WILEY | Hindawi

Research Article

How to Use Advanced Fleet Management System to Promote Energy Saving in Transportation: A Survey of Drivers' Awareness of Fuel-Saving Factors

Changjian Zhang, Jie He, Chunguang Bai ⓘ, Xintong Yan, Jian Gong, and Hao Zhang

School of Transportation, Southeast University, Nanjing 210018, China

Correspondence should be addressed to Chunguang Bai; baichg@seu.edu.cn

Despite the broad application of advanced fleet management systems (FMSs) in third-party logistics (3PL) companies, there is a marginally limited understanding of how to employ them to enhance transport energy efficiency. In a case study of a Chinese 3PL company, this paper analyzed data obtained from the online FMS to assess drivers' awareness of fuel-saving factors. A questionnaire was primarily designed to investigate the drivers' awareness of fuel-saving factors based on the reliability and validity test. Then, Extreme Gradient Boosting (XGBoost), a machine learning algorithm, was utilized to explore the intrinsic impacts of various factors on fuel consumption with the outputs providing the evaluation basis for individual awareness of the drivers. The results show a significant deviation in the driver's awareness of fuel-saving factors, among which the three indicators of engine speed, idling condition, and rolling without engine load are seriously underestimated, while the indicators related to the environment are seriously overestimated due to social expectations. In addition, the average speed was found to be the most important fuel-saving indicator besides the load. Based on these findings, this paper recommends that the 3PL companies choose a route with more freeways when planning, and the mileage should be controlled within 800 km as far as possible.

1. Introduction

So far, fossil fuels remain the primary source of energy for freight movements. Freight transportation is thereby often considered as one of the most difficult economic activities to decarbonize. Among all transportation modes, road freight has been under particular observation given the rapid growth of future demand. As reported by International Energy Agency [1] and International Transport Forum [2], road freight transport consumed roughly 30% of the energy produced and 53% of the total international trade-related carbon emissions. Such an international background puts forward strict requirements for developing low-carbon transportation of the third-party logistics (3PL) company.

There is a close relationship between logistics and low-carbon transportation. On the one hand, the 3PL company, as one of the most important implementations and management subjects of freight transportation, plays a crucial role in energy-saving and low-carbon development [3]. On the other hand, sustainability is becoming a critical success factor in cost reduction and increased sales in the logistics service industry [4, 5]. Therefore, reducing energy consumption has become the inevitable trend for the 3PL company. Information and Communication Technologies (ICTs) have provided new low-cost and easy-to-use tools for 3PL companies to control transportation activities. On-board units are seen as a key technology in monitoring the success of efficiency measures, because they could register fuel consumption and couple it with other vehicle parameters, such as weight load, speed, engine revolution speed, altitude, and mileage [6]. Although on-board units have shown large room for reducing fuel consumption, limited by the development of science and technology, these smart systems and devices are not widely used in Chinese 3PL

companies. Even if on-board units are installed on trucks, most small and medium-sized enterprises do not well understand how to analyze and apply data obtained from them to lower fuel consumption except a few large-scale enterprises such as Jingdong, Shunfeng, and Debang logistics.

In this paper, we constructed a comprehensive methodologic framework to show how the advanced on-board measurement system could be used to mine these internal data relating to fuel consumption, so as to contribute to the decarbonization of logistics transportation. We designed two paths to analyze the excessive fuel consumption of heavy-duty diesel trucks (HDDTs). The first one was to design a questionnaire for the fleet of data providers, and investigate in detail the cognitive level of each driver on the contributing factors of fuel consumption in the real world. The alternative one was to use the machine learning approach to investigate how these contributing factors determine fuel consumption under real driving conditions. Finally, the driver's subjective cognitive level was evaluated based on the results of data mining. Hence, 3PL companies can formulate more targeted eco-driving training strategies according to the drivers' cognitive degree and features, so as to improve the macro-control of the fleet on energy saving and environmental protection.

2. Literature Review

2.1. Application of Advanced On-Board Systems. The advanced on-board fleet management system (FMS) uses intelligent sensing equipment, a mobile communication network, and satellite positioning technology to feedback the position and state of the vehicle to the fleet management personnel. As shown in Figure 1, it consists of on-board units, servers, and clients. The operation data of vehicles can be collected in real-time and feedback to managers through the mobile network. At present, a variety of similar equipment or system has been put into practical application on the market to support scientific research, strategic planning, management policies, etc. Bousonville et al. [7] collected naturalistic driving data (e.g., total weight, truck model, average speed, engine power, height meters, and weather) using on-board telematics systems to analyze and forecast fuel consumption of trucks. Madhusudhanan et al. [8] collected data from a supermarket's FMS interface to model the transport performance and cost of biomethane trucks. The loggers can acquire the real-time data via 3G/4G network, and the uploaded data can be automatically downloaded and stored on a server. Xu et al. [9] utilized the Internet System of Vehicles by the Shaanxi Automotive Group to obtain the truck operational data to calibrate fuel consumption estimation models. Moreover, some scholars did not use the company's FMS but used some similar intelligent equipment to collect driving data and then to carry out fuel consumption analysis of vehicles [10, 11].

The biggest advantage of the advanced on-board FMS is that it could be used to acquire large-scale, real-world driving data. However, to the best of our knowledge, most of the previous studies are only for the purpose of theoretical research and are not combined with the realistic situation of the company that provides data to discuss in detail as to how to use the data to serve the energy-saving management in the real world.

2.2. Environmental Awareness. After reviewing the relevant literature, it can be found that studies focusing on driver's awareness of fuel-saving is relatively scarce. Considering that the overuse of fossil fuels is closely related to the environment, this section mainly discussed the relationship between environmental awareness and behaviors of different subjects.

Environmental awareness is generally regarded as the primary step in helping improve the environment, reflecting people's attention and knowledge of the impacts of their behaviors on the environment [12]. In early behavioral economics, Kahneman et al. [13] made a clear description of the relationship between awareness and behavior: people's decision-making process is influenced by preferences, beliefs, and cognitive biases. This theory suggests that people's cognitive level of environmental knowledge induces their behavior unconsciously. Generally, individuals with higher environmental awareness are more likely to behave in an environmentally sustainable manner [14, 15]. Also, the lack of such awareness may lead to showing indifferent behavior toward the environment [16]. Previous empirical studies mainly focus on the relationship between environmental awareness and consumption behavior. It was found that environmental awareness can increase drivers' preferences to choose products with eco-labels or organic products [17, 18]. Okada et al. [19] summarized that environmental awareness can increase purchasing intentions of electric vehicles, which confirmed that energy conservation awareness increased the likelihood of the occurrence of energy-saving behavior. A similar study can also be found in Li et al. [20], where a simple model was developed incorporating individuals' environmental awareness in their decision-making processes and proposed that awareness does not necessarily reduce energy consumption, but subjective factors can play an important role in energy-saving. These studies implied that correct environmental awareness can invisibly promote environmental behavior. Hence, it seems more reasonable to investigate the fuel-saving awareness of freight drivers in this paper.

In addition, several studies have confirmed this view from other aspects. Environmental awareness could be separated into two distinct categories: general environmental awareness and specific environmental awareness. Compared with the former, specific environmental awareness indicates people are concerned about more specific or localized environmental issues [21]. In this study, drivers' awareness of fuel-saving factors is more closely linked to the latter, as it affects each micro-driving decision during transportation. Mobley et al. [22] have revealed that specific environmental awareness has a stronger power in inducing environmental protection intention and behavior. Although some literature points out that there is a gap between awareness and behaviors [23], effective measures can help transform awareness into behavior, subsequently bridging
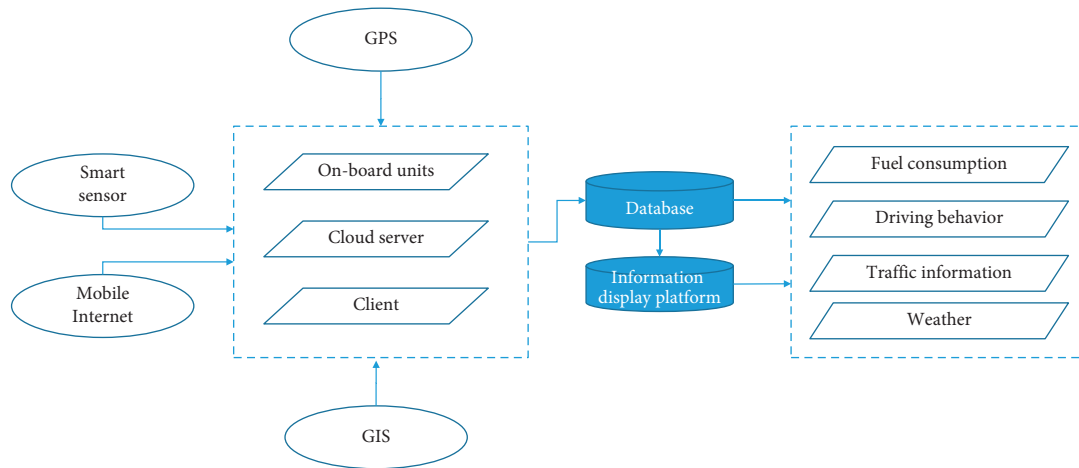
FIGURE 1: The schematic of the advanced on-board fleet management system.

this gap [24]. Arguably, there is feedback between exploring whether there is a bias in driver's fuel-saving awareness and formulating effective driving training and management strategies. That is, fully grasping the driver's fuel-saving awareness helps to formulate corresponding management countermeasures, and effective management countermeasures will further promote the improvement of the driver's fuel-saving awareness.

