

Research Article

Analysis and Prediction of Overloaded Extra-Heavy Vehicles for Highway Safety Using Machine Learning

Yi-Hsin Lin ¹, Suyu Gu,¹ Wei-Sheng Wu,² Rujun Wang,³ and Fan Wu¹

¹School of Civil Engineering, Southeast University, Nanjing 211189, China

²Department of Electrical Engineering, National Cheng Kung University, Tainan 70101, Taiwan

³Changzhou Transportation Comprehensive Administrative Law Enforcement Detachment, Changzhou 213001, China

Correspondence should be addressed to Yi-Hsin Lin; yihsinlin@seu.edu.cn

Received 8 October 2020; Revised 10 December 2020; Accepted 14 December 2020; Published 30 December 2020

Academic Editor: Hsu-Yang Kung

Copyright © 2020 Georgii S. Vasyliov et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Along with the prosperity and rapid development of the national economy, the transportation industry has rapidly developed in China. However, overloaded vehicles have been causing frequent traffic accidents. Thus, to alleviate or resolve the corresponding problems associated with highway engineering safety and the market economy, an improved technique for overload management is urgently required. In this study, to analyze the overload data on expressways and highways in China, we developed a machine learning model by comparing the performances of cluster analysis, backpropagation neural network (BPNN), generalized regression neural network (GRNN), and wavelet neural network (WNN) in analyzing global and local time series overload data. In a case study, our results revealed the trends of overloading on highways in Jiangsu Province. Given sufficient data, BPNN performed better than GRNN and WNN. As the amount of training data increased, GRNN performed better, but the runtime increased. WNN had the shortest runtime among the three methods and could reflect the future trends of the overload rate in the monthly data prediction of overload. Our model provides information with potential value for expressway network management departments through data mining. This information could help management departments allocate resources reasonably and optimize the information utilization rate.

1. Introduction

Along with the prosperity and rapid development of the national economy, the transportation industry has also rapidly developed in China. Nevertheless, the related overloading management system has not developed accordingly and the enforcement of transportation laws has not become stricter. Thus, transport operators hope to expand their profits by overloading their vehicles; however, overloading leads to frequent traffic accidents. According to the Highway Bureau of the Chinese Ministry of Transport, more than 80% of truck road traffic accidents are caused by overloaded transport. Overloading continuously causes road damage [1] and traffic accidents [2], disrupts the normal economic order of the logistics market, and affects the healthy development of the highway transportation economy [3]. Thus, to alleviate or resolve the problems associated

with highway engineering safety and market economy, an improved technique for overload management is urgently required.

In this study, we established a machine learning model for highway overrun and overload. We explored the overload characteristics through cluster analysis and compared the performances of backpropagation neural network (BPNN), a commonly employed method for training neural network, generalized regression neural network (GRNN), a radial basis function neural network, and wavelet neural network (WNN), a novel neural network that combines classical sigmoid neural networks with wavelet analysis (WA), to determine the best method to mine, analyze, and predict the overload data. Furthermore, we provide information with more potential value like overload characteristics, overload rate, and overload trend, for highway network management, and explore the application of data mining analysis results in

daily and instant overload management work using cluster analysis and neural networks (NNs).

This study makes a significant contribution in assisting highway management departments to reasonably allocate resources, optimize information utilization, improve the transportation management methods and management efficiency, and alleviate highway engineering safety problems. Furthermore, this study aids in economic and environmental development and provides a practical reference for future work on overload management.

2. Background

Rapid advancements in the field of science and technology have enabled the constant updation of software and hardware for overload control; consequently, toll collection centers at all levels in China are able to store increasing amounts of detailed historical data [4]. These massive amounts of data contain abundant valuable information. Conventional research methods for analyzing the characteristics of overloading in China mainly include literature review, statistical prediction model, and game analysis model. For example, Li and Wang analyzed the characteristics of freeway overload vehicles via cluster analysis [5]. Ryu et al. analyzed vehicle overrun behaviors using game theory and the risk preference model [6]. In addition, some scholars now study the related characteristics by building data models. Cheng et al. established a behavior judgment and prediction model and studied the vehicle detour behavior [7]. Zhang analyzed the impact of overload on road capacity through VISSIM simulation [8]. It is seen that most of the current research is based on data modeling to conduct data mining; thus, we explore which model should be used in the research as follows.

Past research has revealed that the overload characteristics [4, 9] should first be clarified to further study the problem of overload management. Machine learning, a typical data mining technique, uses the “black box” principle to build a machine learning model to address complex and diverse data [10]. In the field of traffic management, machine learning is currently applied to feature mining, prediction, response, and timely disposal of traffic events. For example, Muhammed Yasin and Ahmet proposed a highway traffic accident prediction model based on an artificial neural network (ANN) to analyze the characteristics of traffic accidents [11]. Li et al. proposed a situational awareness machine learning model to thoroughly investigate the factors affecting the fatigue degree of traffic management personnel [12]. Mohammed Abdulhafedh et al. reviewed data mining and machine learning methods for sustainable intelligent urban traffic classification and indicated that machine learning is extremely effective for the identification and classification of traffic states [13]. Therefore, machine learning can be used for data mining, studying the characteristics of overloading, and forecasting overloading patterns to thoroughly exploit the massive amounts of historical data in the existing domestic highway toll collection centers at all levels; this complies with the current demand for off-site law enforcement and technology. This

could significantly improve highway transportation management and alleviate highway engineering safety problems.

In terms of research on traffic transport characteristics, the currently adopted machine learning methods include support vector machine (SVM), logistic regression (LR), cluster analysis, and ANN.

SVM has been successfully employed for classification, regression, and pattern recognition [14]. Sangare et al. highlighted that SVM can demonstrate better performance with low data volume; however, it lacks the ability to automatically identify relevant features, and its calculation cost is high [15]. Lin and Li suggested that NNs are better than SVM in predicting traffic congestion [16]. The LR is a statistical method used to predict event probability and is a significant model for classified data [17]. However, it is sensitive to the multicollinearity of independent variables and affects regression results. Moreover, LR would be highly suited for feature mining through the analysis of the influencing factors. [18]. Because this research has made sufficient data available, one need not consider to use SVM solely because it performs better than other methods with less data. Furthermore, considering its high computational cost and poor prediction performance compared with that of NNs, SVM is not a practical choice. Moreover, independent variables affecting the overload may cause multicollinearity problems; thus, the regression results of LR are also affected. Therefore, SVM and LR have limitations in mining the characteristics of overload and the prediction of overload.

Cluster analysis is a typical and extremely effective method of exploring characteristics and laws. It can discover the internal distribution structure of data without prior knowledge regarding the correct results [19]. Yassin and Pooja used *K*-means clustering to extract hidden information from traffic accident data and created training sets; the accuracy of this method was 99.86% [19]. Žunić et al. used *K*-means clustering to analyze the impact of road, environment, vehicles, and drivers on traffic accidents and highlighted that clustering has been applied in several cases and instances in current professional and scientific practices in the field of traffic management [20]. The most commonly employed clustering analysis methods include the hierarchical clustering algorithm and dividing-based clustering algorithm. According to the multiperiod original data of this research system, clustering and *K*-means clustering, which are the most classic, universal, and commonly used algorithms in the field of transportation, are selected and are combined. *K*-means is easy to implement; however, the number of classification classes must be specified in advance. The use of system clustering can solve this problem; concurrently, *K*-means can also compensate for any defects in the inaccurate termination conditions of system clustering.

ANN can be used to analyze nonlinear factors, and ANN models demonstrate high accuracy and degree of fit. The commonly used ANNs include BPNN, GRNN, and WNN. For example, Wang et al. established a BPNN model to estimate the probability of collision results for different traffic accidents. The model accuracy was greater than 90%, and the model exhibited a good fitting effect [21]. Zhang and Zhang believed that the GRNN model has higher accuracy and reliability than historical average and vector

autoregression (VAR) in making short-term traffic flow predictions [22]. Hou et al. suggested that WNN has favorable adaptive ability and self-learning ability for forecasting short-term traffic volume [23]. BPNN is relatively mature in network theory and performance and has high self-learning and adaptive abilities, which enables it to extract “reasonable rules” between data in the event of unseen patterns and to adequately mine the data. In contrast to the global approximation algorithm of BPNN, GRNN is characterized by an optimal approximation, which can process unstable data and avoid the problem of the local optimum that could occur due to the BPNN, whereas WNN has significant advantages over sequential data, implying that these three NNs have different functions [3, 24]. Therefore, when sufficient historical data are available for training, BPNN, GRNN, and WNN can suitably adapt to the characteristics of overload behavior, such as the large impact of timing sequence, large randomness, and nonlinear influencing factors, and are considerably suitable for solving the problem of feature mining and prediction modeling of overload vehicles.

Most of the current research is based on data modeling for data mining; therefore, this study explores the models that can be used and finally selects two models of machine learning: clustering analysis and NN. Machine learning has been used in the field of transportation research but rarely used in the analysis of overload characteristics. Therefore, as discussed earlier, we explore the overload characteristics through cluster analysis and compare the performances of BP, GRNN, and WNN to determine the best method to mine, analyze, and predict overload data.

3. Materials and Methods

3.1. Experimental Data. The expressway data of Jiangsu Province (15 565 data units) and the national provincial highway data of Liyang city (30 764 data units), from 2018 to 2019, were collected from the expressway transportation management department of the Jiangsu Province. The datasets included “License plate,” “Lane,” “Time,” “Gross weight of vehicles and goods,” “Axle number,” “Rated weight,” “Overload standard,” “Overload rate,” and “Overload interval.” The details of these datasets are as follows.

License plate. License plate is the number plate identifying the vehicle, and the dataset is in the form of “苏 DETXXX,” where “苏” represents Jiangsu Province and “X” is an integer.

Lane. A lane is the road that a vehicle travels on. If there are three lanes, the one closest to the central zoning is defined as the first lane, the middle lane is the second lane, and the outer lane is the third lane.

Time. Time refers to the time when the car passes through the detection station; the dataset is in the form of “year/month/day, hour: minutes.”

Gross weight of vehicles and goods. This refers to the total weight of the vehicle and cargo; the unit is ton (e.g., “49 t”).

Axle number. Axle number refers to the total number of axles configured in the lower part of the vehicle underframe. Vehicles are usually divided into classes of one to six axles.

Rated weight. Rated weight refers to the maximum allowable vehicle load under the condition of ensuring safe driving. The unit is ton (e.g., “49 t”).

Overload standard. Overload standard determines whether the vehicle weight exceeds the rated weight indicated on the driving license. If it exceeds the rated weight, it is termed overweight.

$$\text{overload rate} = \frac{\text{actual weight} - \text{rated weight}}{\text{rated weight}}. \quad (1)$$

Overload interval. Overload interval is divided into “less than 5%,” “5–30%,” “30–50%,” “50–100%,” and “more than 100%.”

3.1.1. Data Preprocessing. The raw datasets in this research were not correctly formatted and could not be comprehended by computing machines. Before building the model, this problem had to be resolved to ensure favorable data quality. Consequently, data cleaning, missing data handling, encoding, and normalization were performed. Character data, such as “Maximum wheelbase,” of the province highway missed 2007 data in total, and the national highway data do not contain this character data; thus, we deleted these character data. Character data such as “Time,” “License plate,” and “Detection station type” were converted to numerical data (Table 1). All the data were normalized between zero and one to eliminate the differences in the orders of magnitude, which also can reduce the possibility of overfitting in the BPNN model.

3.1.2. Data Extraction. Data were extracted from raw data that were processed and were divided into expressway data and national and provincial highway data.

Cluster analysis extracts two-dimensional (2D) data for analysis. “Overload rate” is the concrete data of “Overload interval”; thus, “Overload interval” was discarded. The attributes were “License plate,” “Gross weight of goods and vehicles,” “Overload standard (rated weight),” “Axle number,” and “Overload rate,” whereas “Overload standard” is similar to the information represented by “Gross weight of goods and vehicles”; thus, we selected “Gross weight of goods and vehicles” to avoid duplication.

