the same time by the same detector
8	Ratio of occupancy to velocity measured at the same time by the same detector
9	Ratio of flow to velocity measured at the same time by the same detector
10	Ratio of measured flow and predicted flow of upstream detector
11	Ratio of measured speed to predicted speed of upstream detector
12	Ratio of measured occupancy rate to predicted occupancy rate of upstream detector
13	Ratio of measured flow to predicted flow of downstream detector
14	Ratio of measured speed to predicted speed of downstream detector
15	Ratio of measured occupancy rate to predicted occupancy rate of downstream detector
16	Flow ratio of adjacent upstream and downstream detectors at the same time
17	Speed ratio of adjacent upstream and downstream detectors at the same time
18	Acquisition occupancy ratio of adjacent upstream and downstream detectors at the same time

Step 5. In order to measure importance of feature X_j, it can be calculated by the following formula:

$$\overline{\mathrm{MDA}}_j = \frac{1}{n} \sum_{i=1}^{n} \left(\mathrm{Acc}_{\mathrm{OOB}_i} - \mathrm{Acc}_{\mathrm{OOB}_i^{(1)}} \right). \qquad (2)$$

3. Method of the Deep Cycle Limit Learning Machine

Cyclic convolution neural network consists of input layer, cyclic convolution layer, and output layer. It can extract classification features layer by layer through cyclic convolution layer and sampling layer. The last layer is softmax nonlinear classifier. The weights and biases of cyclic convolution neural network are trained by Newton algorithm, crossing entropy cost function as the object function of training cyclic convolution neural network and iteratively searching for the minimum optimal solution of the object function of cyclic convolution neural network. The feature extraction is a part of the whole classifier design. Therefore, it is advantageous to extract features from cyclic convolution neural network as the input of limit learning machine. The extreme learning machine is a single hidden layer neural network. Its input layer weights and offset values are randomly generated during the initialization process. Then, the weight biases from the hidden layer to the output [15] layer are calculated by using the generalized inverse method based on the relationship between the target output and the input. Because the weight offset of the extreme learning machine is computed without training, the training speed of the extreme learning machine is faster. In the process of extracting image target features, the circular convolution neural network extracts image target features better, so the classification effect of the extreme learning machine of the circular convolution neural network is good.

In Figure 3, this paper combines the cyclic convolution neural network with the extreme learning machine and constructs the extreme learning machine based on the cyclic convolution neural network to classify the image target, which takes full advantage of the characteristics of the cyclic convolution neural network to extract the abstract and salient features of the image target and the fast calculation of the extreme learning machine. In Figure 4, the limit learning machine model based on cyclic convolution neural network consists of three parts:

(1) Cyclic convolution neural network is composed of input layer, cyclic convolution layer, and output layer of pooling layer. Quasi-Newton method is used to train cyclic convolution neural network to realize feature extraction of image target.

(2) Cyclic convolution neural network is used to extract the features of image objects as the input layer of the extreme learning machine to calculate the parameters of the extreme learning machine.

(3) Use extreme learning machine to classify image objects.

The combined algorithm of convolution neural network and extreme learning machine is as follows: input training sample data, class label information of target, and output training model of cyclic convolution neural network.

Because the initial goal of extreme learning machine is to solve single-hidden-layer feedforward neural network, assuming that a hidden single-layer neural network has N samples, a single-hidden-layer neural network with L hidden nodes can be expressed as

$$\sum_{i=1}^{L} \beta_i g\left(w_i \cdot x_j + b_i \right) = o_j, \quad j = 1, \ldots, N. \qquad (3)$$

In formula (3), $g(x)$ is the activation function; $w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ is the input weight variable; β_i is the output weight; b_i is the bias of the hidden layer neuron.