Until today, studies relating to eco-driving have focused little on driver's fuel-saving awareness. Nearly all the papers focused on the objective data analysis and directly put forward the eco-driving training strategy based on the analysis results [25, 26]. Although this practice has proved to be effective [27, 28], it will achieve better efficacy if the eco-driving practice is carried out on the basis of investigating drivers' awareness of fuel-saving [29]. Therefore, this study aims to analyze whether there are deviations in drivers' awareness of fuel-saving factors and guide 3PL companies to formulate more targeted driver training plans by showing this awareness bias.

2.3. Review of Modeling Methods.

In this paper, we designed two paths to analyze the fuel consumption of HDDTs. Firstly, for the 3PL company that provided data, a cognitive survey of fuel-saving factors (i.e., the importance of different factors in the driver's awareness) was carried out for drivers of HDDTs. Then, the naturalistic driving data obtained from the advanced on-board FMS were mined to evaluate the intrinsic importance ranking of multiple factors. The methodological review in this section primarily focuses on the following two aspects.

2.3.1. Subjective Awareness Survey.

Questionnaires have been widely applied to investigate people's attitudes toward certain things. Identifying the factors that influence human willingness to purchase or accept alternative fuel vehicles is one of the most common directions, such as consumers' preference for new energy vehicles [30, 31], truck fleets' acceptance of alternative fuel vehicles [32], and acceptance

of electric ride-hailing from the driver's perspective [33]. Analyzing driver psychology and behavior is another important application of this method. Yang et al. [34] designed a driving attitude scale, a risk perception scale, and a queue-jumping behavior scale to survey the factors affecting drivers' queue-jumping. After verifying the reliability and validity of the developed scales, a structural equation model (SEM) was established to explore the interrelationships between related characteristics and queue-jumping behaviors. Obst et al. [35] developed surveys to explore participants' past behaviors with regard to driving while fatigued and their perceptions of risk associated with driving fatigued. Castritius et al. [36] employed an online questionnaire to investigate the public's willingness and views on platooning technology in Germany and California with 536 participants. At present, besides the SEM, there are several analysis methods to deal with the questionnaire data, such as factor analysis [37], principal component analysis [36], hypothesis testing [31], and correlation analysis [38]. In this paper, given that our purpose was to sort the importance of the questionnaire items, we choose the principal component analysis (PCA) as the analysis tool [39].

2.3.2. Data Mining Method.

In this paper, data mining was used to process large-scale naturalistic driving data and identify the importance of explanatory variables to fuel consumption. This demand just caters to the advantages of machine learning, which is a better approach to address nonlinearity and large-scale problems [40]. However, most of the machine-learning methods are often treated as "black box" and failed to nonlinearly evaluate the variable importance [41]. Therefore, this study developed an Extreme Gradient Boosting (XGBoost) model to quantitatively capture the importance of variables. XGBoost, as a new machine-learning method for regression and classification, is an efficient implementation of the boosting concept. It has incomparable advantages in terms of accuracy and speed compared with other ensemble learning algorithms [42, 43]. In addition, it can prevent overfitting by

regularization and detect nonlinear effects or interactions when dealing with large-scale data [44, 45].

In a word, this paper developed a method based on the advanced FMS, which can be applied to 3PL companies to investigate the awareness of their fleet drivers on fuel-saving factors, so as to formulate effective eco-driving management strategies. Taking a 3PL company in China as an example, we demonstrated the whole process of this method and made some contributions to the company's decarbonization and energy conservation relying on the survey results.

3. Research Goals and Innovations

Our key objective was to exhibit how 3PL companies can improve their fuel efficiency and consumption by using advanced on-board FMS to analyze their drivers' awareness of fuel-saving factors. The framework is shown in Figure 2. Its innovation and contribution can be summarized in the following aspects:

    (i) Demonstrated how to make full use of the fuel-related data obtained from advanced on-board FMS of 3PL companies to contribute to road freight decarburization.

    (ii) Evaluated the driver's awareness of fuel-saving factors based on the results of data mining, which enhances the reliability of the research conclusion.

    (iii) Questionnaire analysis and mining driving data are common research methods in this field, but previous studies only involve one research method, which greatly weakens the role of conclusions in practical guidance. This paper fills this research gap by combining the two methods including subjective investigation and objective data analysis to provide help for energy-saving and decarburization of road freight transportation.

4. Data Sources and Methods

4.1. Subjective Awareness Survey

4.1.1. Questionnaire Participants. The data provider of this study is a 3PL company named TopChains International Logistics Co., Ltd. (TopChains), which provides naturalistic driving data of 39 HDDTs. Therefore, our questionnaire would also be administered to 116 drivers of these 39 HDDTs. Note that the method used in this paper is sufficient to deal with large-scale data analysis, with strong portability, and can be applied to 3PL companies of different sizes. Due to the prevalence of COVID-19, we chose to conduct an investigation using an online questionnaire. An instructed item was set to filter valid questionnaires [46]. A total of 104 responses to the online questionnaire were collected after screening. Due to the particularity of occupation, almost all of them are male engaged in freight driving in China. What needs to be emphasized is that all participants were voluntary participants and were informed in detail of the procedure and purpose of this survey before participating in it.

4.1.2. Questionnaire Design. In the first part of the questionnaire, two questions were set to investigate the driving age and self-awareness of fuel consumption of the 104 male drivers. As provided in Figure 3, the vast majority of drivers have more than 15 years of driving experience, indicating that the drivers participating in this survey have rich experience and are sufficient to represent the current situation of professional truck drivers in China. Moreover, 89.12% of drivers thought they knew very well about the factors affecting fuel consumption, and 10.88% of drivers thought they knew well, reflecting that drivers tended to not perceive they have a bias in their awareness of the factors affecting fuel consumption. On the contrary, they will subconsciously adhere to their own awareness, which might lead to a negative impact on energy conservation.

In the second part, 16 variables were listed that might have a significant impact on fuel consumption. Participants were asked to score the importance of these variables according to their awareness and experience. The Likert scale method is conducted to quantify each option (5–very important, 4–important, 3–neutral, 2–not important, or 1–not at all important), laying the foundation for the subsequent PCA process [47], as shown in Table 1.

4.1.3. Questionnaire Data Analysis

(1) Reliability and Validity Test. In this study, SPSS 24.0 was used to examine the reliability and validity of the questionnaire. The reliability test refers to the test of the reliability of the result data of the questionnaire. The test result can indicate the quality of the questionnaire design. This study intends to use Cronbach's alpha to test the reliability of the questionnaire. Cronbach's alpha is the most commonly used reliability evaluation tool in psychological or educational tests [48]. According to experience, it is generally contended that when α is lower than 0.6, the reliability of the questionnaire is poor and cannot be used for research; when α is higher than 0.7, the questionnaire can be used to carry out more in-depth research. In this paper, SPSS 24.0 was used to measure the reliability of the importance of various factors on the fuel consumption of HDDTs per 100 km. The result shows that the Cronbach alpha coefficient of this questionnaire is 0.812, indicating that the data of the questionnaire is relatively reliable and has a high degree of internal consistency.

For the validity test, KMO (Kaiser–Meyer–Olkin) was used to analyze the correlation between variables, with the value ranging from 0 (no correlation) to 1 (the strongest correlation). Bartlett's test was applied to measure whether the variables are independent or not. When the KMO value is higher than 0.5 and the p-value of Bartlett's test is less than 0.05, the subsequent PCA can be carried out [34, 49]. In the current study, the KMO value was 0.746, and Bartlett's $\chi^2 = 653.722$ ($p < 0.01$); thus, it can be used for factor analysis or PCA.