The NN can also extract 2D data for analysis. Because of the lack of data for December and speed data from June to November in 2019, for the national and provincial data, only the national and provincial data of Liyang city in 2018 were selected for analysis. In May 2019, there were 738 expressway data points in Jiangsu Province in hours. In this research, the national and provincial highway data of Liyang city in 2018 were selected according to monthly, quarterly, and annual data. The monthly data were measured in hours and contributed 289 pieces of data for the 12 months of 2018. The

TABLE 1: Data conversion table.

| Data | Before conversion | After conversion |
|---------------------------|------------------------------------|---|
| Time | X year x month x day XX:XX:XX | The month is represented with a value in the range 1–12 The hour is represented by a value in the range 0–23 |
| | X year x month x day XX:XX:XX | |
| License plate | License plate in the province | [1, 0, 0] |
| | License plate outside the province | [0, 1, 0] |
| | No license plate | [0, 0, 1] |
| Type of detection station | Ordinary | [1, 0, 0] |
| | Dynamic | [0, 1, 0] |
| | High speed | [0, 0, 1] |

quarterly data were measured in hours, and 964 data points were available in the first quarter of 2018. The annual data were measured in hours, and 3367 pieces of data were available for 306 days in 2018.

In summary, cluster analysis was performed to extract the attributes “Time,” “Gross weight of vehicles and goods,” “Axle number,” and “Overload rate” to analyze the data. The BPNN and GRNN extracted “Time,” “Gross weight of goods and vehicles,” “Axle number,” and “Overload rate” from Jiangsu expressway and the provincial highway data. WNN was only used to process 2D time-series data, i.e., the “Overload rate” data.

3.2. Machine Learning

3.2.1. Cluster Analysis. Clustering, which was proposed in the mid-20th century, is a typical research method of unsupervised learning. It aims to group similar observations into multiple clusters according to the observations of multiple variables of each individual, to maximize the similarity of samples in each cluster, maximize the differences between unknown groups, and reveal the inherent distribution structure of data. The concept underlying cluster analysis originated from traditional taxonomy, in which complex classification problems are encountered. Unlike traditional classification, the data types that need to be divided in cluster analysis are typically unknown and are defined by analyzing specific cluster results.

- (1) Systematic clustering: systematic clustering aims to classify each sample step by step according to the criterion of the distance function calculation until the samples satisfy the requirements of the systematic clustering algorithm. Among them, the distance calculation formulas and criteria are primarily used to calculate the average distance between samples and clusters formed in each iteration and between them and each sample. Different distance calculation formulas and criteria yield different distance calculation results. The common distance calculation criteria include the shortest distance method, longest distance method, intermediate distance method, barycenter method, and class average distance method.
- (2) *K*-means clustering: *K*-means is considered a crucial clustering algorithm [25], which has an iterative process for datasets. In this iterative process, the

datasets are divided into “*k*” predefined nonoverlapping clusters or subgroups. The data points between clusters are as similar as possible, and the distance between clusters is retained for the maximum time possible. It assigns data points to clusters, so that the sum of the square of the distance between the cluster centroid and data points is the smallest. Here, the cluster centroid is the arithmetic average of the data points in the cluster.

In this study, systematic clustering and *K*-means clustering were combined for clustering analysis; the specific process is shown in Figure 1. *K*-means clustering needs to determine the number of clustering classes “*k*” before clustering, and this research aims to cluster the overloading data of the national and provincial highways in Liyang city in every quarter from 2018 to 2019. Therefore, we used systematic clustering to obtain the clustering tree diagram for eight quarters from 2018 to 2019 and determine the same number of clustering classes “*k*” in eight quarters, taking into account the continuity of the data. On this basis, the clustering results were obtained by *K*-means analysis.

3.2.2. Neural Network

- (1) BPNN ANN is a supervised machine learning algorithm [26]. The BPNN is called an error back-propagation ANN. This is a commonly employed method for training NNs, which assists in calculating the gradient of the loss function relative to all weights in the network and enhancing the connectivity between layers to obtain the optimal solution [27]. A BPNN usually has three or more layers. Each layer of a BPNN comprises several neurons.

In this research, a BPNN was used to build a machine learning model for the expressway data and to predict the overload rate and trend. The algorithm flow of the BPNN can be divided into three steps Figure 2: BPNN establishment, training, and prediction.

First, the specific structure of the BPNN was determined according to the characteristics of the fitting function. Because there are four input parameters—“Time,” “Gross weight of vehicles and goods,” “Axle number,” and “Type of inspection station”—and only one output parameter, namely, “Overload rate,” the number of nodes in the three

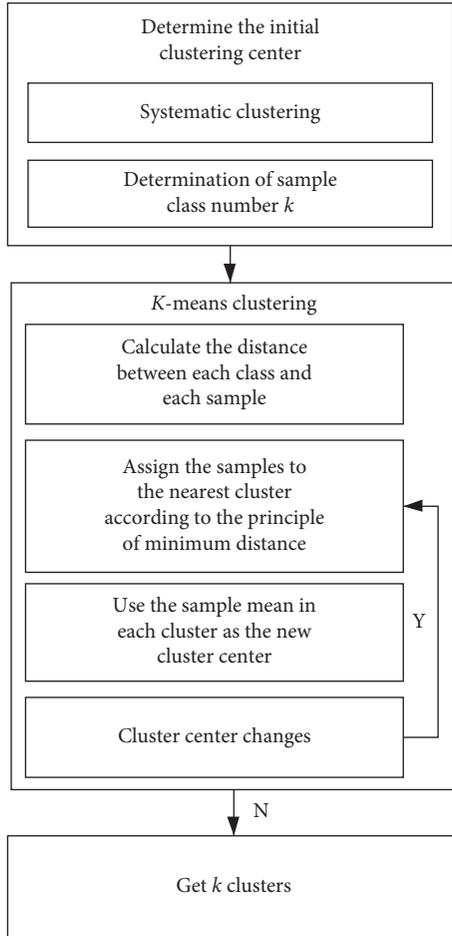


FIGURE 1: Flowchart for cluster analysis.

layers of the BPNN was “4-?-1”; this implies there are four input layer nodes, the number of hidden layer nodes is unknown, and there is one output layer node.

In reality, we can only use data from the first few months that have occurred to predict subsequent results. Therefore, to consider actual situations, in this research, we arranged the data in chronological order and systematically divided them into training data and testing data in a 10:1 ratio.

The number of hidden layer nodes can be calculated as follows:

$$L = \sqrt{n + m + a}, \quad (2)$$

where L denotes the number of hidden layer nodes, n denotes the number of neurons in the input layer, m denotes the number of neurons in the output layer, and a is a constant having a value between 1 and 10. Substituting $n = 4$ and $m = 1$ in equation (2), we obtain a value between $\sqrt{5} + 1$ and $\sqrt{5} + 10$ because the number of nodes L is an integer between 4 and 12. If the fitted nonlinear function is simple, the prediction error of the BPNN decreases with an increase in the number of nodes. If the function is complex, the error decreases first and then increases.

To explore the appropriate number of hidden nodes, a loop was written in the code to evaluate the goodness of fit of the prediction by the degree of fitting to obtain the optimal number of hidden nodes.

Goodness of fit refers to the fitting accuracy of the regression line of the model to the observed value. Generally, the determinable coefficient (also called the deterministic coefficient) is used to determine the goodness of fit. The closer the value is to one, the better the fitting degree of the model is; in contrast, the closer it is to zero, the worse the fitting degree of the model is. In this research, the prediction ability of the model was evaluated by calculating the coefficient of determination, as follows:

$$R^2 = \frac{\left(l \sum_{i=1}^l \hat{y}_i y_i - \sum_{i=1}^l \hat{y}_i \sum_{i=1}^l y_i \right)^2}{\left(l \sum_{i=1}^l \hat{y}_i^2 - \left(\sum_{i=1}^l \hat{y}_i \right)^2 \right) \left(l \sum_{i=1}^l y_i^2 - \left(\sum_{i=1}^l y_i \right)^2 \right)}, \quad (3)$$

where R^2 is the determinable coefficient, \hat{y}_i is the i th predicted value, y_i is the i th true value, and l denotes the number of samples.

- (2) Generalized regression neural network GRNN was proposed by Donald F. Specht, an American scholar, in 1991. It is essentially a radial basis function NN but differs from the conventional radial basis function NN. It is a type of strong regression tool with strong nonlinear mapping ability and flexible network structure and can be employed in regression, prediction, and classification research [28].

In this research, the GRNN was used to build a machine learning model for the expressway data and to predict the overload rate and trend. The algorithm flow of the GRNN can be divided into three steps Figure 3-GRNN establishment, training, and prediction.

First, the GRNN structure is determined according to the characteristics of the fitting function. The number of neuron nodes in the input layer is equal to the dimension of the learning samples. The number of neuron nodes in the mode layer is equal to the number of learning samples, i.e., “Time,” “Gross weight of vehicle and goods,” “Axle number,” and “Type of inspection station,” whereas that in the output layer is equal to the dimension of the output vector of the learning samples, i.e., “Overload rate.” Therefore, the GRNN does not need to determine the number of hidden layer nodes. Unlike the BPNN, the GRNN only needs to confirm the SPREAD parameter and thus exhibits considerable computational advantages. In this study, the parameters of SPREAD were determined by loop training. The loop code starts with “for SPREAD=0.1:0.1:2,” which means SPREAD is assigned

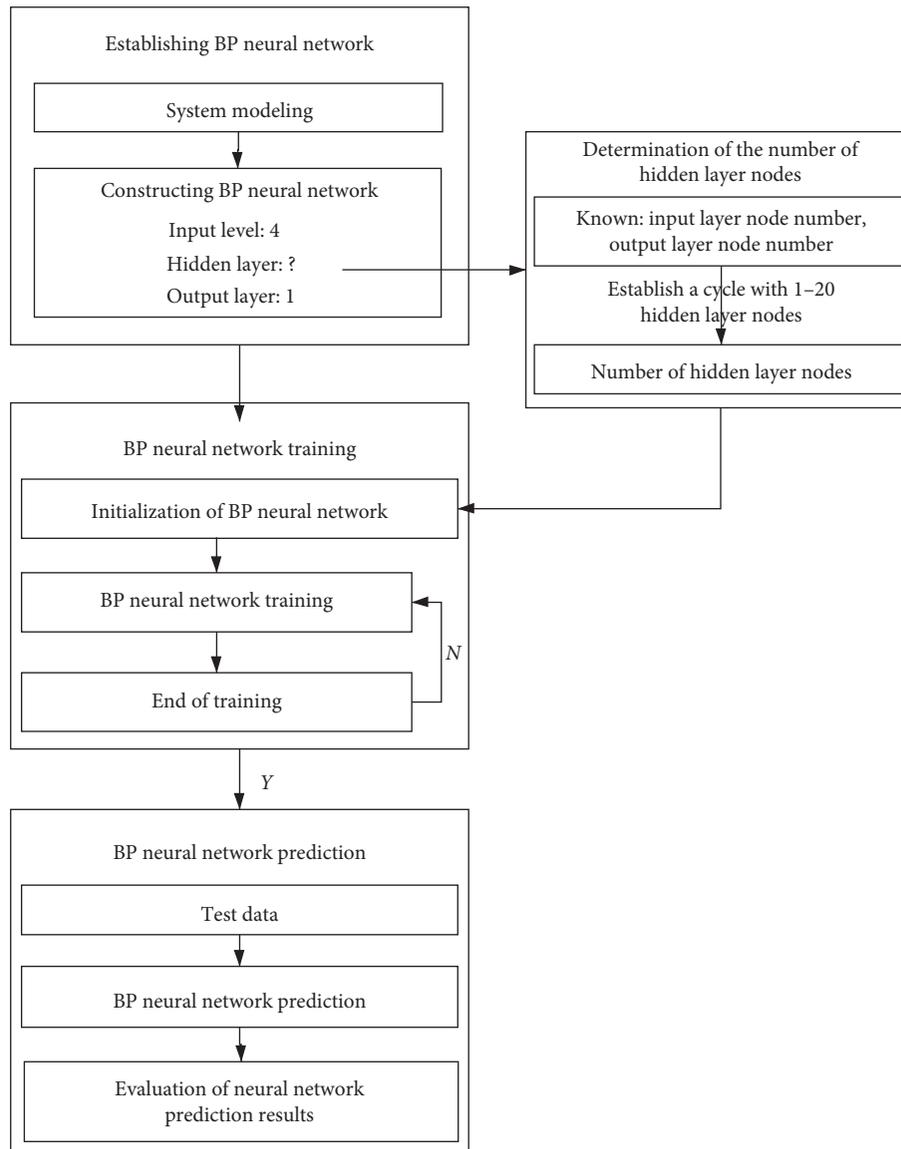


FIGURE 2: Flowchart of the BPNN.

from 0.1 and the step distance is 0.1, and each loop increases by 0.1 until $SPREAD = 2$. Similar to the manner followed for BPNN, for GRNN, the data are arranged in chronological order and systematically divided into training data and testing data in a 10:1 ratio.