Then the objective function of the minimum learning output error of the single-hidden-layer neural network can be expressed as
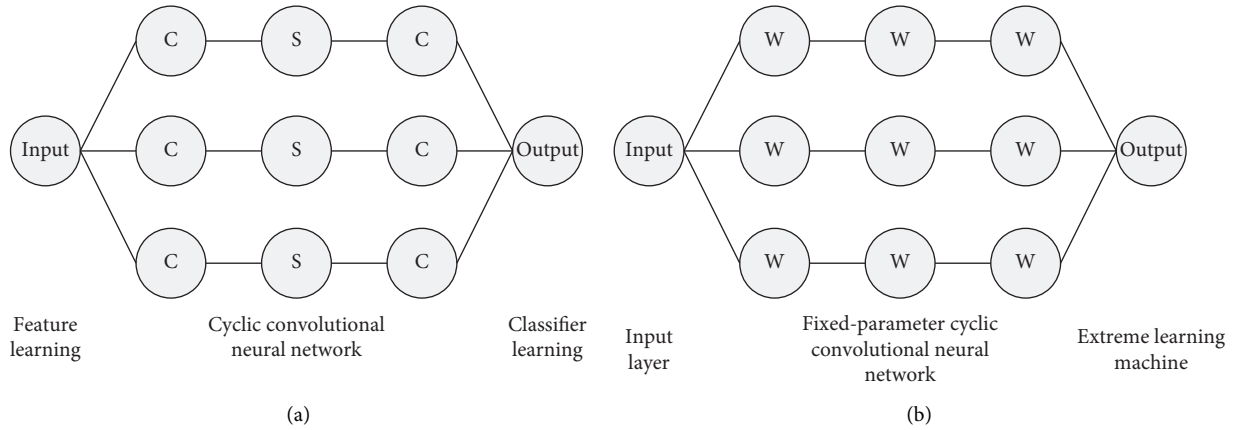
(a)

(b)

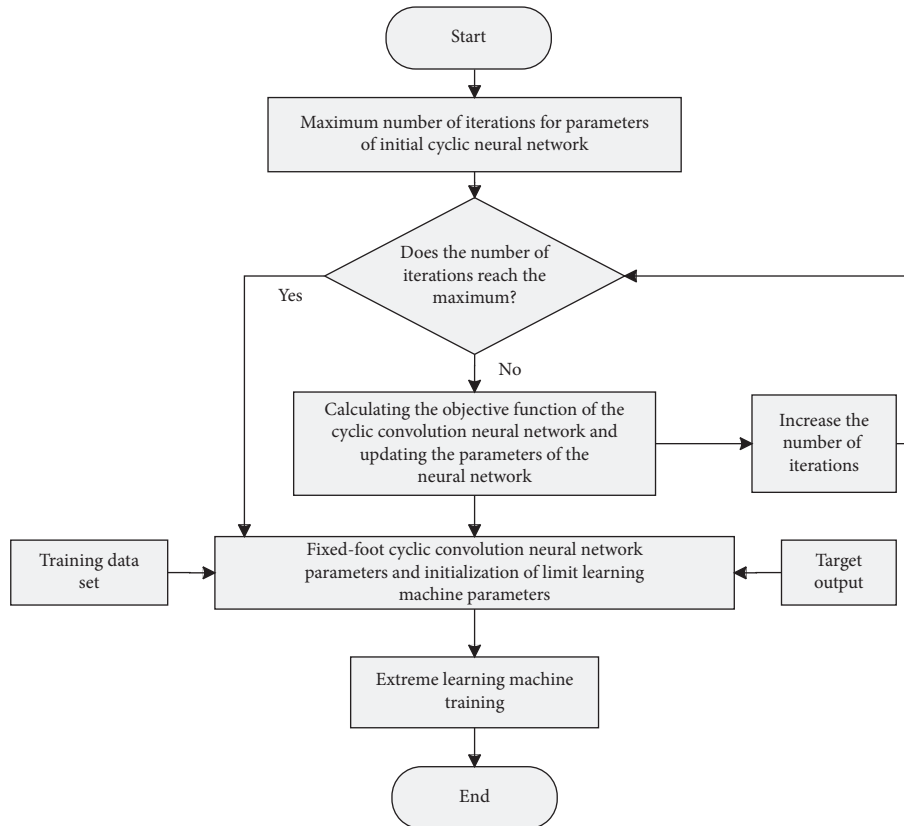FIGURE 3: Deep cycle extreme learning machine training.