(2) Principal Component Analysis (PCA). Before using the PCA method, the variables used are usually transformed into
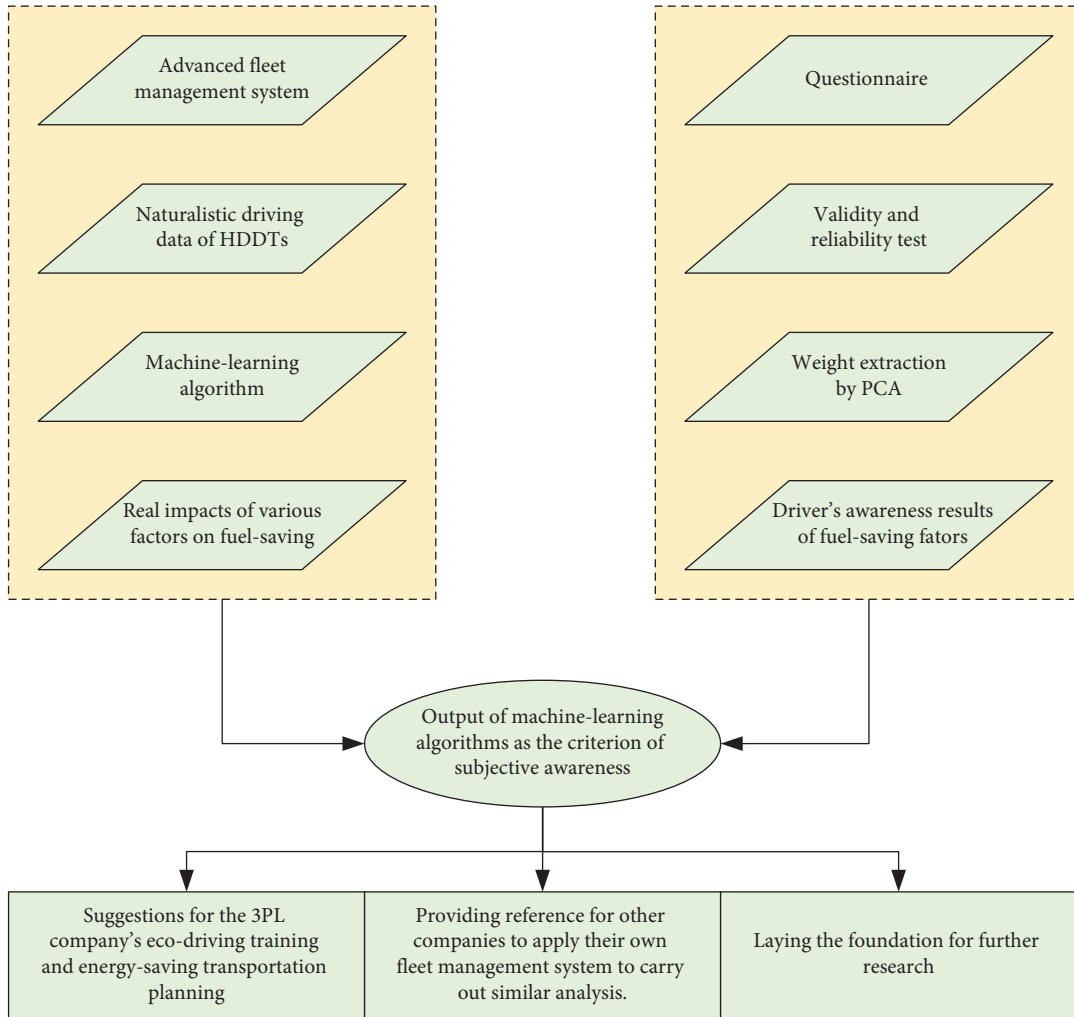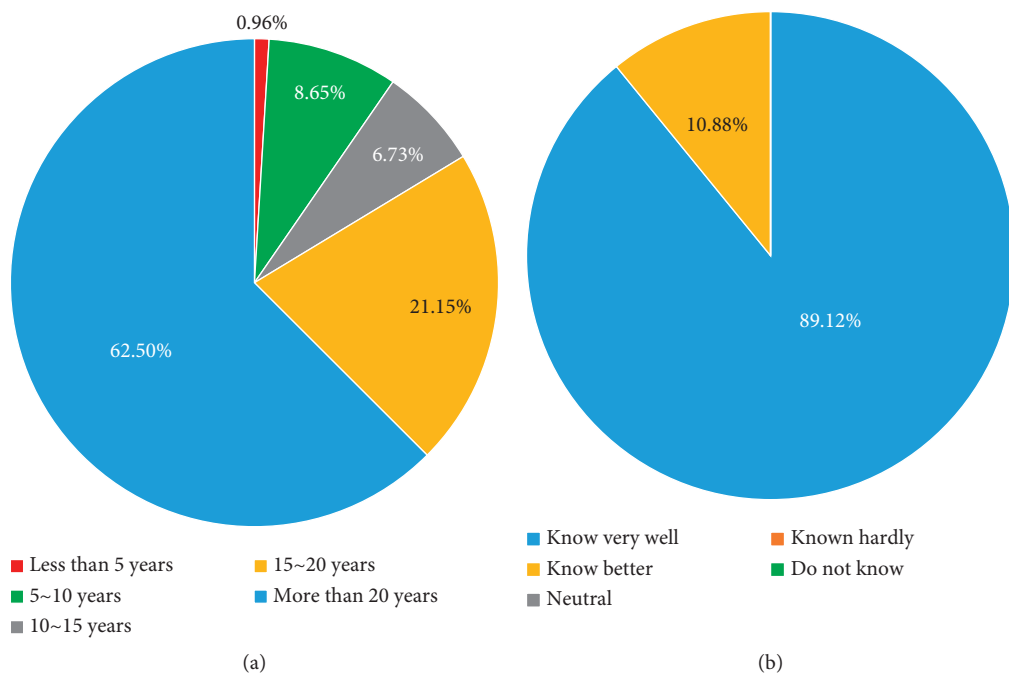
Figure 2: The framework of this study.

Figure 3: The result of part one. (a) Driving experience. (b) The driver's understanding of his vehicle's fuel consumption.

TABLE 1: The second part of the questionnaire.

Items	Scale					Mean	STD
	5 (%)	4 (%)	3 (%)	2 (%)	1 (%)		
Driven distance (mileage)	74.04	5.77	13.46	5.77	0.96	4.46	0.99
Altitude while driving	82.69	11.54	3.85	1.92	0.00	4.75	0.62
Altitude change while driving	11.54	81.73	4.81	1.92	0.00	4.03	0.49
Whether to maintain economic speed	12.5	78.85	8.65	0.00	0.00	4.04	0.46
Vehicle speed	83.65	11.54	3.85	0.96	0.00	4.78	0.49
Engine speed	15.38	78.85	4.81	0.96	0.00	4.09	0.48
The degree of change in vehicle speed	78.85	14.42	4.81	0.96	0.96	4.69	0.70
The degree of change in engine speed	13.46	81.73	4.81	0.00	0.00	4.09	0.42
Whether to skid in neutral	3.85	4.81	4.81	77.88	8.65	2.17	0.82
Whether to skid with gear	3.85	5.77	74.04	11.54	4.81	2.92	0.75
Whether idling	11.54	72.12	9.62	4.81	1.92	3.87	0.72
Proportion of parking	4.81	5.77	78.85	4.81	5.77	2.99	0.73
Whether the driving day is a holiday	7.69	80.77	5.77	1.92	3.85	3.87	0.74
Temperature	5.77	81.73	9.62	2.88	0.00	3.9	0.42
Road types	85.58	11.54	2.88	0.00	0.00	4.83	0.45
Rainfall	18.27	77.88	2.88	0.96	0.00	4.13	0.44

Note: "5, 4, 3, 2, 1" means very important, important, neutral, not important, not at all important, respectively; "STD" means standard deviation.

matrix forms. Assuming that there are n observations and each observation has p variables or measurements, they can be represented by matrix X as follows [50]:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}. \qquad (1)$$

If the PCA is suitable for the analysis, there are $K < n$ principal components. For each principal component i (or j) [51]:

$$Z_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{ip}x_p, \qquad (2)$$

$$a_{i1}^2 + a_{i2}^2 + \cdots + a_{ip}^2 = 1, \qquad (3)$$

$$\mathrm{COR}\left[Z_i, Z_j\right] = 0. \qquad (4)$$

Furthermore, the symmetric $p \times p$ sample variance-covariance matrix is given as [50, 52]:

$$s^2[X] = \begin{bmatrix} s^2(x_1) & s^2(x_1, x_2) & \cdots & s^2(x_1, x_p) \\ s^2(x_2, x_1) & s^2(x_2) & \cdots & s^2(x_2, x_p) \\ \vdots & \vdots & \ddots & \vdots \\ s^2(x_p, x_1) & s^2(x_p, x_2) & \cdots & s^2(x_p) \end{bmatrix}, \qquad (5)$$

where the diagonal elements of this matrix stand for the estimated variance of random variable 1 through p; the off-diagonal elements represent the estimated covariance between variables.

After this, the eigenvalues λ_p were introduced to represent the sample variance.

$$\lambda_1 + \lambda_2 + \cdots + \lambda_p = \mathrm{VAR}(x_1) + \mathrm{VAR}(x_2) + \mathrm{VAR}(x_p). \qquad (6)$$

It can be derived from equation (11) that the sum of the eigenvalues is equal to the sum of variance $(s^2(x))$, so the proportion of total variance interpreted by the jth principal component is as follows [50]:

$$\mathrm{VAR}_j = \frac{\lambda_j}{\lambda_1 + \lambda_2 + \cdots + \lambda_p}, \quad j = 1, 2, \ldots, p. \qquad (7)$$

There is something worth noting; the correlation matrix should use the standardized variables Z_{ij} instead of the original variables X_{ij} [53]:

$$Z_{ij} = \frac{X_{ij} - \overline{X_j}}{\sigma_j}, \quad i = 1, 2, \ldots, n; j = 1, 2, \ldots, p. \qquad (8)$$

According to experience, often choose the principal component with the eigenvalues greater than 1, and ensure that the cumulative variance explanation is at least greater than 60%.

According to the principal component expression, that is, the coefficients a_{ij} in equation (9), and the proportion of total variance interpreted by each principal component, the weight of the ith original indicator W_i can be obtained to measure the relative importance among the indicators:

$$W_i = \frac{\sum a_{ij} \times \mathrm{VAR}_j}{\sum \mathrm{VAR}}. \qquad (9)$$

4.2. Data Obtained from the Advanced On-Board Fleet Management System

4.2.1. Data Description. The type of HDDTs involved in this study is SITRAK engaging in traditional road transportation. SITRAK, belonging to high-end products, is a joint venture brand jointly promoted by China NATIONAL HEAVY DUTY TRUCK GROUP CO., LTD and the German MAN Group. The design life of its gearbox and axle is up to 1

million km. The detailed parameters are introduced as follows (see Table 2).

The advanced on-board fleet management system of TopChains is called SINOTRUK TELEMATICS (see Figure 4).

This system provides the daily naturalistic driving data (contains 1153 trips) of 39 HDDTs from 01.09.2019 to 31.01.2020. A total of 21 variables, including all variables involved in the questionnaire, were obtained as illustrated in Table 3. Furthermore, STATA 16 was utilized to make descriptive statistics on these 21 variables, and the statistical results are shown in Table 3.

In Table 4, all explanatory variables are grouped into three categories according to their properties: driving characteristics, road characteristics, and environmental characteristics. Fuel consumption is determined as the dependent variable for data analysis and modeling.