- (3) Wavelet neural network WNN is a novel NN that combines classical sigmoid NNs with wavelet analysis (WA) [29]. The concept of WA was developed considering the shortcomings of the Fourier transform, which cannot accurately evaluate the time a certain signal occurs. In contrast, WA can analyze the local characteristics of signals through wavelet basis function transformation and select the direction of signals in two dimensions. The WNN can perform time-frequency local analysis, has a fast convergence speed, and can effectively avoid falling into a local optimum.

In this study, WNN was used to build a machine learning model for expressway, national, and provincial highway data and predict the overload rate and trend. The algorithm flow of the WNN can be divided into three steps Figure 4: WNN establishment, training, and prediction.

First, the structure of the WNN was determined according to the characteristics of the fitting function. In this study, according to the periodic characteristics of overload and goodness of fit of the operation results, the number of neuron nodes in the three layers of the NN was determined to be "4-4-1." Four nodes in the input layer indicate the overload rate at four time points before the predicted time node. The hidden layer nodes were composed of wavelet basis functions. There was one node in the output layer, which denoted the overload rate to be predicted. Similar to the manner followed for BPNN and GRNN, for WNN, the data are arranged in chronological order and systematically divided into training data and testing data in a 10:1 ratio.

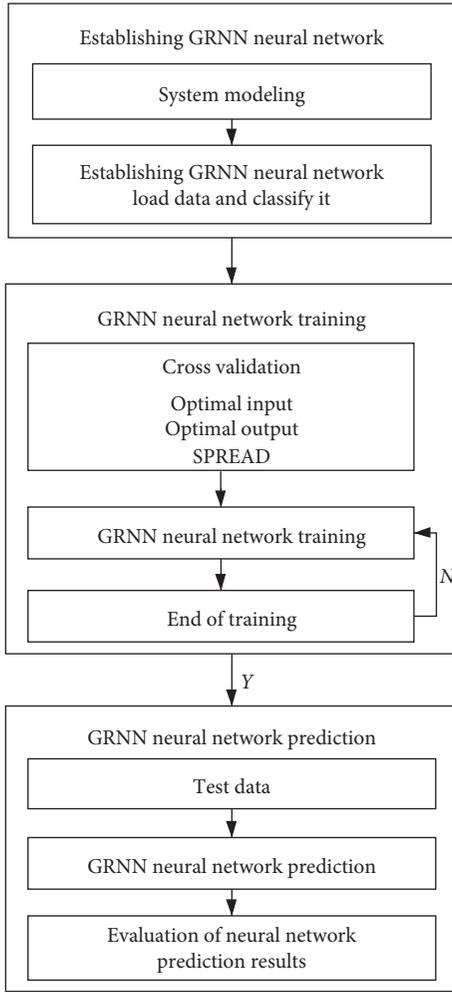


FIGURE 3: Flowchart of GRNN neural network.

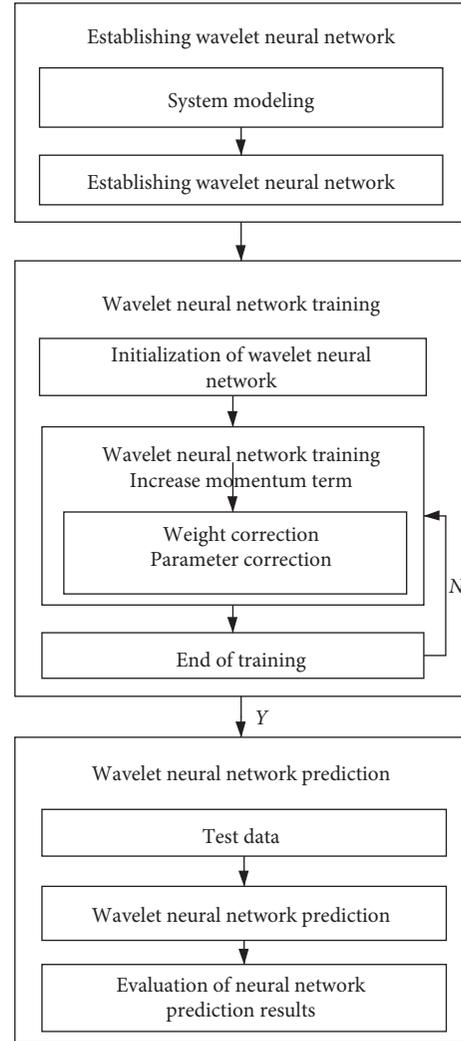


FIGURE 4: Flowchart of the WNN.

4. Results

Table 2 presents the clustering results when the number of clustering classes was six, i.e., $K = 6$. “Gross weight of vehicle and goods” gives the average gross weight of the vehicles and goods at the time of entering highway. “Category of gross weight” and “Proportion of gross weight” are the proportions of the components of the average gross weight. “Axis number” and “Proportion of axis number” are the proportions of the components of the axis number at the time of entering the highway. “Overload rate” gives the average overload rate of the vehicles at the time of entering highway. “Proportion of vehicles” shows the percentage of vehicles passing at the time of entering the highway in the total number of vehicles.

It can be observed that, on the expressways in Jiangsu Province, vehicles with six axles and weighing above 49 t entering the expressways at 23:00–04:00 and 12:00–14:00 are more likely to exhibit overload behavior, and the overload rate of these vehicles is higher than that of others during this period. Most vehicles enter the expressway at 05:00 and 17:00–22:00, among which six-axle vehicles weighing above 49 t are more likely to be overloaded. Among the vehicles

entering the expressway between 06:00–07:00 and 15:00–16:00, two-axle vehicles weighing 18–27 t are more likely to be overloaded. Among the vehicles entering the expressway between 08:00–11:00, four-axle vehicles weighing 37–43 t and 43–49 t are more likely to be overloaded.

Table 3 presents the clustering results of the national and provincial data in the eight quarters of 2018 and 2019. The definitions of “Gross weight of vehicle and goods,” “Category of gross weight,” “Proportion of gross weight,” “Axis number,” “Proportion of axis number,” “Overload rate,” and “Proportion of vehicles” are the same as above. The analysis is as follows.

The data clustering results of the national and provincial highways in the four quarters of 2018 were as follows: (1) most vehicles entered the provincial highway at 00:00–08:00 and 16:00–20:00 in the first quarter, among which six-axle vehicles weighing above 49 t were more likely to be overloaded; the overload rate of six-axle vehicles weighing above 49 t and entering the provincial highway at 02:00–03:00 was higher. Among the vehicles entering the national provincial highway at 11:00, 2-axle, 3-axle, and 4-axle vehicles weighing

TABLE 2: Clustering results of expressway data at $K=6$.

| Group | Time of entering highway | Gross weight of vehicle and goods | Category of gross weight | Proportion of gross weight (%) | Axis number | Proportion of axis number (%) | Overload rate | Proportion of vehicles (%) |
|-------|--------------------------|-----------------------------------|--------------------------|--------------------------------|-------------|-------------------------------|---------------|----------------------------|
| 1 | 23:00–4:00 | 90.62 | Over 49 t | 100.00 | 6 | 100.00 | 0.85 | 1.16 |
| | | | 43 t–49 t | 2.43 | 3 | 0.50 | | |
| 2 | 5:00 | 53.11 | Over 49 t | 97.57 | 3 | 0.50 | 0.11 | 50.83 |
| | | | | | 4 | 3.69 | | |
| | | | | | 5 | 4.25 | | |
| | | | | | 6 | 90.58 | | |
| 3 | 6:00–7:00 15:00–16:00 | 21.64 | 18 t–27 t | 80.66 | 2 | 86.68 | 0.13 | 5.21 |
| | | | 27 t–36 t | 19.34 | 3 | 13.32 | | |
| 4 | 8:00–11:00 | 41.88 | Over 49 t | 97.57 | 27 t–36 t | 2.37 | 0.17 | 6.72 |
| | | | | | 36 t–37 t | 1.17 | | |
| | | | | | 37 t–43 t | 55.47 | | |
| | | | | | 43 t–49 t | 40.99 | | |
| 5 | 12:00–14:00 | 67.04 | Over 49 t | 100.00 | 4 | 1.40 | 0.38 | 9.70 |
| | | | | | 5 | 0.33 | | |
| | | | | | 6 | 97.62 | | |
| | | | | | 9 | 0.66 | | |
| 6 | 17:00–22:00 | 58.44 | Over 49 t | 100.00 | 4 | 0.73 | 0.20 | 26.38 |
| | | | | | 5 | 0.91 | | |
| | | | | | 6 | 97.88 | | |
| | | | | | 9 | 0.48 | | |

27–36 t and 37–49 t were more likely to be overloaded. (2) In the second quarter, most vehicles entered the highway between 00:00–08:00 and between 14:00–21:00, among which 6-axle vehicles weighing above 49 t were more likely to be overloaded; the overload rate of 6-axle vehicles weighing above 49 t and entering the highway between 21:00–23:00 was high. Among the vehicles entering the national provincial highway from 09:00 to 13:00, 3-axle vehicles weighing 27–36 t and 37–49 t were more likely to be overloaded. (3) In the third quarter, most vehicles entered the highway between 00:00–8:00 and between 16:00–20:00, among which 6-axle vehicles weighing above 49 t were more likely to be overloaded; the overload rate of 6-axle vehicles weighing above 49 t and entering the highway between 21:00–23:00 was high. Among the vehicles entering the highway between 09:00–15:00, 2-axle, 3-axle, and 4-axle vehicles weighing 27–36 t and 37–49 t were more likely to be overloaded. (4) In the fourth quarter, most vehicles entered the highway between 00:00–08:00 and 13:00–22:00, among which 6-axle vehicles weighing above 49 t were more likely to be overloaded. At 23:00, the overload rate of 6-axle vehicles weighing above 49 t and entering the highway was high. Among the vehicles entering the highway between 09:00–12:00, 4-axle vehicles weighing 27–36 t and 37–49 t were more likely to be overloaded.

The data clustering results of the national and provincial highways in the four quarters of 2019 were as follows: (1) in the first quarter, most vehicles entered the national and provincial highway between 04:00–13:00 and between 18:00–23:00, among which 6-axle vehicles weighing above 49 t were more likely to be overloaded; the overload rate of 6-axle vehicles weighing above 49 t and entering the provincial highway between 00:00–03:00 was high. Among the vehicles

entering the highway between 14:00–17:00, 2-axle vehicles weighing 18–27 t and 27–36 t were more likely to be overloaded. (2) In the second quarter, most vehicles entered the highway between 00:00–15:00, among which 6-axle vehicles weighing above 49 t were more likely to be overloaded. From 00:00 to 03:00, the overload rate of 6-axle vehicles weighing above 49 t and entering the highway was high. Among the vehicles entering the national provincial highway between 06:00–15:00, 4-axle vehicles weighing 37–43 t were more likely to be overloaded. Among the vehicles entering the national highways between 16:00–23:00, 4-axle and 6-axle vehicles weighing above 49 t were more likely to be overloaded. (3) In the third quarter, most vehicles entered the highway between 04:00–05:00 and between 14:00–18:00, among which 6-axle vehicles weighing above 49 t were more likely to be overloaded. From 02:00 to 03:00, the overload rate of 6-axle vehicles weighing above 49 t and entering the national provincial highway was high. Among the vehicles entering the highways between 22:00–01:00 and between 07:00–13:00, 4-axle vehicles weighing 37–43 t were more likely to be overloaded. Among the vehicles entering the highway at 06:00, 2-axle and 3-axle vehicles weighing 27–36 t were more likely to be overloaded. (4) In the fourth quarter, most vehicles entered the highway between 08:00–15:00, among which 6-axle vehicles weighing above 49 t were more likely to be overloaded. The overload rate of 6-axle vehicles weighing above 49 t and entering the highway between 00:00–04:00 was high. Among the vehicles entering the highway from 05:00 to 07:00, 2-axle vehicles weighing 18–27 t and 27–36 t were more likely to be overloaded. Among the vehicles entering the highway from 16:00 to 23:00, 4-axle and 6-axle vehicles weighing above 49 t were more likely to be overloaded.