FIGURE 4: Deep cycle extreme learning machine training.

$$\sum_{j=1}^{N} \left\| O_j - t_j \right\| = 0. \tag{4}$$

Therefore, the objective function of the extreme learning machine can be expressed as

$$H\beta = T, \tag{5}$$

where H is the output of hidden nodes, β is the output weight, and T is the expected output.

In the training process, the gradient of Newton method near saddle point decreases slowly and easily falls into local optimum, which makes it difficult to train the model. Newton's method needs to calculate the black matrix of the objective function (Hessian) in the calculation process, and it cannot guarantee that the black matrix of the objective function is always positive definite.

The second-order partial derivative of the objective function needs to be calculated, which is too large and difficult to store, so that the direction of the algorithm does

not always descend. Thus, the Newton method fails. In order to solve this problem, an improved BFGS algorithm based on quasi-Newton method is adopted.

Newton's algorithm updates the parameters of the deep-loop neural network model in the optimization process as follows:

$$w_{i+1} = w_i - \frac{\nabla j_w}{\nabla^2 j_w}. \tag{6}$$

In order to overcome the shortcomings of Newton's method, quasi-Newton equation is adopted, and the initial value of second derivative is replaced by b approximation:

$$b_{k+1} s_k = y_k. \tag{7}$$

In order to better optimize the nonconvex objective function, the improved Newton algorithm is adopted as follows:

$$b_{k+1} = b_k - \frac{b_k s_k s_k^T b_k}{s_k^T b_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}. \tag{8}$$

Generally speaking, the weights w and offset b of the hidden neurons in the extreme learning machine are set by a single-hidden-layer neural network using a random method. In this paper, the training set of the cyclic convolution neural network is used to learn the features of the target as the input of the single-hidden-layer limit learning machine, and the weights of the neurons of the single-hidden-layer neural network are set by calculating the expected output of the target.

4. Case Study

4.1. Experimental Environment. In order to verify the performance test experiment of the proposed expressway traffic event detection algorithm, the experimental environment is shown in Table 3.

The database includes 107 traffic events (10294 samples in total). 55 traffic event data are randomly selected for training and the remaining 52 traffic event data are used for testing. Because the amount of nonevent data is too large, nonevent samples are usually randomly selected to build training set and test set. In order to retain the information of nonevent samples to a large extent, the training set and test set are used. The proportion of event samples in test set is set to 20%. The composition of training set and test set is shown in Table 4.

In practice, the number of traffic event samples is far less than the number of nonevent samples, and the number of two types of samples is unbalanced. Therefore, traffic incident detection can be regarded as a two-classification problem of unbalanced data. Synthetic minority sampling technique (SMOTE) is a commonly used oversampling technique. SMOTE can generate new samples that do not exist in the original sample. Therefore, to a certain extent, it avoids the hyperfitting of the classification algorithm. The standard SMOTE is used to balance the traffic incident detection sample. The specific steps are as follows:

TABLE 3: Experimental environment configuration.

Project	Environmental values
Server	Dell PowerEdge T640
Processor	Intel® Xeon® Bronze 3106 CPU @ 1.70 GHz
Memory	128 GB
Video card	4 ∗ GeForce RTX 2080 Ti
Operating system	Ubuntu 16.04
Development environment	Python 3.6, TensorFlow, Keras

(1) For each sample in the event sample set x_i, use the European distance as a measure to search for K samples in the event sample set closest to the sample.

(2) The sampling rate N is determined based on the ratio of the number of nonevent samples to the number of event samples, and N samples are randomly selected from the K nearest neighbor samples of each event sample, denoted as x_{ij}.

(3) According to formula (1), random linear interpolation between randomly selected nearest neighbor samples x_{ij} and event samples x_i is used to construct a new event sample:

$$x^{new} = x_i + \text{rand}(0, 1) \times x_{ij} - x_i. \tag{9}$$

In the equation, rand(0, 1) represents a random number belonging to a large interval of 0,1.

(4) Merge the newly generated event sample with the original sample set to obtain a relatively balanced training sample set.