4.2.2. Extreme Gradient Boosting (XGBoost). Given the complex relationships among various factors, a machine-learning algorithm was developed to reveal the underlying influence of factors on the fuel consumption of HDDTs. As illustrated in previous studies [54, 55], gradient boosting, which was proposed by Friedman [56], is an ensemble machine-learning method for regression and classification with a tree structure. Compared with traditional tree-based algorithms, XGBoost has some obvious advantages, such as introducing a second-order Taylor expansion in the loss function and adding a regularization term to control overfitting [57]. In addition, it has high efficiency when dealing with missing values or large-scale data.

Referring to some current papers [43, 56], the objective function is as follows:

$$F(t) = \sum_{i=1}^{n} l\left(y_i, \widehat{y}^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) + c, \qquad (10)$$

where t represents the tth tree; x_i is the set of features classified in leaf nodes; $\Omega(f_t)$ is the regularization term; c is a constant part.

After adopting the Taylor Expansion second order to equation (3), the objective function can be rewritten as follows:

$$F(t) = \sum_{i=1}^{n} \left[l\left(y_i, \widehat{y}^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) + c. \qquad (11)$$

In this study, we will demonstrate the rank of feature importance using XGBoost. The importance of features in a single XGBoost tree is calculated by the amount of information gain after splitting the tree using the feature. Its calculation is shown in the following equation [43]:

$$IG(T, F) = H(T) - H(T|F) = -\sum_{i=1}^{J} p_i \log_2 p_i - \sum_{F} p(F) \times \sum_{i=1}^{J} - p(i|F) \log_2 p(i|F), \qquad (12)$$

where $H(T)$ and $H(T|F)$ are the entropy of the parent node and the child nodes of the split based on the F feature and p_i is the fraction of each labeled sample in one node.

It should be noted that the grid-search approach was employed to tune the best hyper-parameters in the XGBoost model. Grid-search refers to building a model for each combination of hyper-parameters separately to determine the parameters corresponding to the best performing model. In this study, the set of hyper-parameters was presented as follows, and the best choice for each one was [0, 1, 0.01, 5, 750, 0.7].

(i) Parmeters = [gamma, lambda, learning_rate, max_ *depth depth*, *n*_estimators, subsample]

The importance ranking can intuitively reflect the impact of different factors on fuel consumption. However, it is unable to evaluate how various factors affect fuel consumption (whether it is positive or negative, linear or nonlinear). For this purpose, SHAP values were utilized to interpret these impacts by calculating the performance of the model with and without the feature. In this manner, the individual impact of each factor on fuel consumption can be estimated, as well as the interactive effects between different factors.

SHAP (Shapley additive explanation) was first presented by Štrumbelj and Kononenko [58] based on economic game theory. SHAP is essentially a model interpreter that can visualize the output of the black-box model (e.g., XGBoost, random forest, and decision tree), thereby enhancing the application value of research. It compares the distance between the contribution of each feature and the average effect by calculating the marginal effect. In other words, a larger SHAP value (absolute) indicates that the feature has a greater impact on model performance [59]. As such, the importance and contribution of each feature can be obtained. Another advantage is that it can not only show the global effect of the feature set but also calculate the local effect of each feature and the correlation effect between features.

The algorithm's solution speed is exponentially improved after improvement by [60], which has demonstrated to provide better identification of influential features compared to other traditional approaches [55, 61]. In this study, The SHAP library in Python was utilized to implement the writing and solving of the SHAP algorithm. We sort the fuel-saving performance of all factors according to the SHAP value, and quantitatively give the intrinsic impacts of factors on fuel consumption. For several factors of interest, the overall effect and interaction effect of them are further analyzed in detail.

5. Results and Discussion

Table 5 and Figure 3 demonstrate the estimated results of the PCA and XGBoost model, respectively. The PCA reveals the relative importance of each factor in the driver's subjective cognition, and the XGBoost gives the intrinsic importance ranking of each factor in the real world. Indeed, many factors, including vehicle type, road type, climate, geographical environment, etc., will cause

TABLE 2: The specific parameters of the HDDTs.

| Parameters	Content	Parameters	Content
Drive form	6 × 4	Curb weight	8.54 t
Engine	CNHTC: MC13.54-50	Total mass	25 t
Maximum horsepower	540	Fuel type	Diesel
Emission standards	National V emission standard	Passengers	3 people
Gearbox	ZF Friedrichshafen: ZF16S2530 TO	Displacement	12.419 L
Tread front	2022 mm	Tread rear	1830 mm
Axle-base	3200 + 1400 mm	Maximum power	397 kw
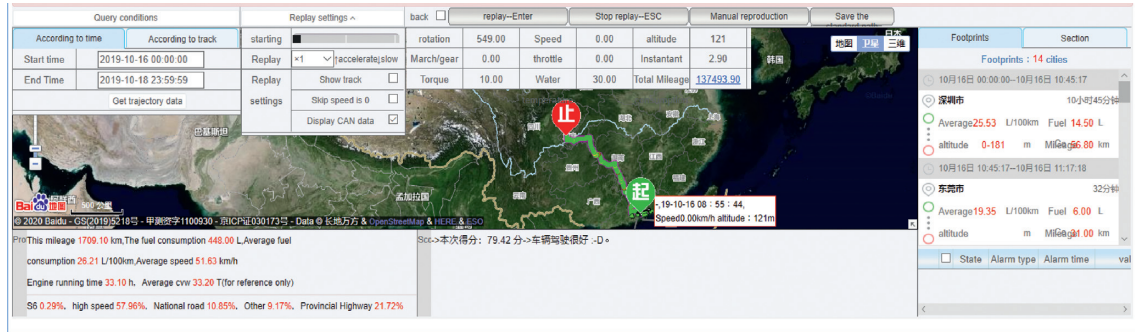
FIGURE 4: Advanced on-board measurement systems.

different fuel consumption performance. At the same time, the characteristics of drivers in different regions are also the main uncontrollable factors. Their education level, driving habits, family, and other conditions are very likely to affect their awareness of fuel-saving factors. It can be said that the driver's awareness bias of fuel-saving factors has spatial and temporal differences. Therefore, different 3PL companies should use the methods proposed in this study to conduct independent surveys of drivers' cognitive biases on fuel-saving factors based on their own advanced FMS. Companies with similar fleet situations can benefit directly from the conclusions of this paper, but it is still recommended that these companies carry out independent research to obtain more accurate results. In any case, from a macro perspective, drivers in a certain country or region may have similar cognitive biases on fuel-saving factors (the difference is only the magnitude of the deviation or the partial ranking of factors). Of course, such speculation needs more in-depth research to support.

In the XGBoost model, more variables were considered to increase the reliability of the model outcomes, because in the real world, these factors exist simultaneously and affect each other significantly. In order to have an intuitive visualization of the impact relating to these determinants, the SHAP Python library was utilized to provide individualized factor attributions. After adjustment, the final importance ranking and corresponding effectiveness of each variable were obtained, which laid the foundation for further discussion. In Figure 5, variables are sorted by their global impact on the y-axis, and each dot is colored from low (blue) to high (red), with a smooth gradual change in color [55].

When multiple samples have similar effects, they overlap vertically.

To compare the influence of different variables conveniently, each variable was labeled with similar meanings, as shown in Table 6 and Figure 6. In this paper, the indicator of Freeway was chosen to represent four road characteristics in the XGBoost model as freeways are the main channels for long-distance road freight. After calibrating the variables with similar meanings in the two methods, a total of 16 indicators were determined to analyze the final results. To compare the performance of the PCA and XGBoost, the subtracting technique was utilized to visualize the match value of the two rankings (see equation (13)), and the outcome is shown in Table 5.

$$\text{Match value} = \text{Rank}_S - \text{Rank}_o. \tag{13}$$

where Rank_S and Rank_o represent the indicator importance from the perspectives of subjective and objective, respectively. After the transformation of this equation, a greater Match value implies a greater underrating of this indicator by drivers. Equally, a smaller Match value denotes that the indicator is overestimated by drivers.

More generally, some indicators could lead to fuel reduction, but the company has little control over them in real transport organization and fleet management, such as weight load, weather, altitude, and holiday. Also, route choice is out of its control in many cases because there could be only one, or if lucky, two routes available. For these reasons, the following discussion will concentrate on indicators that are easily improved through management and training.

TABLE 3: The data obtained from the advanced on-board fleet management system.