TABLE 3: Clustering results of national and provincial highway data when $K = 4$.

| Year | Quarter | Group | Time of entering highway | Gross weight of vehicle and goods | Category of gross weight | Proportion of gross weight | Axis number | Proportion of axis number (%) | Overload rate | Proportion of vehicles (%) | |
|------|---------|-------------|--------------------------|-----------------------------------|--------------------------|----------------------------|-------------|-------------------------------|---------------|----------------------------|--|
| 2018 | 1 | 1 | 0:00-1:00 4:00-10:00 | 53.96 | 37t-43t | 1.57 | 2 | 0.02 | | | |
| | | | | | 43t-49t | 3.18 | 3 | 0.75 | | | |
| | | | | | Over 49t | 95.25 | 4 | 4.59 | 0.14 | 40.36 | |
| | | | | | Over 49t | 100 | 5 | 0.29 | | | |
| | 2 | 2 | 2:00-3:00 | 81.75 | 18t-27t | 21.16 | 2 | 0.08 | | | |
| | | | | | 27t-36t | 36.83 | 3 | 5.25 | 0.13 | 9.56 | |
| | | | | | 36t-37t | 2.17 | 4 | 25.61 | | | |
| | | | | | 37t-43t | 39.84 | 6 | 43.64 | | | |
| | 3 | 3 | 11:00 | 29.02 | 37t-43t | 2.76 | 2 | 0.08 | | | |
| | | | | | 43t-49t | 3.6 | 4 | 5.25 | 0.12 | 40.57 | |
| | | | | | Over 49t | 93.65 | 6 | 93.11 | | | |
| | | | | | Over 49t | 100 | 2 | 0.01 | | | |
| 4 | 4 | 12:00-23:00 | 53.19 | 37t-43t | 2.76 | 2 | 0.01 | | | | |
| | | | | 43t-49t | 3.61 | 3 | 1.56 | 0.13 | 47.95 | | |
| | | | | Over 49t | 93.63 | 4 | 4.97 | | | | |
| | | | | Over 49t | 100 | 6 | 92.91 | | | | |
| 2 | 2 | 9:00-13:00 | 28.53 | 18t-27t | 2.243 | 2 | 26.58 | | | | |
| | | | | 27t-36t | 37.78 | 3 | 44.84 | 0.12 | 8.73 | | |
| | | | | 36t-37t | 1.99 | 4 | 28.58 | | | | |
| | | | | 37t-43t | 37.79 | 4 | 28.58 | | | | |
| 3 | 3 | 14:00-21:00 | 51.97 | 37t-43t | 4.78 | 2 | 0.02 | | | | |
| | | | | 43t-49t | 5.3 | 3 | 3.26 | 0.14 | 32.99 | | |
| | | | | Over 49t | 89.92 | 5 | 1.22 | | | | |
| | | | | Over 49t | 100 | 6 | 89.09 | | | | |
| 4 | 4 | 22:00-23:00 | 77.5 | 37t-43t | 3.06 | 2 | 0.58 | | | | |
| | | | | 43t-49t | 4.09 | 3 | 2.56 | 0.62 | 10.33 | | |
| | | | | Over 49t | 92.85 | 5 | 0.32 | | | | |
| | | | | Over 49t | 100 | 6 | 96.52 | | | | |
| 2018 | 1 | 1 | 0:00-8:00 | 53.37 | 37t-43t | 3.06 | 2 | 0.01 | | | |
| | | | | | 43t-49t | 4.09 | 3 | 0.31 | | | |
| | | | | | Over 49t | 92.85 | 4 | 6.23 | 0.12 | 44.59 | |
| | | | | | Over 49t | 100 | 5 | 1.08 | | | |
| | 2 | 2 | 9:00-15:00 | 28.01 | 18t-27t | 25.81 | 2 | 30.12 | | | |
| | | | | | 27t-36t | 33.61 | 3 | 37.33 | 0.11 | 8.36 | |
| | | | | | 36t-37t | 1.07 | 4 | 32.55 | | | |
| | | | | | 37t-43t | 39.51 | 6 | 92.37 | | | |
| | 3 | 3 | 16:00-20:00 | 53.46 | 37t-43t | 2.57 | 3 | 0.31 | | | |
| | | | | | 43t-49t | 3.09 | 4 | 4.61 | 0.12 | 41.12 | |
| | | | | | Over 49t | 94.34 | 5 | 0.96 | | | |
| | | | | | Over 49t | 100 | 6 | 94.12 | | | |
| 4 | 4 | 21:00-23:00 | 88.53 | 37t-43t | 2.57 | 3 | 1.55 | | | | |
| | | | | 43t-49t | 3.09 | 4 | 98.45 | 0.82 | 5.93 | | |
| | | | | Over 49t | 100 | 6 | 98.45 | | | | |
| | | | | Over 49t | 100 | 2 | 0.2 | | | | |
| 2018 | 1 | 1 | 0:00-8:00 | 57.05 | 43t-49t | 6.81 | 3 | 1.2 | 0.25 | 41.18 | |
| | | | | | Over 49t | 93.19 | 4 | 10.7 | | | |
| | | | | | Over 49t | 100 | 5 | 2.01 | 0.25 | 41.18 | |
| | | | | | Over 49t | 100 | 6 | 85.89 | | | |
| | 2 | 2 | 11:00 | 97.49 | 37t-43t | 48.45 | 3 | 32.11 | 0.2 | 17.9 | |
| | | | | | 43t-49t | 3.73 | 4 | 42.82 | | | |
| | | | | | 37t-43t | 0.55 | 3 | 0.4 | | | |
| | | | | | 43t-49t | 5.69 | 4 | 12.03 | 0.23 | 30.47 | |
| | 3 | 3 | 13:00-22:00 | 57.03 | Over 49t | 93.76 | 5 | 1.12 | | | |
| | | | | | Over 49t | 100 | 6 | 86.45 | | | |
| | | | | | Over 49t | 100 | 3 | 0.38 | 1.02 | 10.45 | |
| | | | | | Over 49t | 100 | 4 | 2.52 | | | |
| 4 | 4 | 11:00 | 97.49 | Over 49t | 99.79 | 5 | 1.26 | | | | |
| | | | | Over 49t | 100 | 6 | 95.85 | | | | |
| | | | | Over 49t | 100 | 5 | 1.26 | | | | |
| | | | | Over 49t | 100 | 6 | 95.85 | | | | |

TABLE 3: Continued.

| Year | Quarter | Group | Time of entering highway | Gross weight of vehicle and goods | Category of gross weight | Proportion of gross weight (%) | Axis number | Proportion of axis number (%) | Overload rate | Proportion of vehicles (%) | | |
|------|-------------|-------------|--------------------------|-----------------------------------|--------------------------|--------------------------------|-------------|-------------------------------|---------------|----------------------------|-------|--|
| 2019 | 1 | 1 | 0:00-3:00 | 75.63 | Over 49t | 100 | 4 | 2.5 | 0.56 | 16.59 | | |
| | | | | | 371-431 | 8.73 | 6 | 97.5 | | | | |
| | | 2 | 4:00-10:00 | 51.33 | 431-491 | 7.48 | 4 | 14.29 | | | | |
| | | | | | Over 49t | 83.79 | 6 | 85.71 | 0.1 | 35.21 | | |
| | 3 | 11:00-13:00 | 49.99 | 371-431 | 9.13 | 3 | 0.72 | | | | | |
| | | | | 431-491 | 8.52 | 4 | 12.56 | 0.08 | 43.28 | | | |
| | | 4 | 14:00-17:00 | 21.84 | Over 49t | 82.35 | 5 | 1.21 | | | | |
| | | | | | 18t-27t | 67.86 | 2 | 85.51 | 0.11 | 4.92 | | |
| | 2 | 1 | 0:00-3:00 | 84.11 | Over 49t | 100 | 3 | 0.6 | | | | |
| | | | | | 271-361 | 32.14 | 3 | 27.27 | 0.75 | 33 | | |
| | | | 2 | 4:00-5:00 | 57.2 | 431-491 | 3.15 | 3 | 0.65 | | | |
| | | | | | | Over 49t | 96.85 | 5 | 10.34 | 0.24 | 31.67 | |
| | | 3 | 6:00-15:00 | 34.57 | 18t-27t | 9.84 | 2 | 87.93 | | | | |
| | | | | | 271-361 | 14.77 | 3 | 15.48 | | | | |
| | | | 4 | 16:00-23:00 | 50.42 | 371-431 | 58.29 | 2 | 17.42 | 0.17 | 21.41 | |
| | | | | | | 431-491 | 17.1 | 4 | 67.1 | | | |
| | | 3 | 22:00-1:00 | 39.83 | 371-431 | 23.9 | 3 | 1.51 | | | | |
| | | | | | 431-491 | 11.95 | 4 | 38.19 | 0.19 | 13.92 | | |
| | | | 4 | 7:00-13:00 | 55.37 | Over 49t | 64.14 | 6 | 60.3 | 0.19 | 13.92 | |
| | | | | | | 271-361 | 4.29 | 2 | 2.8 | | | |
| | | 4 | 8:00-15:00 | 51.71 | 361-371 | 2.86 | 3 | 8.41 | | | | |
| | | | | | 431-491 | 75 | 4 | 85.98 | 0.16 | 29.29 | | |
| | | | 1 | 0:00-4:00 | 78.89 | Over 49t | 12.86 | 6 | 2.8 | | | |
| | | | | | | Over 49t | 5 | | | | | |
| 3 | | 2:00-3:00 | 25.87 | Over 49t | 95.96 | 6 | 90.91 | 0.204 | 0.4655 | | | |
| | | | | Over 49t | 100 | 4 | 7.69 | 0.72 | 13.18 | | | |
| | | 4 | 0.25 | 81.43 | 18t-27t | 40 | 2 | 57.14 | 0.25 | 10.98 | | |
| | | | | | 271-361 | 60 | 3 | 42.86 | | | | |
| 4 | | 5:00-7:00 | 25.4 | Over 49t | 100 | 3 | 1.96 | | | | | |
| | | | | 18t-27t | 48.53 | 2 | 90.2 | 0.72 | 7.45 | | | |
| | | 3 | 8:00-15:00 | 50.15 | 271-361 | 45.58 | 2 | 71.68 | | | | |
| | | | | | 361-371 | 3.63 | 3 | 26.32 | 0.29 | 17.38 | | |
| 4 | 16:00-23:00 | 51.71 | 371-431 | 2.27 | 4 | 2.01 | | | | | | |
| | | | 371-431 | 16.39 | 2 | 0.16 | | | | | | |
| | 1 | 0:00-3:00 | 84.11 | Over 49t | 100 | 3 | 5.88 | | | | | |
| | | | | 431-491 | 10.02 | 4 | 22.57 | 0.19 | 63.71 | | | |
| 3 | 4:00-10:00 | 51.33 | Over 49t | 73.59 | 5 | 0.39 | | | | | | |
| | | | 371-431 | 13.75 | 6 | 71 | | | | | | |
| | 4 | 7:00-13:00 | 50.42 | 431-491 | 16.49 | 3 | 10.67 | | | | | |
| | | | | Over 49t | 69.76 | 4 | 17.78 | 0.25 | 11.47 | | | |
| 4 | 11:00-13:00 | 49.99 | 431-491 | 8.52 | 4 | 2.22 | | | | | | |
| | | | Over 49t | 82.35 | 5 | 69.33 | | | | | | |

Clustering results of national and provincial highway data when $K = 4$ (1). Clustering results of national and provincial highway data when $K = 4$ (2). Clustering results of national and provincial highway data when $K = 4$ (3). Clustering results of national and provincial highway data when $K = 4$ (4). Clustering results of national and provincial highway data when $K = 4$ (5). Clustering results of national and provincial highway data when $K = 4$ (6)