In order to make the two types of samples relatively balanced, SMOTE is used to increase the sample of traffic events in the training set. The specific parameters of SMOTE are set as follows: the number of adjacent sample points is 5, and the oversampling rate is 30,000. The number of samples in the balanced training set is 20360. In order to eliminate the effects of different dimensions, improve the training speed and classification effect, and normalize the data to the interval [0, 1], the normalization formula is

$$y_i = \frac{x_i - x_{min}}{x_{max} - x_{min}}, \tag{10}$$

In the formula, x_i is the original data, y_i is the normalized data, x_{max} is the maximum value of the original data, and x_{min} is the minimum value of the original data.

4.2. Data Set Partition and Variable Screening. Three basic traffic flow parameters, namely, traffic flow, speed, and occupancy, can be obtained by using remote microwave detector. The sampling interval of data is 5 minutes. By analyzing the changing trend of traffic flow parameters, 123 main road traffic incidents were screened out artificially, of which 71 were on the east side and 52 were on the West side. Traffic incident data are classified according to the east main line, the west main line, and the whole road section. Three

TABLE 4: Composition of training set and test set sample.

Data set categories	Total sample size	Event sample number	Event number
Training set	26190	5240	55
Test set	22978	4595	52

sample data sets are formed with corresponding normal state data. Two-thirds of each data set are used as training samples, and the rest are used as test samples.

Random forest algorithm is used to measure the importance of initial variables, and then key variables that are more sensitive to traffic incidents are selected. Among them, m_{try} takes the square root of the number of characteristic variables recommended by Breiman and sets it to 3. The number of CART in random forest is set to 1000. The importance of random forest variables is calculated by using the program of Python. Eighteen initial variables are normalized and input into the program of random forest. Figure 5 shows the importance of each initial variable.

In order to reflect the role of key variables screening, we should not only select as few variables as possible but also ensure the correct rate of traffic incident detection. Through comparative analysis, the four [16] variables with the highest importance were selected as the key variables. Figure 1 shows that the four most important variables are the ratio of occupancy to speed of the same detector, the ratio of occupancy of adjacent upstream and downstream detectors at the same time, the ratio of occupancy to flow of the same detector, and the ratio of speed of adjacent upstream and downstream detectors at the same time.

4.3. Particle Swarm Optimization.

Particle swarm optimization (PSO) is used to obtain the optimal parameters of the combined kernel function. For general problems, the set range of particles is 20–50. For specific problems, the number of particles can be taken to 100–200. The larger the spatial range of search is, the easier it is to find the global optimal solution and of course the longer the algorithm runs. Considering that the number of particles in the event detection problem is 20, it can solve the problem of traffic incident detection and improve the training efficiency of the algorithm. The specific parameters of PSO are as follows: the number of particles is 20, the dimension of particles is 3, and the acceleration factor $c_2 = c_2 = 2$. The inertia weight coefficient decreases linearly from 0.9 to 0.4 with the number of iterations, and the maximum number of iterations is 100. The average detection rate of traffic events verified by 5-fold cross validation is used as the fitness function value. Take the sample data set of the east main line as an example to optimize the parameters. Figure 6 shows the applicability curve of PSO optimization.

As can be seen from Figure 6, the optimum parameters of the combined kernel function of the east main line sample data set are as follows: $\lambda = 0.53$, $\sigma = 0.25$, and $d = 3$.

4.4. Result Analysis.

In order to better evaluate the detection performance of the deep cycle extreme learning machine algorithm (DELM) established in this paper, long short term memory (LSTM) algorithm, deep belief network (DBN) algorithm, convolutional neural network (CNN) algorithm, and gradient boosting decision tree (GBDT) algorithm are selected for comparison.