Variables	Definition	Variable type
Fuel consumption	Fuel consumption per 100 km for each trip: 0 if low, 1 if medium, 2 if high	Discrete variable
Weight	Average weight load of freight in tones per trip	Continuous variable
AVG_Rotating	Average engine revolution speed per trip/(100 r/min)	Continuous variable
SD_Rotating	Standard deviation of engine revolution speed per trip	Continuous variable
AVG_V	Average speed per trip/(km/h)	Continuous variable
SD_V	Standard deviation of speed per trip	Continuous variable
PR_Eco_Rotating	Percentage of driving time that the engine runs in the economic revolution speed range per trip	Continuous variable
PR_Neutral_Skidding	Percentage of driving time that the truck is rolling without engine load per trip	Continuous variable
PR_Gear_Skidding	Percentage of driving time that the truck is rolling with engine load per trip	Continuous variable
PR_Idling	Percentage of driving time spent running idle per trip	Continuous variable
PR_Parking	Percentage of time spent parking per trip	Continuous variable
Mileage	Mileage per trip	Continuous variable
Freeway	Percentage of distance that the truck is driven on freeways per trip	Continuous variable
O_N_F	Percentage of distance that the truck is driven on ordinary national roads per trip	Continuous variable
O_P_F	Percentage of distance that the truck is driven on ordinary provincial roads per trip	Continuous variable
Other_Road	Percentage of distance that the truck is driven on other low-grade roads per trip	Continuous variable
AVG_Altitude	Average altitude per trip/(100 m)	Continuous variable
Altitude_Change	The change of altitude per trip/(100 m)	Continuous variable
Holiday	A dummy variable indicating whether the driving day is a holiday: 0 if no, 1 if yes)	Discrete variable
Tem	Average temperature outside the trucks: 0 if $\leq 10°C$, 1 if $10–15°C$, 2 if $15–20°C$, 3 if $20–25°C$, 4 if $25–30°C$, 5 if $> 30°C$ (the statistical interval is left open and right closed)	Discrete variable
Rainfall	Average rainfall per trip: 0 if $0–1$ mm, 1 if $1–8$ mm, 2 if $9–20$ mm, 3 if ≥ 20 mm (the statistical interval is left closed and right open)	Discrete variable

5.1. The Overestimated Indicators.

(i) Whether to maintain economic speed. It can be seen from the calculation results that the indicator of economic speed is not so powerful in the real world. One possible explanation why drivers overestimate the impact of this indicator is that they have been accustomed to the original knowledge that the economic speed will not be changed. As is known to all, an economic speed range will be calibrated before engines leave the factory. Generally, the manufacturer will tell the buyer that it will be more fuel-efficient within this speed range. However, the range of economic speed is usually calibrated under ideal experimental conditions, and the engine almost has no loss at this time. With the increase of working time, the engine loss is also increasing. In addition, under realistic driving conditions, the change of fuel consumption is restricted by multiple other factors, which will change the range of economic speed, exceeding the original concept of the drivers, thereby weakening the effectiveness of this indicator on fuel-saving. Therefore, more relevant information should be provided to drivers to enhance their cognition. With the loss of engine parts, fleet managers should inform the drivers of the dynamic changes of economic speed in driving training. But, to succeed in fuel-saving, the 3PL company is best able to use naturalistic driving data to obtain the latest economic speed range. In this case, the global impact of each factor on fuel consumption was characterized, as shown in Figure 7. Although increasing the proportion of economic speed can still play a role in reducing fuel consumption, it is obvious that this effect is weak, because the distribution of these scatters is too discrete.

TABLE 4: Descriptive statistics.

Variables	Obv.	Mean	SD	Median	Min	Max
Driving characteristics						
Weight	1153	36.280	6.280	36.450	13.840	58.910
AVG_Rotating	1153	11.311	0.858	11.489	8.039	13.772
SD_Rotating	1153	1.952	0.369	1.929	0.945	3.086
AVG_V	1153	57.760	10.360	59.210	23.030	83.490
SD_V	1153	20.260	4.030	19.742	4.354	31.417
PR_Eco_Rotating	1153	0.260	0.150	0.242	0.001	0.645
PR_Neutral_Skidding	1153	0.020	0.009	0.014	0.000	0.043
PR_Gear_Skidding	1153	0.008	0.006	0.007	0.000	0.027
PR_Idling	1153	0.038	0.021	0.034	0.000	0.104
PR_Parking	1153	0.650	0.170	0.653	0.249	0.997
Mileage	1153	479.500	276.250	477.100	2.700	1247.400
Road characteristics						
Freeway	1153	410.500	234.330	422.000	1.000	814.000
O_N_F	1153	117.470	133.590	56.000	1.000	435.000
O_P_F	1153	145.560	155.400	74.000	1.000	486.000
Other_Road	1153	346.010	202.500	344.000	1.000	681.000
Environmental characteristics						
AVG_Altitude	1153	1.953	1.594	1.520	0.050	6.830
Altitude_Change	1153	4.36	3.288	3.790	0.100	13.270
Holiday	1153	0.036	0.190	0.000	0.000	1.000
Tem	1153	1.560	1.140	1.383	0.000	5.000
Rainfall	1153	0.190	0.350	0.000	0.000	2.000
Dependent variable						
Fuel consumption	1153	1.050	0.615	1.000	0.000	2.000

TABLE 5: Factor importance ranking of driver's awareness perspective.

Indicator	Weights (%)	Rank
Altitude while driving	8.665681	1
Road types	8.249529	2
Vehicle speed	7.924107	3
Whether to maintain economic speed	7.849868	4
Altitude change while driving	7.559500	5
Temperature	7.454759	6
Rainfall	7.305892	7
Driven distance (mileage)	6.338620	8
The degree of change in vehicle speed	6.215343	9
Whether the driving day is a holiday	5.972035	10
The degree of change in engine speed	5.775637	11
Whether idling	5.274959	12
Whether to skid with gear	4.595463	13
Proportion of parking	4.514446	14
Engine speed	4.231072	15
Whether to skid in neutral	2.073099	16

(ii) Environmental-related indicators. Interestingly, drivers were found to overestimate the power of all environmental-related indicators, which might be explained by "social expectations" [22], that is, drivers subjectively regard uncontrollable environmental factors as more important indicators, while driving behavior factors that are closely related to them are underestimated. In this way, the main responsibility for high fuel consumption has been transferred to the environment. In addition, an important conclusion can be drawn that optimizing drivers' behavior might be more critical to fuel-saving than the external environment, contributing to the greater success of eco-driving research [62].

5.2. The Underrated Indicators

(i) Average engine speed. In this case, it is surprising that the drivers underrated the power of engine speed. Therefore, the company should pay more attention to the education of engine speed optimal control in driving training. It can be seen from Figure 8 that there is a negative correlation between engine speed and fuel consumption. Moreover, with the increase of engine speed, fuel consumption decreases faster.

(ii) Percentage of driving time spent running idle. It can be seen from the output of the XGBoost model that reducing the proportion of idle state in transportation is the key to saving oil. But, the driver seriously underestimated its impact; such cognitive bias is likely to lead to the driver increasing some unnecessary idle conditions. From the global impact of idling on fuel consumption (see Figure 9), it can be concluded that the indicator of idling has a significant positive correlation with fuel consumption, which is consistent with the previous research [63, 64]. Therefore, the fleet manager should focus on the characteristics of idle conditions in the actual eco-driving training, so as to assist drivers to avoid the occurrence of this operating condition.
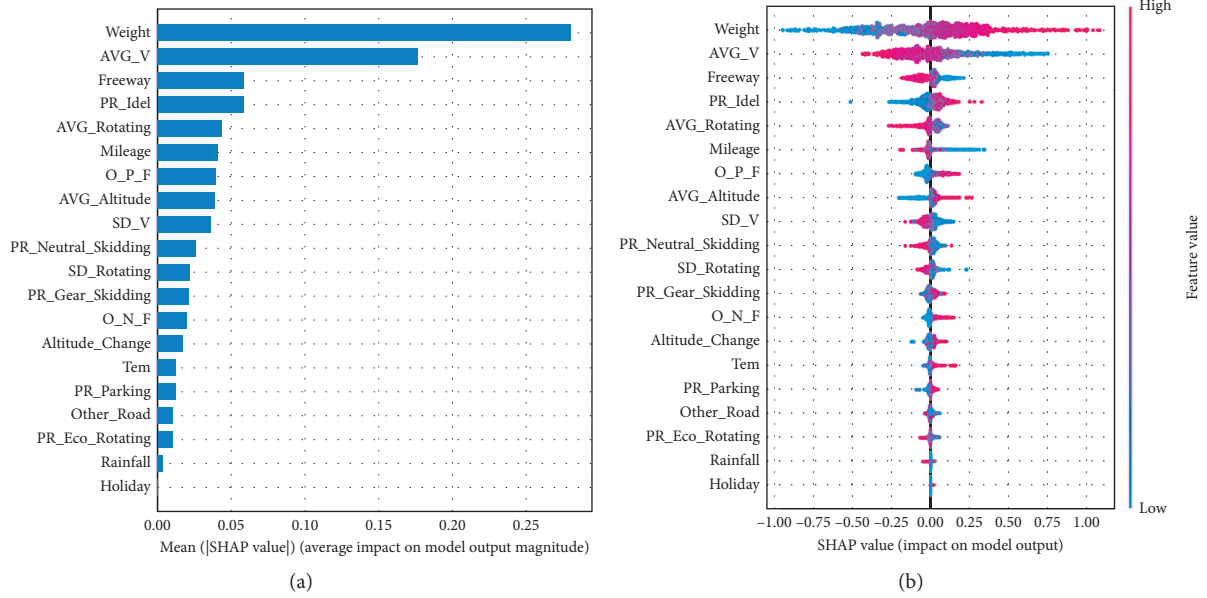
FIGURE 5: SHAP summary plot for fuel consumption based on the XGBoost model.

TABLE 6: Result of indicator matching.

Label	The corresponding variable in PCA	The corresponding variable in XGBoost	Match value
(1)	Vehicle speed	AVG_V	2
(2)	Engine speed	AVG_Rotating	11
(3)	The degree of change in vehicle speed	SD_V	2
(4)	The degree of change in engine speed	SD_Rotating	2
(5)	Whether to maintain economic speed	PR_Eco_Rotating	−10
(6)	Whether idling	PR_Idling	9
(7)	Whether to skid in neutral	PR_Neutral_Skidding	8
(8)	Whether to skid with gear	PR_Gear_Skidding	3
(9)	Driven distance (mileage)	Mileage	3
(10)	Proportion of parking	PR_Parking	1
(11)	Road types	Freeway	0
(12)	Altitude while driving	AVG_Altitude	−5
(13)	Altitude change while driving	Altitude_Change	−6
(14)	Rainfall	Rainfall	−10
(15)	Whether the driving day is a holiday	Holiday	−6
(16)	Temperature	Tem	−6

(iii) Percentage of driving time that the truck is rolling without engine load. As presented in Figure 10, the results revealed that this indicator has a significantly negative impact on fuel consumption, also confirmed by Walnum and Simonsen [6] in their research. It should be noted that in the subjective survey of this case, the fuel-saving effect of this indicator ranked last due to "social expectations" that such driving is dangerous, especially for heavy trucks, and can easily lead to serious traffic accidents. Therefore, it is unnecessary to emphasize the

fuel-saving role of this indicator in fleet ecological management for safety reasons.