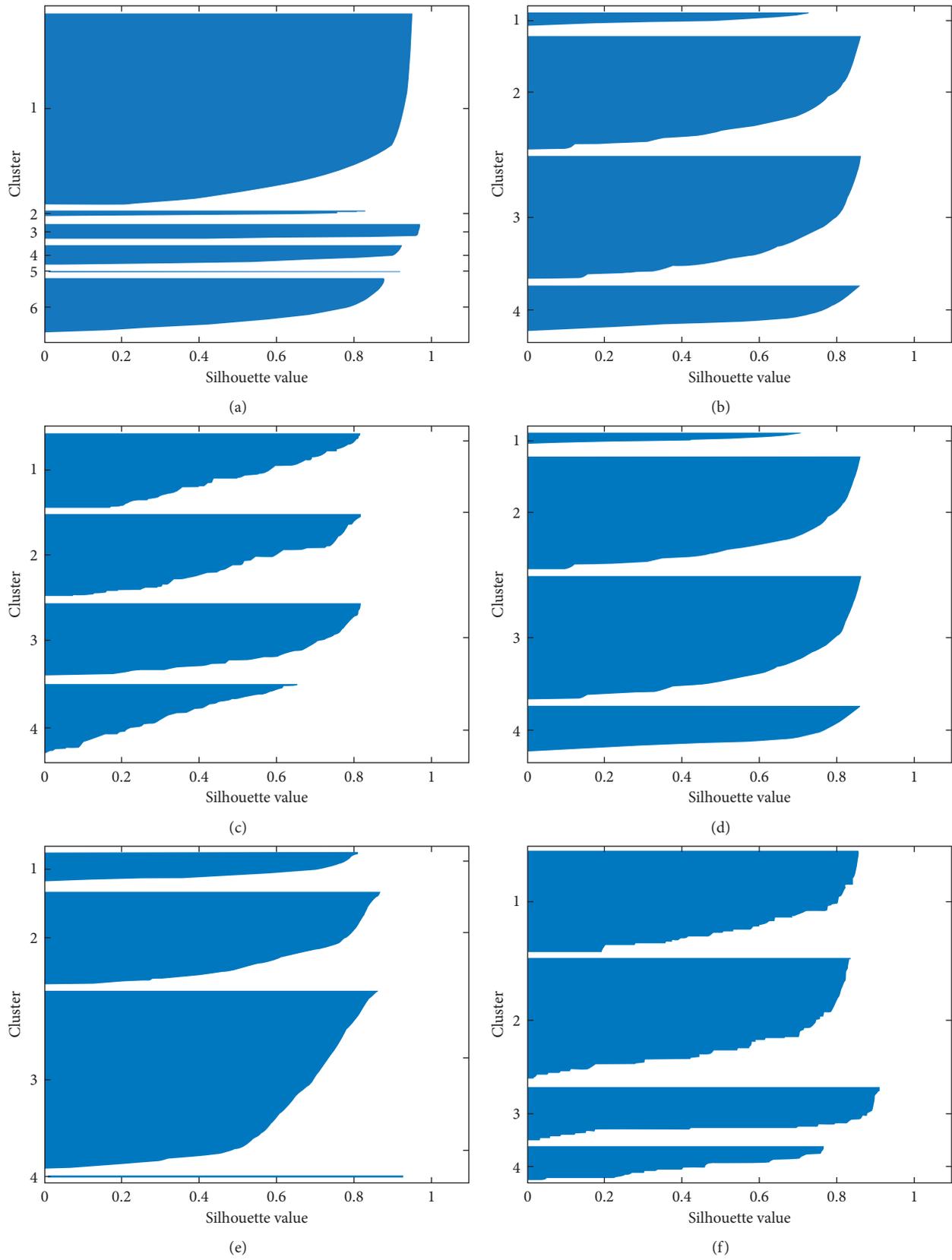


FIGURE 5: Continued.

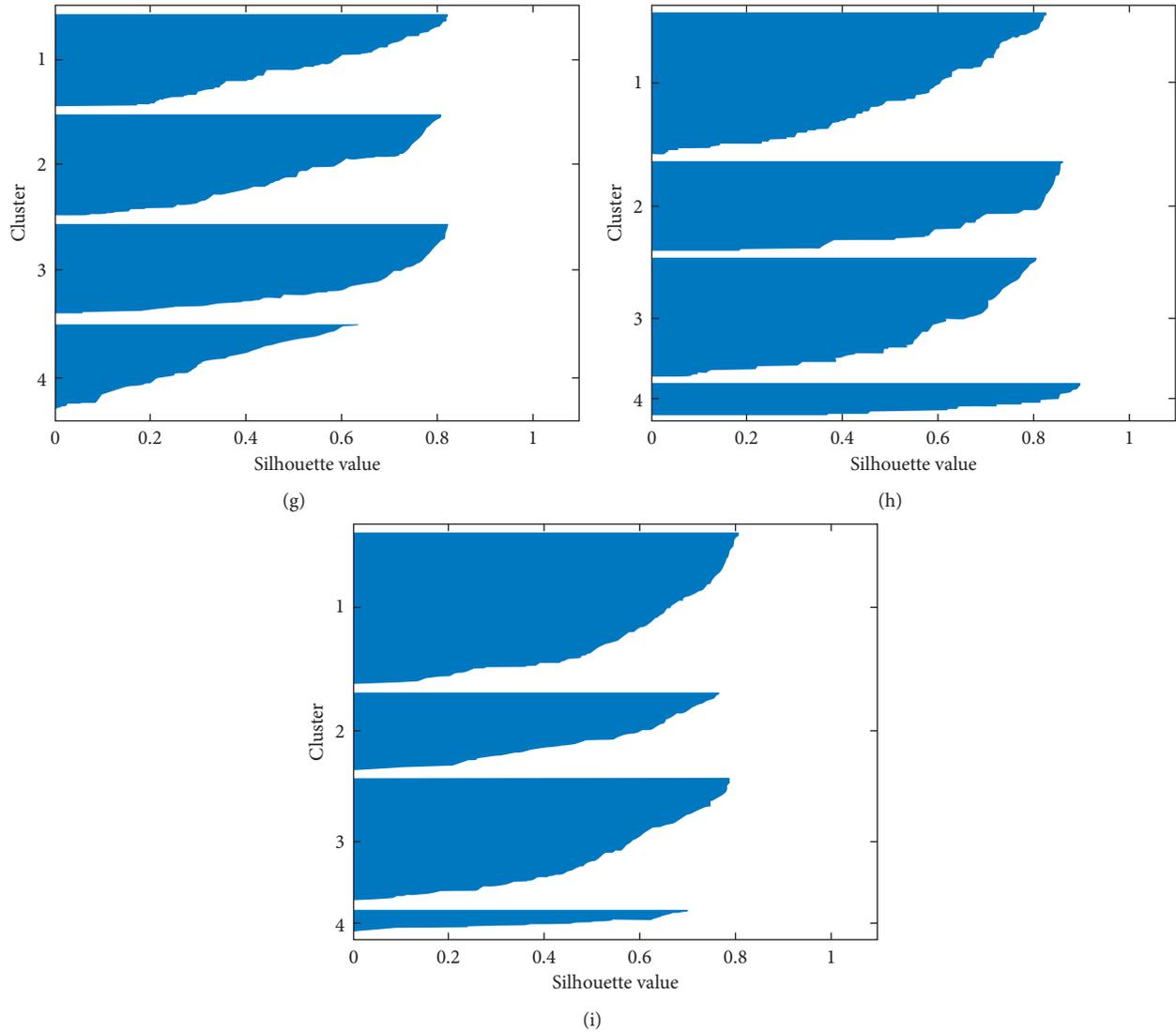


FIGURE 5: (a) Silhouette value of the expressway data in Jiangsu Province. Silhouette values of national and provincial highways data in (b) quarter 1 of 2018, (c) quarter 2 of 2018, (d) quarter 3 of 2018, (e) quarter 4 of 2018, (f) quarter 1 of 2019, (g) quarter 2 of 2019, (h) quarter 3 of 2019, (i) quarter 4 of 2019.

Figures 5(a)–5(i) show the silhouette values of the expressway data and the national and provincial highways data. A silhouette value is a value between -1 and 1 . When the cluster cohesion degree is equal to the separation degree, the silhouette value is 0 . The closer the silhouette value is to 1 , the better the cohesion degree and separation degree are. If the silhouette value is closer to -1 , the point should not be classified into a group. The figures show that all the silhouette values are above 0 , and most of them are above 0.8 , indicating that the result of the clustering is good.

The prediction results of BP, GRNN, and WNN of expressway data are shown in Figures 6(a)–6(c). The prediction results of monthly, quarterly, and annual BP, GRNN, and WNN of national and provincial highway data in Liyang city in 2018 are shown in Figures 7(a)–7(c), 8(a)–8(c), and 9(a)–9(c), respectively.

Figures 6(a), 7(a), 8(a), and 9(a) show that the predicted overload rate fits the actual overload rate well when using BPNN with the fitting curves “ $y = 0.99x + 0.0068$,” “ $y = 0.97x + 0.0088$,” “ $y = 0.96x + 0.011$,” and “ $y = 0.97x + 0.0021$,” respectively. The “ R^2 ” values are always larger than 0.95 ; this indicates the validity of the BPNN model. Figures 6(b), 7(b), 8(b), and 9(b) show the fitting degree of the predicted overload rate and the actual overload rate when using GRNN with the fitting curves “ $y = 0.53x + 0.045$,” “ $y = 0.49x + 0.036$,” “ $y = 0.83x + 0.0236$,” and “ $y = 0.86x + 0.0126$,” respectively. The “ R^2 ” values are approximately 0.5 , which does not correspond to a fitting degree as good as that exhibited by the results obtained using BPNN. Figures 6(c), 7(c), 8(c), and 9(c) show the fitting degree of the predicted overload rate and the actual overload rate when using WNN with the fitting curves “ $y = 0.16x + 0.151$,” “ $y = 0.18x + 0.1748$,”

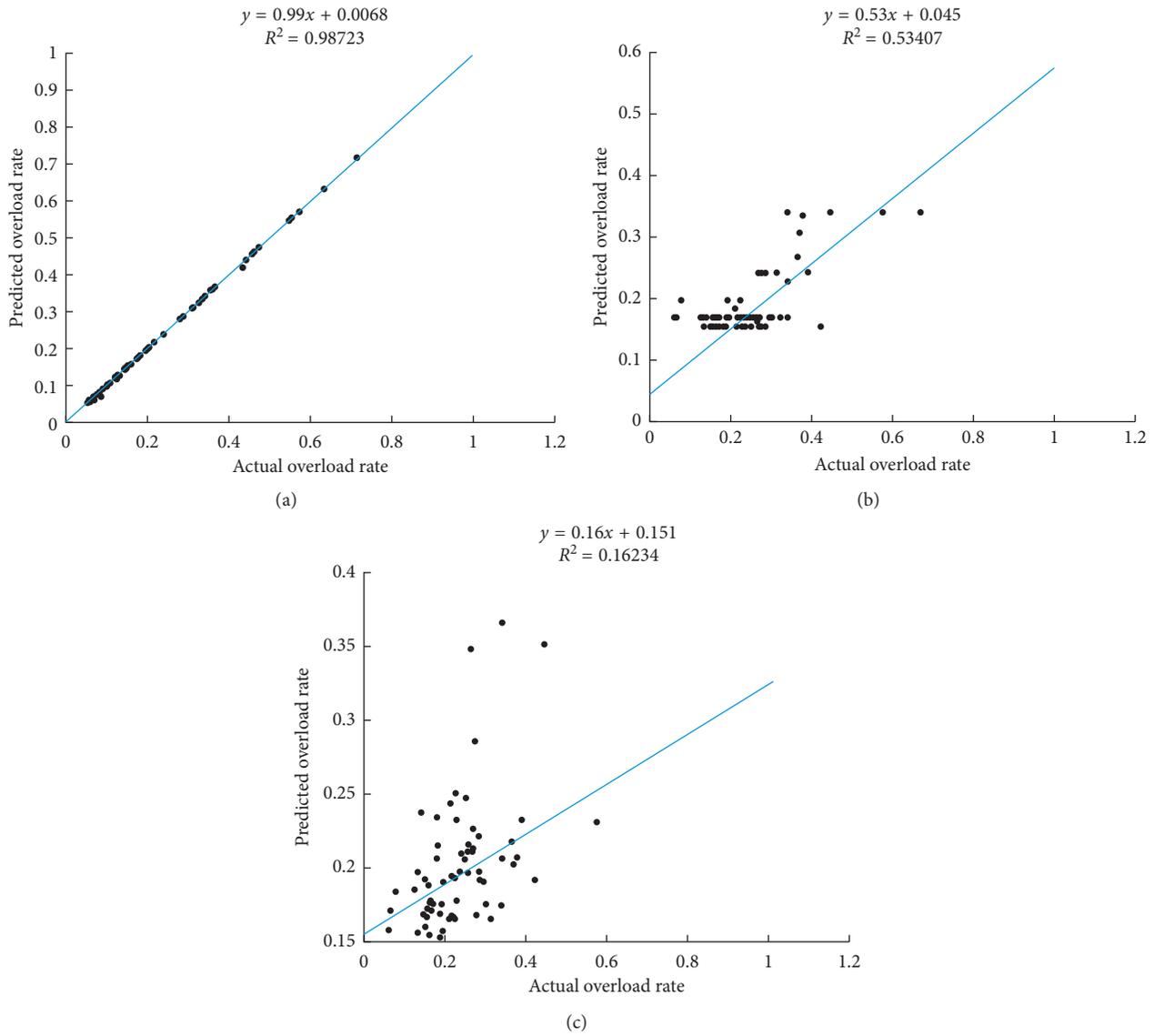


FIGURE 6: Prediction results of the (a) BPNN, (b) GRNN, and (c) WNN in Jiangsu province.