The performance evaluation indexes of traffic incident detection algorithm include detection rate (DR), false alarm rate (FAR), and mean time to detection (MTTD). DR represents the percentage of the number of events detected in a given period of time to the actual number of events. FAR represents the percentage of false alarm events in all decision-making times over a given period of time. MTTD denotes the arithmetic mean of the difference between the detected event occurrence time and the actual event occurrence time. The calculation formulas of the three evaluation indexes are as follows:

$$DR = \frac{DN}{AN} \times 100\%,$$

$$FAR = \frac{FN}{RN} \times 100\%, \quad (11)$$

$$MTTD = \frac{1}{n} \sum_{i=1}^{n} [TI(i) - AT(i)].$$

In the above formula, DR is traffic incident detection rate; DN is the number of traffic events detected; AN is the total number of traffic events in the corresponding time; FAR is false alarm rate; FN is number of misreported traffic events in corresponding time; RN is all decision times in corresponding time; MTTD is average detection time; $TI(i)$ is the time of the ith traffic incident detected by the algorithm; $AT(i)$ is the actual time of the ith traffic incident.

DR and FAR are used to measure the detection performance of automatic traffic incident detection algorithms. MTTD is used to describe the detection efficiency of automatic traffic incident detection algorithms. The three indicators of DR, FAR, and MTTD are related to each other and depend on each other. If a higher DR is obtained, then it must lead to a higher FAR; meanwhile, reducing FAR will inevitably reduce DR. Similarly, if a shorter MTTD is obtained, the FAR is relatively higher. When evaluating a new automatic traffic incident detection algorithm, there is usually a trade-off in these 3 indicators.

Generally speaking, the three indicators of DR, FAR, and MTTD are affected by the determination of the start time of the traffic event. The occurrence time of the traffic event detected by the algorithm refers to the time corresponding to the abnormality of traffic flow data caused by the impact of the traffic event, and the end time of the traffic event impact detected by the algorithm refers to the time when the traffic flow is not affected by the traffic event and the traffic flow data returns to normal. Because this paper is an offline test and the starting time of traffic incidents is not like using simulation software to set the time of occurrence to directly
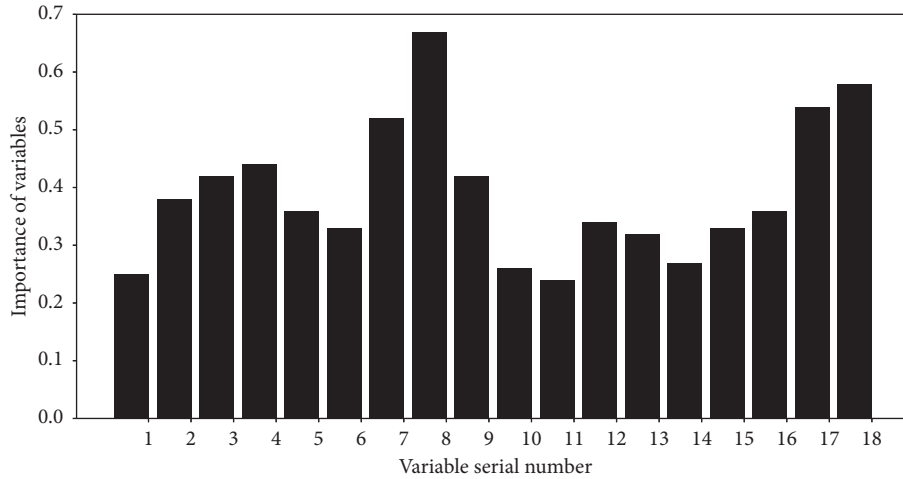
FIGURE 5: Importance of variable.
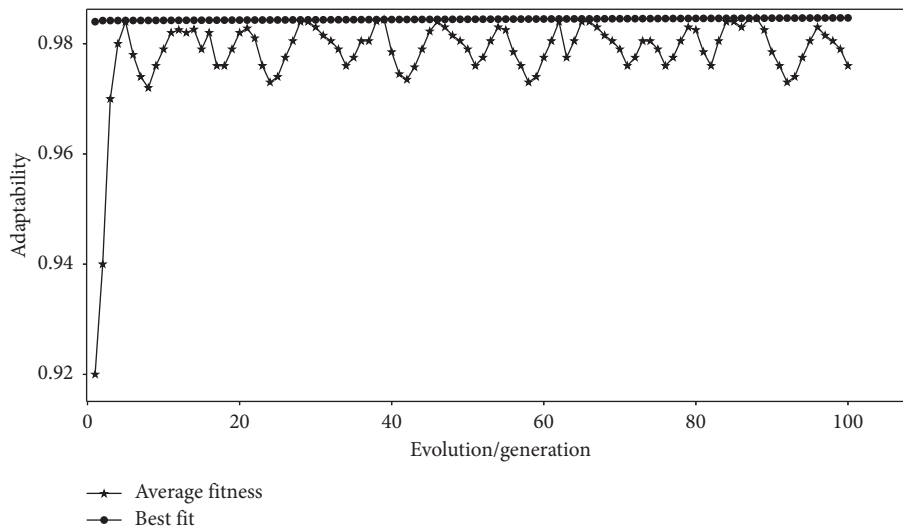