5.3. Other Factors Worth Noting

(i) Average speed. Vehicle speed has a powerful force on fuel-saving, which can be intuitively obtained in the estimation results of XGBoost, as shown in Figure 11. Also, this indicator has also been reasonably recognized in the driver's awareness. Most of the current studies have confirmed that
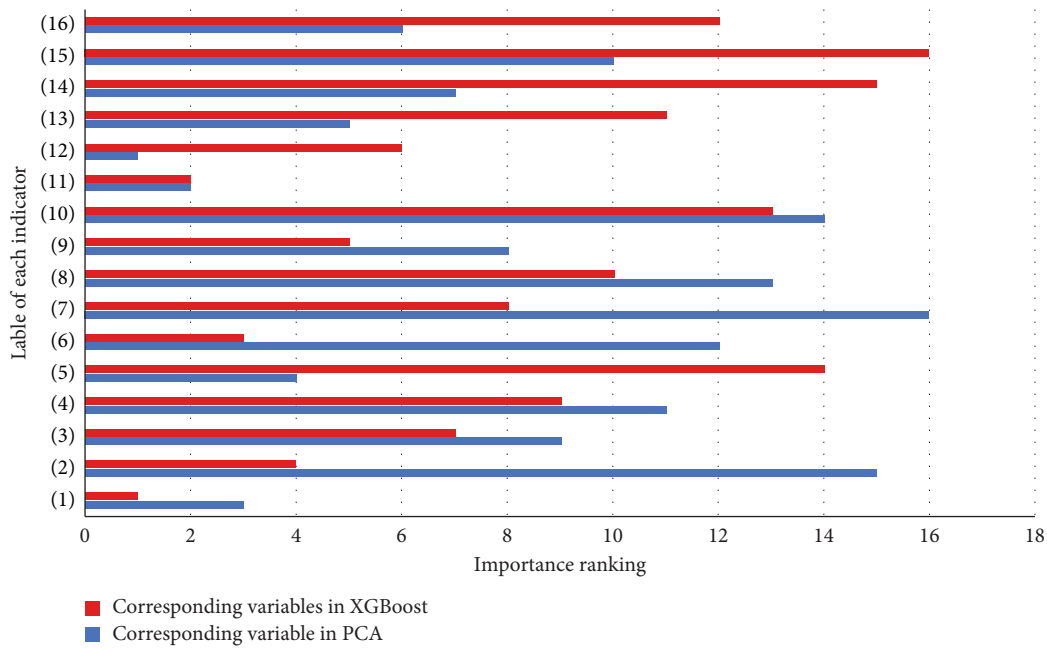
FIGURE 6: Comparison of subjective cognitive survey results and machine-learning algorithm output.
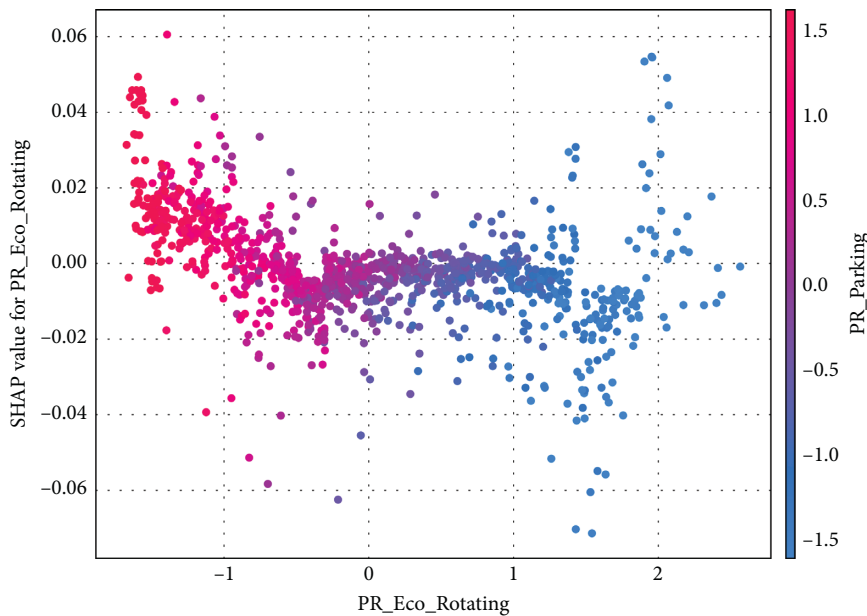

FIGURE 7: The global impact of economic speed on fuel consumption.

maintaining a high and stable speed is significantly related to the ecological driving modes [63, 65].

(ii) Driven distance (mileage). Although there is no deviation in the driver's awareness of mileage, an interesting phenomenon was still captured. The global impact of mileage on fuel consumption shows obvious segmentation, as shown in Figure 12. As a whole, the longer mileage is conducive to fuel-saving when other conditions remain unchanged. However, a significant inflection point appeared near 800 km, suggesting a sharp rise in fuel

consumption forming a local peak. This could be explained by several psychological factors such as fatigue, which reminds team managers and drivers to try to control a single trip within 800 km. If the total mileage of a transport mission exceeds 800 km, the driver is advised to take adequate rest before driving 800 km without affecting timeliness.

(iii) Road types. Relying on good conditions, smooth traffic, and no signal obstruction, the influence of road type on fuel reduction is easy to understand. If there is more than one route available, it is helpful to
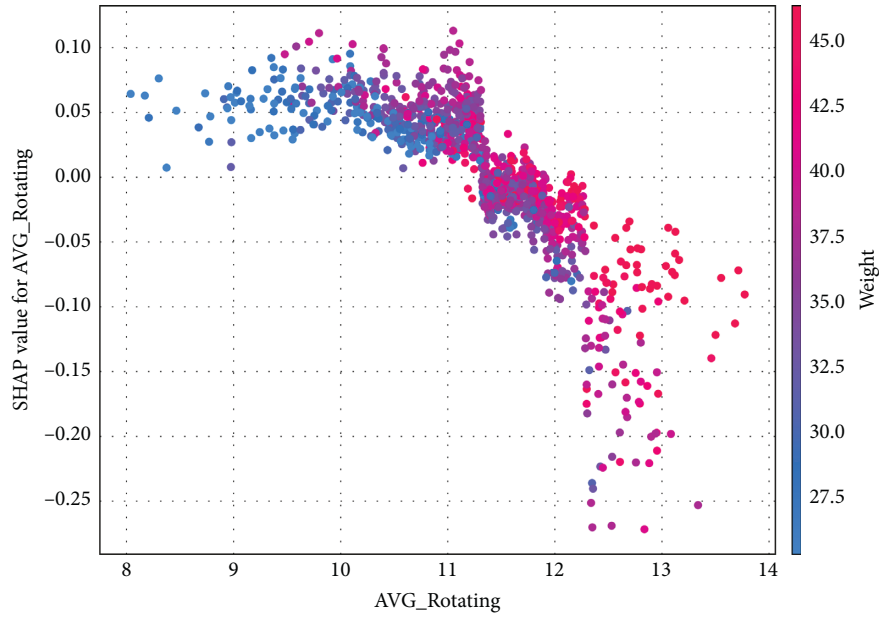
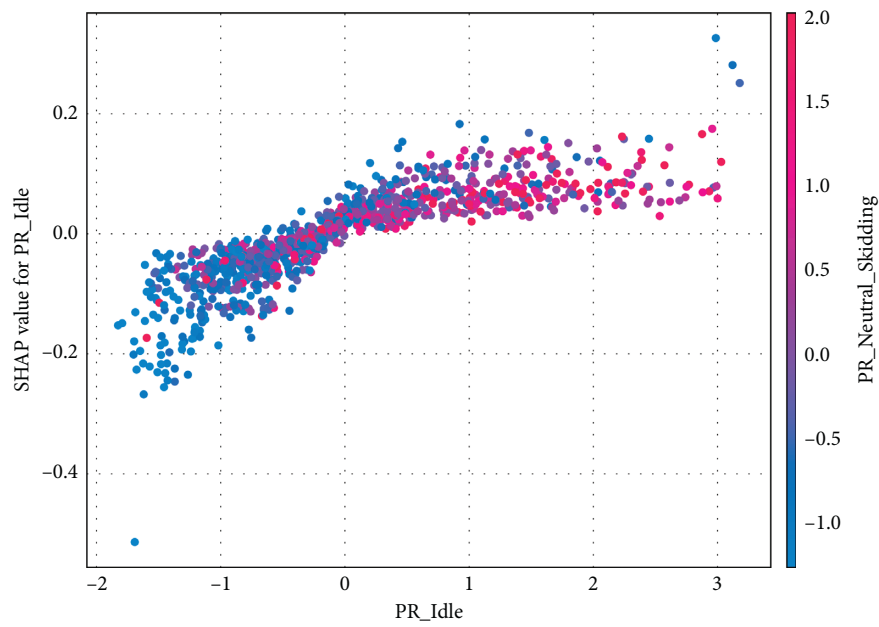FIGURE 8: The global impact of average engine speed on fuel consumption.