“ $y = 0.17x + 0.198$,” and “ $y = 0.02x + 0.217$,” respectively. The “ R^2 ” values are always less than 0.2, which corresponds to the lowest fitting degree among those of the three methods.

Tables 4–6 present the prediction results of the BPNN, GRNN, and WNN, respectively, indicating the data and results in the process of machine learning. All data are divided into training data and test data in a ratio of 10:1 and run according to the process described in Section 3.2. The prediction results of BPNN exhibit a high degree of fitting of all the validation data, training data, and test data, which also shows that the model performed well without overfitting. GRNN prediction required a long runtime owing to its inclusion of cross-validation and cycle training. WNN could not accurately predict the overload rate, and there was a gap between its prediction value and the actual value. It is speculated that WNN is typically used for short-term time-series prediction, whereas quarterly data and annual data prediction belong to long-term time-series prediction; thus,

the prediction results are not ideal, regardless of the overload rate and overload trend. Although the overload rate obtained from monthly data was not sufficiently accurate, it could reflect the overload trend. Furthermore, the WNN omitted attributes that were useful for prediction by BPNN, such as “License plate,” “Lane,” “Gross weight of vehicles and goods,” “Axle number,” and “Speed,” and only used “Overload rate” and short-term time series for the prediction. Thus, the prediction results reveal the weak correlation between the overload rate and short-term time.

5. Discussion

According to the clustering results of the expressways in Jiangsu Province, 6-axle vehicles weighing above 49 t and entering the expressways between 23:00–4:00 and 12:00–14:00 are more likely to exhibit overloaded behavior, and the overload rate of these vehicles is higher than that of others

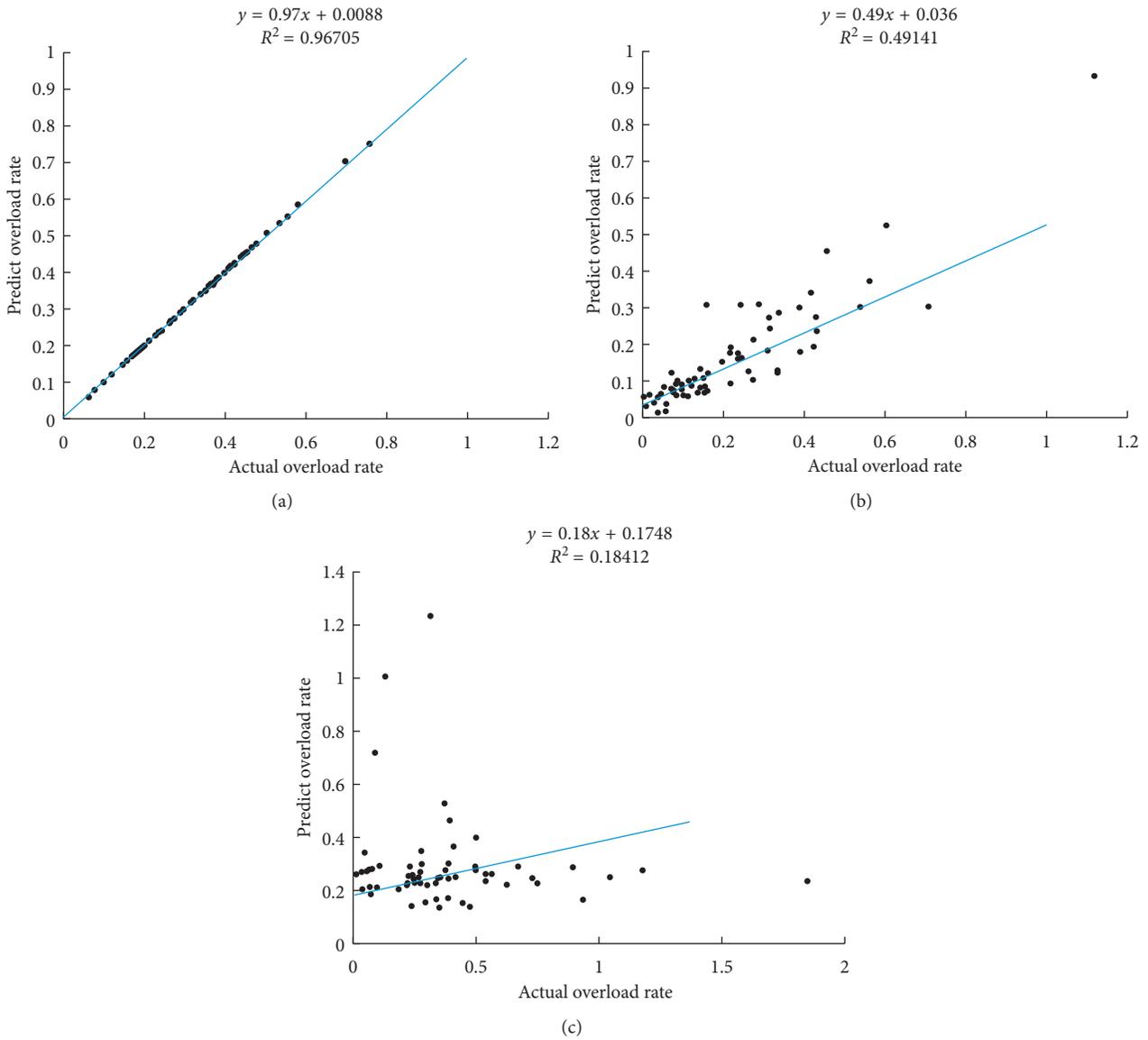


FIGURE 7: Prediction results of the (a) BPNN, (b) GRNN, and (c) WNN in Liyang city in 2018 (month).

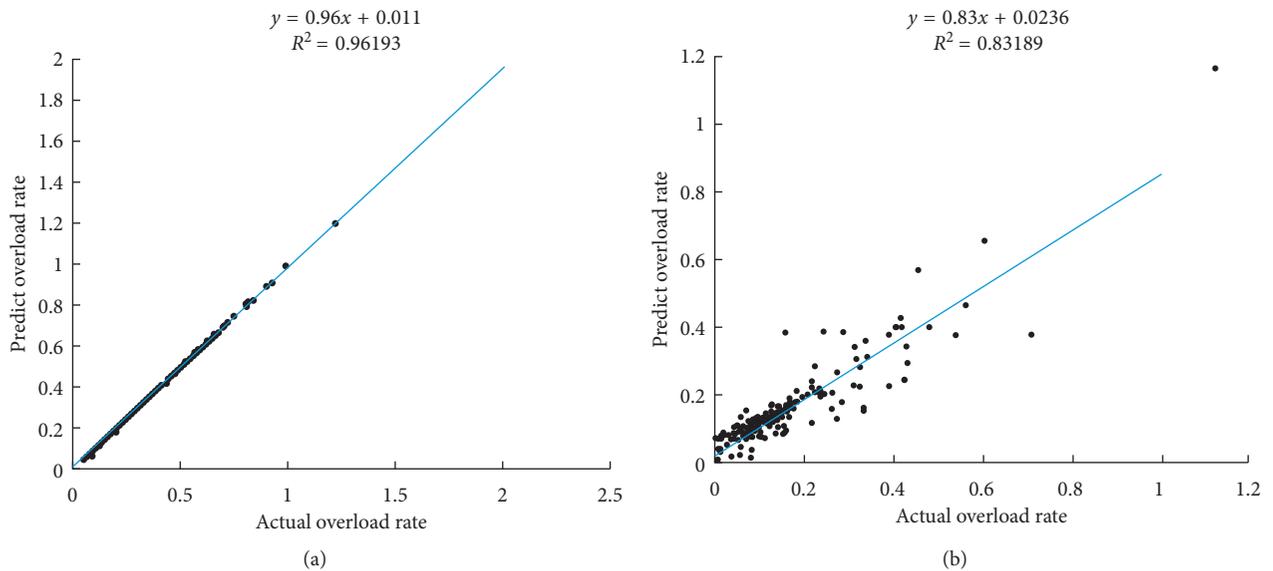


FIGURE 8: Continued.

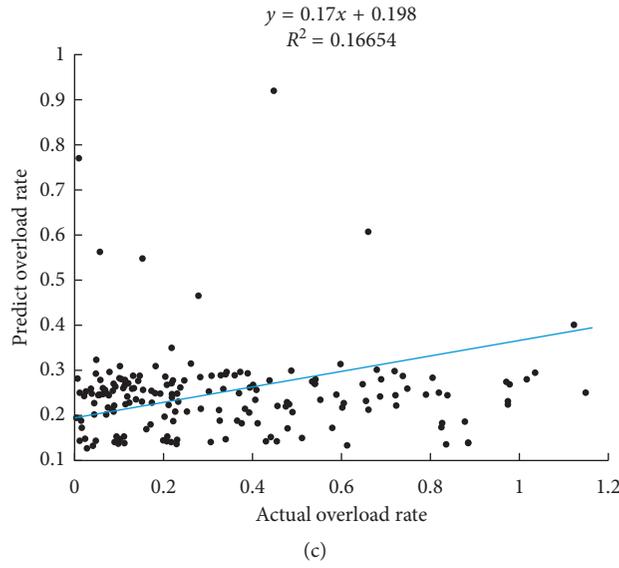


FIGURE 8: Prediction results of the (a) BPNN, (b) GRNN, and (c) WNN in Liyang city in 2018 (quarter).

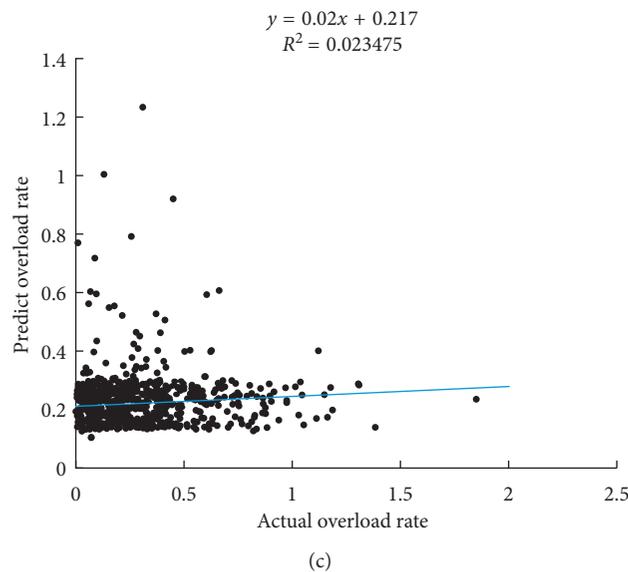
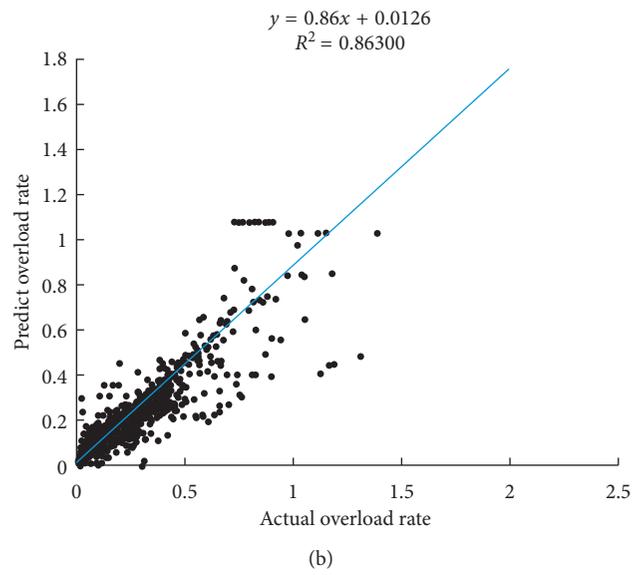
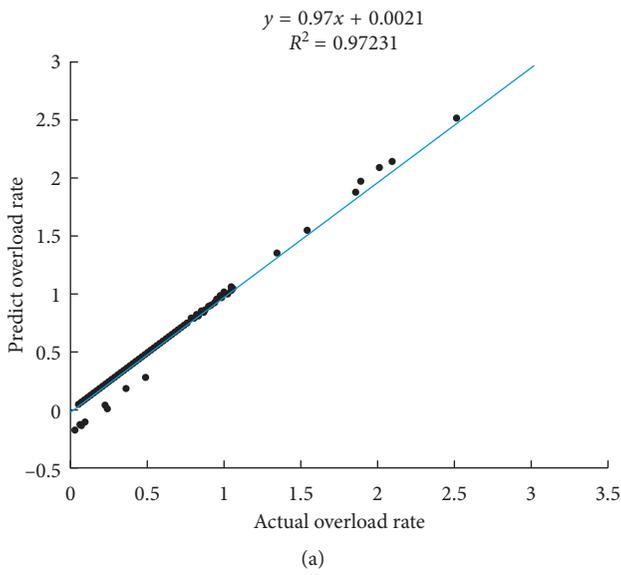


FIGURE 9: Prediction results of the (a) BPNN, (b) GRNN, and (c) WNN in Liyang city in 2018 (year).