Average fitness
Best fit

FIGURE 6: Fitness curves of PSO optimization.

extract data from the database, its determination method is based on data threshold judgment. In a certain period, the abnormal moment of the traffic flow data is the first time when the traffic event occurs. After a period of time, when the traffic flow data returns to normal, the moment when there is no abnormality is the time when the impact of the traffic event ends. However, in this article, the start time of the traffic incident has a limited impact on DR, FAR, and MTTD. In the offline test, the detection time is related to the running speed of the computer and the speed of connecting to the server, and this test is on the same hardware platform, and the calculation time is not much different; while the algorithm is running online, the detection time is mainly affected by the [17] network. The transmission data speed is affected; its detection rate and false alarm rate are mainly affected by the algorithm classification performance; the higher the algorithm classification accuracy, the more correctly detected events and the fewer false alarm traffic events. The start time of the traffic incident mainly affects

the duration of the traffic incident. Therefore, the start time of the traffic incident in this article is a secondary indicator.

The initial variables and important variables were used to construct the training set to test the performance of these five algorithms. The results are shown in Table 5. It can be seen that, compared with the automatic traffic event detection algorithm using initial variables, the automatic traffic event detection algorithm using important variables has been improved in DR, FAR, and MTTD, with higher DR, lower FAR, and shorter MTTD. Among them, the DR of DELM event detection algorithm using important variables is improved by 3.99%, 2.08%, 2.27%, and 2.72% compared to the DR of LSTM, DBN, CNN, and GBDT; FAR difference is small; MTTD has been reduced by 24.44%, 22.72%, 27.66%, and 26.09%, respectively. So the DELM algorithm using important variables has better performance than the LSTM, DBN, CNN, and GBDT algorithms and can effectively solve the problem of expressway traffic incident detection.

TABLE 5: Comparison of detection effects of different detection algorithms.

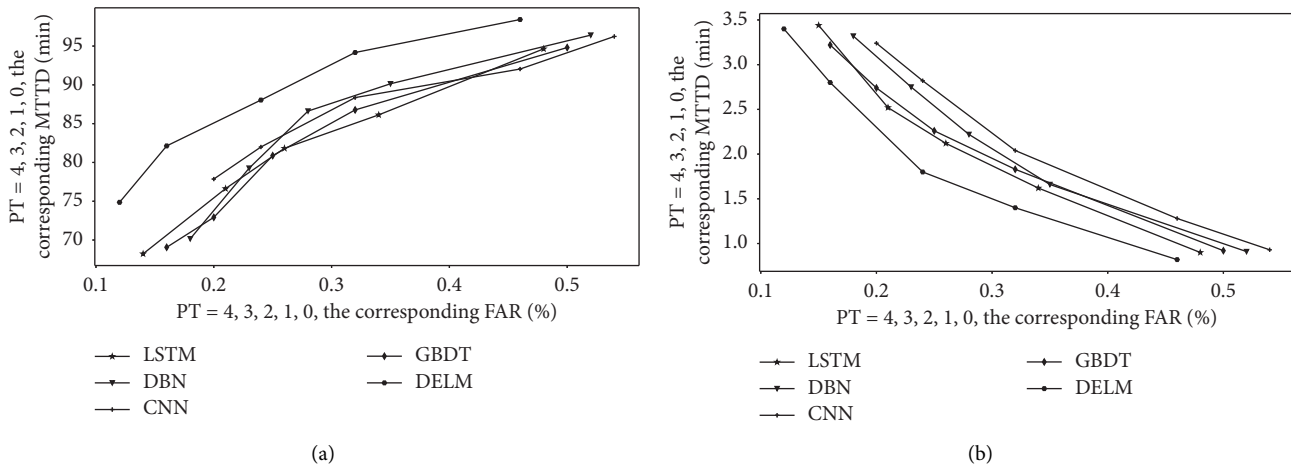| Detection algorithm	Variable selection			Initial variable		
	DR (%)	FAR (%)	MTTD (min)	DR (%)	FAR (%)	MTTD (min)
DELM	98.42	0.56	0.68	92.20	0.63	1.09
LSTM	94.64	0.48	0.90	86.52	0.65	1.24
DBN	96.41	0.50	0.88	88.86	0.70	1.12
CNN	96.24	0.54	0.94	86.42	0.66	1.14
GBDT	95.81	0.50	0.92	87.86	0.76	1.21