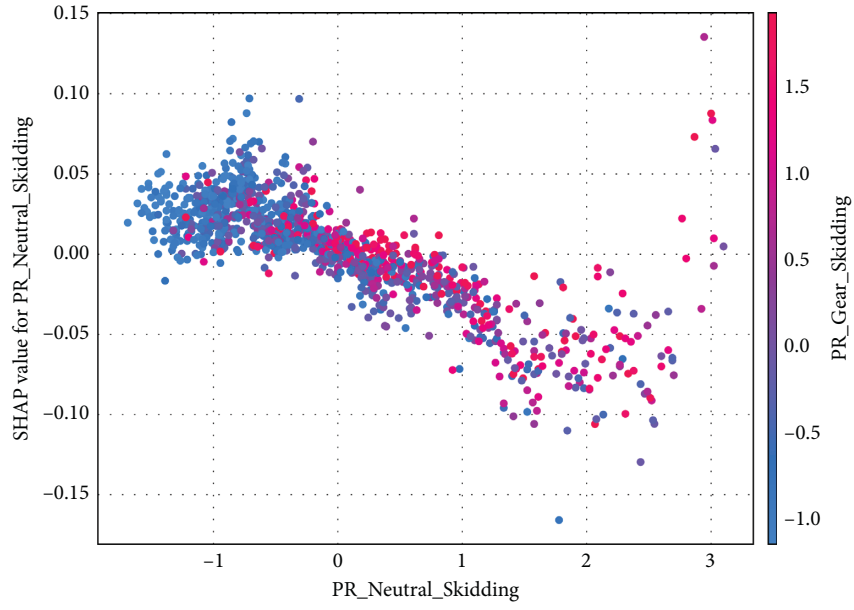FIGURE 9: The global impact of idling on fuel consumption.

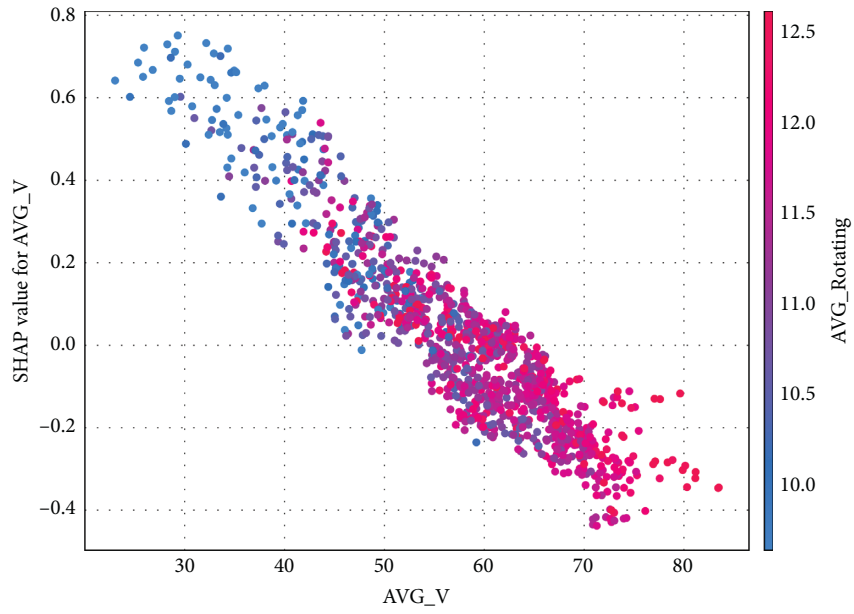Figure 10: The global impact of rolling without engine load on fuel consumption.

Figure 11: The global impact of average speed on fuel consumption.
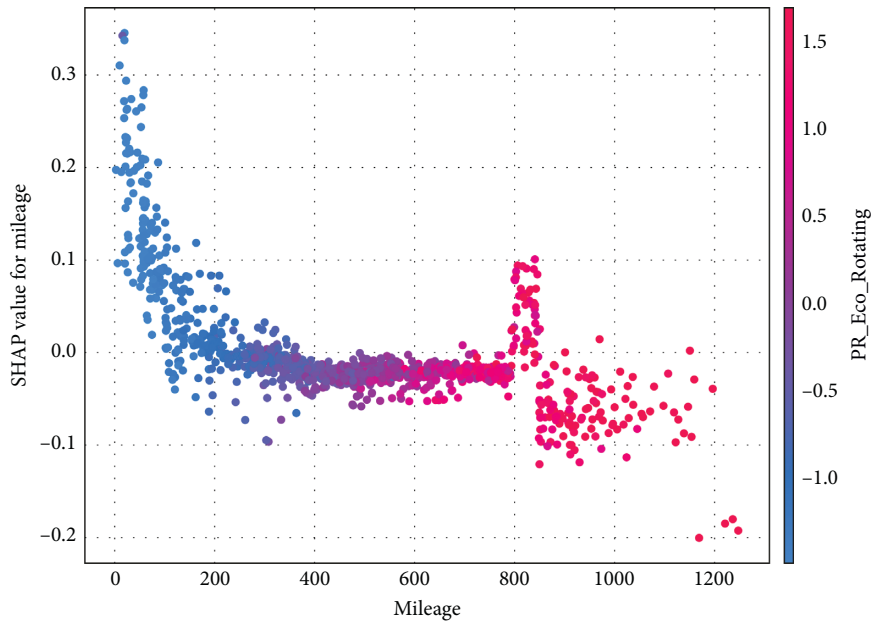
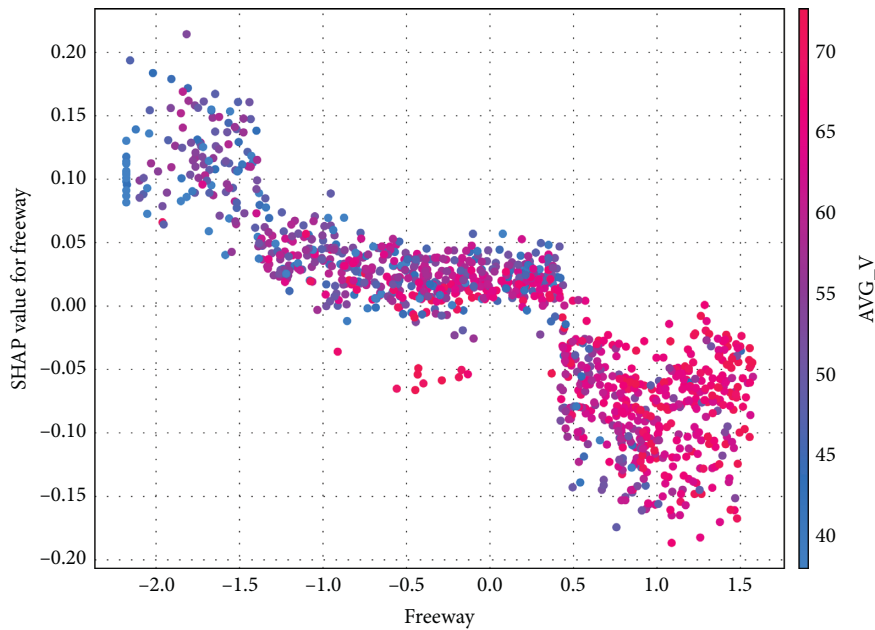FIGURE 12: The global impact of mileage on fuel consumption.

FIGURE 13: The global impact of freeways on fuel consumption.

save fuel by choosing the route with a higher percentage of freeways as far as possible [66]. The global impact of freeways on fuel consumption is provided in Figure 13.

6. Conclusions

This study performed an integral framework to demonstrate how to use the advanced on-board measurement system to analyze drivers' awareness biases on factors that affect the fuel consumption of HDDTs. By reviewing the literature, it can be seen that there is a mutually reinforcing feedback relationship between drivers' fuel-saving awareness and fuel-saving behavior. Therefore, the driver's awareness bias will not only lead to excessive fuel consumption but also weaken the effectiveness of fleet ecological management.

In the specific case and analyses, it was found that the drivers have a long driving experience, which means they are experienced and have relatively fixed driving habits. Worse still, these drivers think they are very familiar with the factors affecting fuel consumption. But in fact, through the investigation, they were found to possess large awareness biases about these factors, which hinder the development of eco-driving training and fuel-saving behavior. Note that

even with the same indicators, there are differences in the effects of ideal experimental conditions and real transport environments on fuel consumption, especially with the fact of engine losses.

To strengthen the reliability of the research, the objective data obtained from the system were used as the basis to judge the driver's subjective awareness. Finally, the main findings can be summarized as the following:

   (i) Drivers tend to overestimate the impact of economic speed on fuel consumption. According to the previous analysis, the economic speed of the engine will be changed according to different driving conditions. If the driver is too dependent on maintaining the engine speed within the original economic speed range, the effect of fuel-saving will be poor. Where possible, the range of engine economic speed under different loss levels should be further explored and the corresponding education content should be updated in driving training.

   (ii) The overestimated indicators were found to be usually related to the environment, while the underrated indicators were mostly related to driving behavior. This might be due to the survey deviation caused by social expectations. Drivers tend to attribute the causes of excessive fuel consumption to uncontrollable environmental indicators. Therefore, in future research or the application of other 3PL companies, this issue should be fully considered by improving the design of questionnaires. In addition, how to optimize the control of engine operating conditions and idling conditions should be the focus of ecological management and training of the fleet, and the overall impact of these two indicators on fuel consumption should be shown to the drivers based on data mining results.