TABLE 4: Prediction results of BPNN.

| Data name | Highway data | National and provincial highway data (month) | National and provincial highway data (quarter) | National and provincial highway data (year) |
|------------------------|--------------|--|--|---|
| Number of hidden nodes | 16 | 3 | 8 | 6 |
| Iteration times | 3 | 3 | 5 | 6 |
| Error descent gradient | 0.008 | 0.005 | 0.019 | 0.034 |
| R^2 | 0.987 | 0.967 | 0.962 | 0.972 |
| MSE ($\%^2$) | 0.090 | 0.055 | 0.093 | 0.059 |
| RMSE (%) | 0.960 | 2.960 | 3.060 | 1.460 |
| ME (%) | 2.700 | -0.150 | 38.750 | -3.200 |
| MAE (%) | 5.200 | 2.120 | 44.330 | 3.840 |
| MAPE (%) | 0.089 | 0.124 | 0.116 | 1.046 |
| RMSPE (%) | 0.059 | 0.026 | 0.058 | 0.894 |
| Running time (s) | 2.016 | 1.938 | 0.001 | 2.953 |

TABLE 5: Prediction results of GRNN.

| Data name | Highway data | National and provincial highway data (month) | National and provincial highway data (quarter) | National and provincial highway data (year) |
|------------------|--------------|--|--|---|
| SPREAD | 1.6 | 0.7 | 1.1 | 0.9 |
| R^2 | 0.534 | 0.491 | 0.832 | 0.863 |
| MSE ($\%^2$) | 0.240 | 0.037 | 0.210 | 0.100 |
| RMSE (%) | 49.860 | 74.132 | 0.560 | 0.288 |
| ME (%) | 70.610 | 86.100 | 7.480 | 5.370 |
| MAE (%) | -10.150 | -12.150 | -3.020 | -8.960 |
| MAPE (%) | 0.542 | 0.622 | 0.048 | 0.102 |
| RMSPE (%) | 0.640 | 0.684 | 0.403 | 0.871 |
| Running time (s) | 5.156 | 7.016 | 12.313 | 19.031 |

TABLE 6: Prediction results of WNN.

| Data name | Highway data | National and provincial highway data (month) | National and provincial highway data (quarter) | National and provincial highway data (year) |
|------------------|--------------|--|--|---|
| R^2 | 0.162 | 0.184 | 0.169 | 0.020 |
| MSE ($\%^2$) | 18.250 | 48.570 | 41.620 | 59.050 |
| RMSE (%) | 0.810 | 1.110 | 0.930 | 7.684 |
| ME (%) | 9.000 | 10.520 | 9.640 | 12.524 |
| MAE (%) | 5.480 | -3.790 | -5.710 | -9.324 |
| MAPE (%) | 0.079 | 0.069 | 0.073 | 9.413 |
| RMSPE (%) | 0.510 | 0.352 | 0.388 | 3.443 |
| Running time (s) | 0.484 | 1.219 | 0.563 | 1.656 |

during this period. Most vehicles enter the expressway at 05:00 and 17:00–22:00, among which 6-axle vehicles weighing above 49 t are more likely to be overloaded. This means that the inspection workers should constantly pay more attention to the vehicles weighing above 49 t during this period. Further, more resources, such as manpower, should be allocated between 23:00–04:00 and 12:00–14:00.

Comparing the clustering results of 2018 and 2019 in Table 3, certain changes were observed in the group characteristics over the past two years. (1) The difference in the peak hour between 2018 and 2019 is approximately 2 h, and the difference in the trough hour between 2018 and 2019 is approximately 2 h. This is speculated to be because the working hours of overload management in 2018 and 2019

had changed. (2) In 2018 and 2019, the overload rate in the first quarter was lower than that in the other three quarters. Compared with that in 2018, the overload rate in 2019 was reduced. This is speculated to be because Changzhou performed several joint actions of overloading management in 2019, because of which the overloading governance had been strengthened. (3) Between 2018 and 2019, there were no evident changes in the “Gross weight of vehicles and goods” and “Axle number.”

Table 7 summarizes the clustering results of national highways in 2018 and 2019. In 2018, most overloaded vehicles entered the highways between 00:00–08:00 and between 16:00–20:00, and vehicles with higher overload rates were concentrated between 21:00–23:00 and between 02:

TABLE 7: Summary of the laws of clustering results of the national and provincial highways when $k = 4$.

| Year | Quarter | Most overloaded vehicles | | | Vehicles with high overload rate | | |
|------|---------|--------------------------|-----------------------------------|-------------|----------------------------------|-----------------------------------|-------------|
| | | Entry time | Gross weight of vehicle and goods | Axle number | Entry time | Gross weight of vehicle and goods | Axle number |
| 2018 | 1 | 0:00–8:00 | Over 49 t | 6 | 2:00–3:00 | Over 49 t | 6 |
| | 2 | 0:00–8:00 | Over 49 t | 6 | 21:00–23:00 | Over 49 t | 6 |
| | 3 | 0:00–8:00 | Over 49 t | 6 | 21:00–23:00 | Over 49 t | 6 |
| | 4 | 0:00–8:00 | Over 49 t | 6 | 23:00 | Over 49 t | 6 |
| 2019 | 1 | 4:00–13:00 | Over 49 t | 6 | 0:00–3:00 | Over 49 t | 6 |
| | 2 | 0:00–15:00 | Over 49 t | 6 | 0:00–3:00 | Over 49 t | 6 |
| | 3 | 4:00–5:00 | Over 49 t | 6 | 2:00–3:00 | Over 49 t | 6 |
| | 4 | 8:00–15:00 | Over 49 t | 6 | 0:00–4:00 | Over 49 t | 6 |

00–03:00. In 2019, most overloaded vehicles entered the highway from 04:00 to 15:00, and vehicles with a high overload rate were concentrated between 00:00–03:00. In 2018–2019, most overloaded vehicles were 6-axle vehicles weighing above 49 t, and the vehicles with high overload rates were also 6-axle vehicles weighing above 49 t.

Comparing the prediction results of the BPNN, GRNN, and WNN, the common evaluation indexes of the regression prediction model include mean-squared error (MSE), the deterministic coefficient R^2 , and goodness of fit. The numerical value of the deterministic coefficient R^2 reflects the goodness of fit. Therefore, we selected the deterministic coefficient, MSE, and the running time t of the entire code as the model performance criteria. In addition, this study used the root-mean-squared error (RMSE), mean error (ME), mean absolute error (MAE), mean absolute percentage error (MAPE), and root-mean-squared percentage error (RMSPE) as five different performance criteria Table 8. The averages of $(1 - R^2)$, MSE, RMSE, ME, MAE, MAPE, and RMSPE are calculated Table 9 to eliminate the prediction error of the BPNN, GRNN, and WNN [30].

The average values and all the seven performance criteria show that the prediction of BPNN is better than that of GRNN and WNN. Its performance is more accurate, especially for the “highway data” and “national and provincial highway data (month).” This is because the amount of data increases with time. Although the overall fitting degree is high, the individual prediction errors are greater, and certain performance criteria also increase with the increase in the number of errors. To clearly evaluate the adequacy of this performance, the performance of BPNN in this study was compared with that in existing studies. For example, Kumar used an ANN to predict the short-term traffic flow of non-urban expressways. In this study, when the amount of data was 160, the fitting degree was 0.9984, and when the data amount was 480, the fitting degree was 0.9988 [31]. The fitting degree of the BPNN in this study was larger than 0.90 and was nearly 0.99, which shows that the BPNN in this study has a very high degree of fitting and the model performs well.

To detect the defects of GRNN and WNN, the performances of GRNN and WNN were also compared with those of other studies. Zhang and Zhang used the spatial

relationship of traffic flow near U.S. Highway and developed two multivariate forecasting approaches, in which the GRNN was also used [22]. In this research, 91.32% of the forecasting error was less than 20%. Tang et al. used WNN to predict the short-term traffic flow with a fitting degree of 0.9453 and used 230 datasets in modeling [31]. They used the WNN to predict short-term traffic flow with a fitting degree of 0.936, 864 datasets for modeling, and running time of 0.995. Compared with that of Tang et al. [32], the fitting degree of WNN in this research was extremely low. However, when the datasets were the same, the running time was similar to that in other studies. The reason is that although the subjects of the study are different, the traffic flow has a strong correlation with time. In terms of overload, compared with other attributes such as “Time,” “License plate,” “Lane,” “Gross weight of vehicles and goods,” “Axle number,” and “Speed,” the overload rate has a weak correlation with time.

By comparison, the combined results of Tables 8 and 9 are as follows.

- (1) The prediction performance of BPNN is superior to that of GRNN and WNN. Given sufficient data, BPNN is a better forecasting method with a high fitting degree, minimum MSE, and short running time. BPNN performs well in hourly highway data prediction and annual overload rate prediction, and the prediction results of BPNN can help determine the centralized overload management time.
- (2) With the increasing amount of data, the fitting degree of GRNN increases, and the running time also increases. In this study, the method of adding a “for loop” to find the best SPREAD greatly increases the running time. Compared with other networks, GRNN had the longest running time.
- (3) Compared with BPNN and GRNN, WNN has the shortest running time in overload rate prediction. Although the fitting degree of WNN is low, it only requires an overload rate for prediction. It can reflect the future trends of the overload rate in the monthly data prediction of overload, and its prediction results can help determine the centralized overload management time when the data are insufficient.