FIGURE 7: (a)DR-FAR curve. (b)MTTD-FAR curve.

Persistence test (PT) is an effective way to reduce FAR. Continuity test refers to continuous and interval detection of traffic incidents and ultimately determines the occurrence of traffic incidents.

When there is no continuous test, PT = 0. MTTD becomes longer as the number of continuous tests increases, and DR decreases as the number of continuous tests increases. PT affects the detection effect and detection efficiency to a certain extent. When PT = 0–4, the performance curves of the three algorithms are shown in Figure 7. As shown in Figure 7(a), it is DR-FAR curve, and, as shown in Figure 7(b), it is MTTD-FAR curve. As shown in Table 5, when PT = 0, the three indicators correspond to maximum values.

As can be seen from Figure 7(a), when PT = 4, the DR-FAR curve of DELM is closer to the upper left corner, indicating that the performance of the DELM algorithm is better. That is, under the same DR conditions, the FAR of the DELM algorithm is better than LSTM, DBN, CNN, and GBDT. The FAR of the algorithm is lower; under the same FAR condition, the DR of DELM algorithm is higher than that of LSTM, DBN, CNN, and GBDT algorithms. For the DR-FAR curve of each algorithm, when PT = 1, the FAR drop is larger and the DR drop is smaller. When the number of PT continues to increase, the decline in FAR gradually slows, and the decline in DR gradually accelerates.

As can be seen from Figure 7(b), when PT = 4, the MTTD-FAR curve of the DELM algorithm is closer to the lower left corner, indicating that the performance of the DELM algorithm is better, That is, under the same FAR, the

MTTD of the DELM algorithm is shorter than the MTTD of LSTM, DBN, CNN, and GBDT algorithms; under the same MTTD conditions, the MTTD of the DELM algorithm is shorter than the MTTD of LSTM, DBN, CNN, and GBDT algorithms. For the MTTD-FAR curve of each algorithm, when PT = 1, although the MTTD is increased to a certain extent, the decrease in FAR is larger. As the number of PT increases, the rate of decrease in FAR becomes slow, and the rate of increase in MTTD becomes faster.

In summary, the event detection results of the DELM algorithm continuous [18] inspection are better than those of LSTM, DBN, CNN, and GBDT. When PT = 1, each algorithm can better balance the three indicators of DR, FAR, and MTTD.

5. Conclusion

(1) The random forest algorithm used in the paper can effectively select important variables for traffic incident detection, reduce the input dimension of traffic incident detection algorithm, and improve the performance of traffic incident detection algorithm. (2) The performance of DELM algorithm is better than that of LSTM, DBN, CNN, and GBDT algorithms. (3) When PT = 1, each algorithm can better balance the three indicators of DR, FAR, and MTTD.

In order to get a more general conclusion, in future research, the DELM algorithm needs to be used in other traffic incident data sets and theoretically analyze and demonstrate the superiority of DELM for traffic incident detection. In addition, to construct a more comprehensive

traffic, the initial variable set for event detection needs further discussion.