TABLE 8: Results of the NN.

| | Neural network | Highway data (hour) | National and provincial highway data (month) | National and provincial highway data (quarter) | National and provincial highway data (year) |
|-----------------------|----------------|---------------------|--|--|---|
| Running time (s) | BPNN | 2.016 | 1.938 | 2.125 | 2.953 |
| | GRNN | 5.156 | 7.016 | 12.313 | 19.031 |
| | WNN | 0.484 | 1.219 | 0.563 | 1.656 |
| R^2 | BPNN | 0.987 | 0.967 | 0.962 | 0.972 |
| | GRNN | 0.534 | 0.558 | 0.832 | 0.863 |
| | WNN | 0.162 | 0.184 | 0.169 | 0.021 |
| MSE (% ²) | BPNN | 0.960 | 2.960 | 3.060 | 1.460 |
| | GRNN | 49.860 | 74.130 | 0.560 | 0.290 |
| | WNN | 0.810 | 1.110 | 0.930 | 7.684 |
| RMSE (%) | BPNN | 2.700 | -0.150 | 3.880 | -3.200 |
| | GRNN | 70.610 | 86.100 | 7.480 | 5.370 |
| | WNN | 9.000 | 10.520 | 9.640 | 12.524 |
| ME (%) | BPNN | 5.200 | 2.120 | 4.430 | 3.840 |
| | GRNN | -10.150 | -12.150 | -3.020 | -8.960 |
| | WNN | 5.480 | -3.790 | -5.710 | -9.324 |
| MAE (%) | BPNN | 0.090 | 0.060 | 0.090 | 0.060 |
| | GRNN | 0.240 | 0.040 | 0.210 | 0.100 |
| | WNN | 18.250 | 48.570 | 41.620 | 59.050 |
| MAPE (%) | BPNN | 0.089 | 0.124 | 0.116 | 1.046 |
| | GRNN | 0.542 | 0.622 | 0.048 | 0.102 |
| | WNN | 0.079 | 0.069 | 0.073 | 9.413 |
| RMSPE (%) | BPNN | 0.059 | 0.026 | 0.058 | 0.894 |
| | GRNN | 0.640 | 0.684 | 0.403 | 0.871 |
| | WNN | 0.510 | 0.352 | 0.388 | 3.443 |

TABLE 9: Average of seven performance criteria.

| | Highway data (hour) | National and provincial highway data (month) | National and provincial highway data (quarter) | National and provincial highway data (year) |
|------|---------------------|--|--|---|
| BPNN | 0.175 | 0.167 | 0.179 | 0.419 |
| GRNN | 0.403 | 0.478 | 0.191 | 0.258 |
| WNN | 0.155 | 0.167 | 0.156 | 29.214 |

6. Conclusions

In this study, we used machine learning to establish a model for highway overrun and overload, considering Jiangsu Province as an example. The characteristics of overloading were summarized by clustering the historical data of overloading, and a forecasting model of overloading with a high fitting degree based on BPNN was obtained. We provided information with potential value for expressway network management departments through data mining. This information could help these management departments allocate resources reasonably and optimize the information utilization rate. The conclusions of this study can be grouped into two aspects: daily overload management based on cluster analysis and immediate overload management based on NNs.

The overload characteristics based on cluster analysis, human resources, and other related resources should be allocated rationally. Targeted monitoring should be performed for overload control and management level, and the governance efficiency should be enhanced using information technology. Currently, numerous large loopholes remain in time and space in law enforcement inspection, and law

enforcement personnel and equipment are limited. Therefore, the thorough use of historical overload data is necessary to satisfy the demand for taking severe measures on overloaded vehicles according to the law of overload characteristics. For example, based on the conclusion of this study, we can focus on monitoring 6-axle vehicles weighing above 49 t and entering an expressway between 23:00–04:00 and 12:00–14:00, and 6-axle vehicles weighing above 49 t entering the national provincial highway of a city between 21:00–23:00 and 02:00–03:00. The allocation of manpower and other related resources should be increased in this period, and related resources should be saved during other periods. Based on the overloading characteristics, different types of vehicles in different periods must be thoroughly investigated to improve the management efficiency of the transportation departments. For example, special attention must be paid to 2-axle vehicles weighing 18–27 t and entering a highway between 06:00–07:00 and 15:00–16:00 and to 4-axle vehicles weighing 37–43 t and 43–49 t and entering a highway between 08:00–11:00. Concurrently, the traffic management department can reasonably adjust the working hours of staff in different seasons. For example, the working hours of staff in the first quarter can

be reduced according to the conclusion of this study to effectively enhance the work efficiency.

Based on the prediction model obtained by the NN, we can predict the overload rate and overload trend, develop auxiliary programs or software, and improve the existing law enforcement system. The information resources of administrative departments in fundamental areas have been in a closed state. The information resources of administrative law enforcement and administrative law enforcement supervision are integrated and shared, and real-time information, such as that on personnel and equipment, is added based on the existing system data to improve the overloaded vehicles database. Based on the BPNN model used in this study, the real-time information on personnel and equipment can be integrated to develop auxiliary programs or software to realize online inquiry by law enforcement departments and online supervision of law enforcement activities. Thus, all departments can inquire about the current overload rate and trend at any time, especially in centralized law management, and can reasonably determine the time and place of centralized law management with the help of prediction results and integrated information of manpower and equipment. By checking changes in manpower, equipment, and other information, we can determine real-time actions for various departments, ensure business linkage, promote the linkage work of administrative law enforcement, further promote the governance of overloading, and improve the efficiency of law enforcement systems.

The overload data mining of expressways in this research is limited to expressways in Jiangsu Province, and the overload data mining of national and provincial highways is limited to Liyang city, which is not universal. Furthermore, the spatial attributes, namely, "Entrance site" and "Exit site" were not analyzed. Our follow-up research will focus on a considerably comprehensive feature analysis and mining of the spatiotemporal law of overloading. Using geographic information system (GIS), global positioning system, spatial statistical models, and other advanced technologies, we aim to explore the spatiotemporal law of overloaded vehicles in highways, visualize it with the help of GIS, predict the destination area of overloading, overloading transportation volume, and overloading transportation routes, and conduct more in-depth analysis to reveal the existence of overloading of freight vehicles in highway networks in different airspaces.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Number: 71573037) and the

Priority Academic Program Development of Jiangsu Higher Education Institutions in China (Grant Number: 1105007002).

References

- [1] C. L. Dos Santos Romeiro Júnior, L. A. Teixeira Brito, L. F. Heller et al., "Impact on pavement deterioration due to overload vehicle regulation in Brazil," *Transportation Research Procedia*, vol. 45, pp. 842–849, 2020.
- [2] K. Bucshházy, E. Matuchová, R. Zůvala, P. Moravcová, M. Kostíková, and R. Mikulec, "Human factors contributing to the road traffic accident occurrence," *Transportation Research Procedia*, vol. 45, pp. 555–561, 2020.
- [3] Y. Dai, J. Guo, L. Yang, and W. You, "A new approach of intelligent physical health evaluation based on GRNN and BPNN by using a wearable smart bracelet system," *Procedia Computer Science*, vol. 147, pp. 519–527, 2019.
- [4] W. Han, J. Wu, C. S. Cai, and S. Chen, "Characteristics and dynamic impact of overloaded extra heavy trucks on typical highway bridges," *Journal of Bridge Engineering*, vol. 20, no. 2, Article ID 5014011, 2015.
- [5] B. Li and X. Wang, "Analysis of the freeway overrun limit based on system clustering," *Modern Computer*, vol. 59, pp. 39–42, 2018.
- [6] S. Ryu, A. Chen, and K. Choi, "A modified gradient projection algorithm for solving the elastic demand traffic assignment problem," *Computers & Operations Research*, vol. 47, pp. 61–71, 2014.
- [7] M. Cheng, C. Li, L. Wu, Li Zhang, and X. Cheng, "Analysis on the detour behavior of freight vehicles based on expressway toll data," *Journal of Highway and Transportation Research and Development*, vol. 15, pp. 299–301, 2019.
- [8] R. Zhang, *Analysis on the Influence of Overload Control on Road Capacity*, Transpo World, Beijing, China, 2019.
- [9] K. Mirsad, I. Zoran, D. Slavko, and P. Ivo, "Modelling truck weigh stations' locations based on truck traffic flow and overweight violation: a case study in Bosnia and Herzegovina," *Promet*, vol. 30, no. 2, pp. 163–171, 2018.
- [10] G. Kostopoulos, K. Neureiter, D. Papatou, M. Tscheligi, and C. Chrysoulas, "ProMe: a mentoring platform for older adults using machine learning techniques for supporting the "live and learn" concept," *Mobile Information Systems*, vol. 2018, Article ID 9723268, 8 pages, 2018.
- [11] Ç. Muhammed Yasin and Ü. Ahmet, "An estimation of transport energy demand in Turkey via artificial neural networks," *Promet*, vol. 31, no. 2, pp. 151–161, 2019.
- [12] F. Li, C.-H. Chen, P. Zheng et al., "An explorative context-aware machine learning approach to reducing human fatigue risk of traffic control operators," *Safety Science*, vol. 125, Article ID 104665, 2020.
- [13] M. Mohammed Abdulhafedh, H. Mazen Radhe, and D. Shankar, "Analysis of performance limitations in optical wireless communication system due to laser source impairment (RIN) and background noise," *Advances in Intelligent Systems and Computing*, vol. 452, pp. 521–531, 2016.
- [14] H. Yu and X. Li, "On the chaos analysis and prediction of aircraft accidents based on multi-timescales," *Physica A: Statistical Mechanics and Its Applications*, vol. 534, Article ID 120828, 2019.
- [15] M. Sangare, S. Gupta, S. Bouzeffrane, S. Banerjee, and P. Muhlethaler, "Exploring the forecasting approach for road accidents: analytical measures with hybrid machine learning,"

- Expert Systems with Applications*, 2020, In press, Article ID 113855.
- [16] Y. Lin and R. Li, "Real-time traffic accidents post-impact prediction: based on crowdsourcing data," *Accident Analysis & Prevention*, vol. 145, Article ID 105696, 2020.
- [17] S. Onder, "Evaluation of occupational injuries with lost days among opencast coal mine workers through logistic regression models," *Safety Science*, vol. 59, pp. 86–92, 2013.
- [18] W. Jiang, J. Josse, and M. Lavielle, "Logistic regression with missing covariates-parameter estimation, model selection and prediction within a joint-modeling framework," *Computational Statistics & Data Analysis*, vol. 145, Article ID 106907, 2020.
- [19] S. S. Yassin and Pooja, "Road accident prediction and model interpretation using a hybrid K-means and random forest algorithm approach," *SN Applied Sciences*, vol. 2, Article ID 1576, 2020.
- [20] E. Zunic, A. Djedovic, and D. Donko, "Cluster-based analysis and time-series prediction model for reducing the number of traffic accidents," in *Proceedings of the International Symposium ELMAR*, pp. 25–29, Zadar, Croatia, September 2017.
- [21] Y.-F. Wang, L.-T. Wang, J.-C. Jiang, J. Wang, and Z.-L. Yang, "Modelling ship collision risk based on the statistical analysis of historical data: a case study in Hong Kong waters," *Ocean Engineering*, vol. 197, Article ID 106869, 2020.
- [22] Y. Zhang and Y. Zhang, "A comparative study of three multivariate short-term freeway traffic flow forecasting methods with missing data," *Journal of Intelligent Transportation Systems*, vol. 20, no. 3, pp. 205–218, 2016.
- [23] Q. Hou, J. Leng, G. Ma, W. Liu, and Y. Cheng, "An adaptive hybrid model for short-term urban traffic flow prediction," *Physica A: Statistical Mechanics and Its Applications*, vol. 527, Article ID 121065, 2019.
- [24] J.-C. Huang, K.-M. Ko, M.-H. Shu, and B.-M. Hsu, "Application and comparison of several machine learning algorithms and their integration models in regression problems," *Neural Computing and Applications*, vol. 32, no. 10, pp. 5461–5469, 2020.
- [25] T. Arsan and M. M. N. Hameez, "A clustering-based approach for improving the accuracy of UWB sensor-based indoor positioning system," *Mobile Information Systems*, vol. 2019, Article ID 6372073, 13 pages, 2019.
- [26] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms," *Mobile Information Systems*, vol. 2018, Article ID 3860146, 21 pages, 2018.
- [27] S. Amin, "Backpropagation—artificial neural network (BP-ANN): understanding gender characteristics of older driver accidents in west midlands of United Kingdom," *Safety Science*, vol. 122, Article ID 104539, 2020.
- [28] L. Xiao, J. Wang, R. Hou, and J. Wu, "A combined model based on data pre-analysis and weight coefficients optimization for electrical load forecasting," *Energy*, vol. 82, pp. 524–549, 2015.
- [29] A. Miglani and N. Kumar, "Deep learning models for traffic flow prediction in autonomous vehicles: a review, solutions, and challenges," *Vehicular Communications*, vol. 20, Article ID 100184, 2019.
- [30] W. Wang, R. Tang, C. Li, P. Liu, and L. Luo, "A BP neural network model optimized by mind evolutionary algorithm for predicting the ocean wave heights," *Ocean Engineering*, vol. 162, pp. 98–107, 2018.
- [31] K. Kumar, M. Parida, and V. K. Katiyar, "Short term traffic flow prediction for a non urban highway using artificial neural network," *Procedia—Social and Behavioral Sciences*, vol. 104, pp. 755–764, 2013.
- [32] R. Tang, Q. Chen, and X. Lei, "GQPSO-WNN short-term traffic flow forecasting based on phase space reconstruction," *Computer Applications and Software*, vol. 36, no. 7, pp. 311–316, 2019.