Data Availability

The data supporting the conclusions of this study are presented in the figures and tables of the article. The code and details involved in this paper are available upon request from the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] C. L. Dudek, C. J. Messer, and N. B. Buckles, *Incident Detection on Urban Freeway*, Transportation Reseach Board, Washington, DC, USA, 1974.

[2] A. I. GalI and F. L. Hall, *Distinguishing between Incident Congestion and Recurrent Congestion: A Proposed Logic*, Transportation Research Board, Washington, DC, USA, 1989.

[3] C.-H Hsiao, C.-T. Lin, and M. Cassidy, "Application of fuzzy logic and neural network to automatically detect freeway traffic incident," *Journal of Transportation Engineering*, vol. 120, no. 5, pp. 753–772, 1994.

[4] B. Abdulhai and S. G. Ritchie, "Enhancing the universality and transferability of freeway incident detection using a Bayesian-based neural network," *Transportation Research Part C: Emerging Technologies*, vol. 7, no. 5, pp. 261–280, 1999.

[5] F. Yuan and R. L. Cheu, "Incident detection using support vector machines," *Transportation Research Part C: Emerging Technologies*, vol. 11, no. 3-4, pp. 309–328, 2003.

[6] G. B Huang, H. Zhou, X. Ding et al., "Extreme learning machine for regression and multicllass classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 42, no. 2, pp. 513–529, 2012.

[7] L. L. Chamara, H. Zhou, and G. B. Huang, "Extreme learning machines-representational learning with ELMS for big data," *Intelligent Systems, IEEE*, vol. 28, no. 6, pp. 31–34, 2013.

[8] J. T. Springenberg, A. Dosovitskiy, T. Brox et al., "Striving for simplicity: the all convolutional net," 2014, https://arxiv.org/abs/1412.6806.

[9] D. Zhang and D. Wang, "Relation classification via recurrent neural network," *Computer Science*, 2015.

[10] C. E. Hatri and J. Boumhidi, "Fuzzy deep learning based urban traffic incident detection," *Cognitive Systems Research*, vol. 50, pp. 206–213, 2018.

[11] L. Zhu, F. Guo, R. Krishnan, and J. W. Polak, "A deep learning approach for traffic incident detection in urban networks," in *Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1011–1016, IEEE, Maui, HI, USA, November 2018.

[12] M. Thaika, S. Tasneeyapant, and S. Cheamanunkul, "A fast, scalable, unsupervised approach to real-time traffic incident detection," in *Proceedings of the 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp. 205–210, IEEE, Nakhonpathom, Thailand, July 2018.

[13] T. Kongkhaensarn and M. Piantanakulchai, "Comparison of probabilistic neural network with multilayer perceptron and support vector machine for detecting traffic incident on expressway based on simulation data," in *Proceedings of the 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp. 221–226, Nakhonpathom, Thailand, July 2018.

[14] L. Li, Y. J. Zhang, Y. Zheng, and B. Ran, "Real-time traffic incident detection with classification methods," in *Green Intelligent Transportation Systems*, vol. 419, pp. 777–788, Springer, Berlin, Germany, 2018.

[15] Y. Sun and Z. X. Hou, "A novel abnormal traffic incident detection method based on improved support vector machine," *Journal of Applied Science and Engineering*, vol. 21, no. 1, pp. 45–50, 2018.

[16] L. C. Li, X. Qu, J. Zhang, and B. Ran, "Traffic incident detection based on extreme machine learning," *Journal of Applied Science and Engineering*, vol. 20, no. 4, pp. 409–416, 2017.

[17] F. Alesiani, L. Moreira-Matias, and M. Faizrahnemoon, "On learning from inaccurate and incomplete traffic flow data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 11, pp. 3698–3708, 2019.

[18] H. Benaddi, K. Ibrahimi, and A. Benslimane, "Improving the intrusion detection system for NSL-KDD dataset based on PCA-fuzzy clustering-KNN," in *Proceedings of the 2018 6th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, pp. 234–239, IEEE, Marrakesh, Morocco, October 2018.