   (iii) Although some indicators are accurately understood by drivers, some interesting rules captured by the current study are worth mentioning. Average speed is the most important indicator besides weight load, and further analysis should be carried out on fuel consumption distribution under different vehicle speeds. Compared with other roads, freeways are undoubtedly conducive to fuel-saving. The mileage presents a nonlinear effect on fuel consumption, with a local maximum of 800 km. Therefore, it is recommended that 3PL should choose the route with more freeways and control the mileage below 800 km when planning a transportation route. Furthermore, drivers should be warned not to choose low-grade roads to avoid paying freeway tolls, and it is best to have ample rest before the mileage reaches 800 km. However, the transportation time must be adequate for the customer's requirements, otherwise, it may cause greater losses.

Fleet management systems, which can collect driving behavior indicators, have the potential to increase energy efficiency in 3PL companies. In this paper, in order to enhance the portability of the research, a medium-sized 3PL company in China was used as an example. However, seriously speaking, the driver status of different 3PL companies might be different, and their fleet management systems are not the same. As such, the main contribution of this study is to provide a solution to the problem of 3PL companies to understand the driver's awareness level of fuel consumption, but to succeed, 3PL companies must act on the results and maintain their focus on energy-saving practices through eco-driving training and management attention.

Due to the limitation of data availability, we have not obtained the fuel-related data of other vehicle types, such as electric vehicles and other energy vehicles, which restricts us from carrying out larger-scale research. Another drawback is that this study only preliminarily investigated the driver's awareness of the importance of fuel-saving factors, and did not further analyze whether there is a deviation in the driver's perception of how these factors affect fuel consumption. These are issues that need to be focused on in future research. If the advanced FMS data of 3PL companies in multiple regions can be obtained in the future, the spatiotemporal instability of different fuel-saving factors can be further studied, so as to investigate their fuel-saving mechanism.

Data Availability

The data used to support the findings of this study were supplied by TopChains International Logistics Co., Ltd. under license and so cannot be made freely available. This is because the data involve third-party rights, patient privacy, and commercial confidentiality.

Conflicts of Interest

The authors declare no conflicts of interest

Acknowledgments

References

[1] International Energy Agency, *Energy Technology Perspectives 2020*, IEA Publications, Paris, France, 2020.

[2] International Transport Forum, The Carbon Footprint of Global Trade, 2016, https://www.itf-oecd.org/sites/default/files/docs/cop-pdf-06.pdf.

[3] A. McKinnon, "Green logistics: the carbon agenda," *Electronic Scientific Journal of Logistics*, vol. 6, no. 3, pp. 1–9, 2010.

[4] R. van Hoek and M. Johnson, "Sustainability and energy efficiency," *International Journal of Physical Distribution & Logistics Management*, vol. 40, no. 1/2, pp. 148–158, 2010.

[5] C. Lammgård, "Intermodal train services: a business challenge and a measure for decarbonisation for logistics service providers," *Research in Transportation Business & Management*, vol. 5, pp. 48–56, 2012.

[6] H. J. Walnum and M. Simonsen, "Does driving behavior matter? An analysis of fuel consumption data from heavy-duty trucks," *Transportation Research Part D: Transport and Environment*, vol. 36, pp. 107–120, 2015.

[7] T. Bousonville, D. Cheubou Kamga, T. Krüger, and M. Dirichs, "Data driven analysis and forecasting of medium and heavy truck fuel consumption," *Enterprise Information Systems*, pp. 1–22, 2020.

[8] A. K. Madhusudhanan, X. Na, A. Boies, and D. Cebon, "Modelling and evaluation of a biomethane truck for transport performance and cost," *Transportation Research Part D: Transport and Environment*, vol. 87, no. September, Article ID 102530, 2020.

[9] Z. Xu, T. Wei, S. Easa, X. Zhao, and X. Qu, "Modeling relationship between truck fuel consumption and driving behavior using data from Internet of vehicles," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 3, pp. 209–219, 2018.

[10] M. V. Faria, G. O. Duarte, R. A. Varella, T. L. Farias, and P. C. Baptista, "How do road grade, road type and driving aggressiveness impact vehicle fuel consumption? Assessing potential fuel savings in Lisbon, Portugal," *Transportation Research Part D: Transport and Environment*, vol. 72, pp. 148–161, 2019.

[11] F. Rosero, N. Fonseca, J.-M. López, and J. Casanova, "Effects of passenger load, road grade, and congestion level on real-world fuel consumption and emissions from compressed natural gas and diesel urban buses," *Applied Energy*, vol. 282, Article ID 116195, 2021.

[12] J. M. Ramsey, H. R. Hungerford, and T. L. Volk, "Environmental education in the k-12 curriculum: finding a niche," *The Journal of Environmental Education*, vol. 23, no. 2, pp. 35–45, 1992.

[13] D. Kahneman, A. Tversky, and A. Tversky, "Prospect theory: an analysis of decision under risk," *Econometrica*, vol. 47, no. 2, pp. 263–292, 1979.

[14] L. Sekhokoane, N. Qie, and P.-L. P. Rau, "Do consumption values and environmental awareness impact on green consumption in China?" *Cross-Cultural Design*, vol. 10281, pp. 713–723, 2017.

[15] H. W. J. Lee, Y. C. E. Lee, Y. C. E. Lee, and K. Hung Wong, "Differential equation approximation and enhancing control method for finding the PID gain of a quarter-car suspension model with state-dependent ODE," *Journal of Industrial & Management Optimization*, vol. 16, no. 5, pp. 2305–2330, 2020.

[16] H. Bulkeley, "Common knowledge? Public understanding of climate change in Newcastle, Australia," *Public Understanding of Science*, vol. 9, no. 3, pp. 313–333, 2000.

[17] T. Laureti and I. Benedetti, "Exploring pro-environmental food purchasing behaviour: an empirical analysis of Italian consumers," *Journal of Cleaner Production*, vol. 172, pp. 3367–3378, 2018.

[18] E. Kikuchi-Uehara, J. Nakatani, and M. Hirao, "Analysis of factors influencing consumers' proenvironmental behavior based on life cycle thinking. Part I: effect of environmental awareness and trust in environmental information on product

choice," *Journal of Cleaner Production*, vol. 117, pp. 10–18, 2016.

[19] T. Okada, T. Tamaki, and S. Managi, "Effect of environmental awareness on purchase intention and satisfaction pertaining to electric vehicles in Japan," *Transportation Research Part D: Transport and Environment*, vol. 67, pp. 503–513, 2019.

[20] X. Li, D. Zhang, T. Zhang, Q. Ji, and B. Lucey, "Awareness, energy consumption and pro-environmental choices of Chinese households," *Journal of Cleaner Production*, vol. 279, Article ID 123734, 2021.

[21] J. Shen and T. Saijo, "Reexamining the relations between socio-demographic characteristics and individual environmental concern: evidence from Shanghai data," *Journal of Environmental Psychology*, vol. 28, no. 1, pp. 42–50, 2008.

[22] C. Mobley, W. M. Vagias, and S. L. DeWard, "Exploring additional determinants of environmentally responsible behavior: the influence of environmental literature and environmental attitudes," *Environment and Behavior*, vol. 42, no. 4, pp. 420–447, 2010.

[23] Y. Bai and Y. Liu, "An exploration of residents' low-carbon awareness and behavior in Tianjin, China," *Energy Policy*, vol. 61, pp. 1261–1270, 2013.

[24] L. Fu, "Environmental awareness and pro-environmental behavior within China's road freight transportation industry: moderating role of perceived policy effectiveness," *Journal of Cleaner Production*, vol. 252, Article ID 119796, 2020.

[25] K. Boriboonsomsin, M. J. Barth, W. Zhu, and A. Vu, "Eco-routing navigation system based on multisource historical and real-time traffic information," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1694–1704, 2012.

[26] D. Lois, Y. Wang, A. Boggio-Marzet, and A. Monzon, "Multivariate analysis of fuel consumption related to eco-driving: interaction of driving patterns and external factors," *Transportation Research Part D: Transport and Environment*, vol. 72, pp. 232–242, 2019.

[27] C. Andrieu and G. S. Pierre, "Comparing effects of eco-driving training and simple advices on driving behavior," *Procedia - Social and Behavioral Sciences*, vol. 54, pp. 211–220, 2012.

[28] S.-H. Ho, Y.-D. Wong, and V. W.-C. Chang, "What can eco-driving do for sustainable road transport? Perspectives from a city (Singapore) eco-driving programme," *Sustainable Cities and Society*, vol. 14, pp. 82–88, 2015.

[29] H. Strömberg, I. C. M. Karlsson, and O. Rexfelt, "Eco-driving: drivers' understanding of the concept and implications for future interventions," *Transport Policy*, vol. 39, no. 2007, pp. 48–54, 2015.

[30] J.-L. Fan, Q. Wang, L. Yang, H. Zhang, and X. Zhang, "Determinant changes of consumer preference for NEVs in China: a comparison between 2012 and 2017," *International Journal of Hydrogen Energy*, vol. 45, no. 43, pp. 23557–23575, 2020.

[31] M. G. Lenné, C. C. Liu, P. M. Salmon, M. Holden, and S. Moss, "Minimising risks and distractions for young drivers and their passengers: an evaluation of a novel driver-passenger training program," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 14, no. 6, pp. 447–455, 2011.

[32] Y. Zhang, Y. Jiang, W. Rui, and R. G. Thompson, "Analyzing truck fleets' acceptance of alternative fuel freight vehicles in China," *Renewable Energy*, vol. 134, pp. 1148–1155, 2019.

[33] M. Du, L. Cheng, X. Li, and J. Yang, "Acceptance of electric ride-hailing under the new policy in Shenzhen, China: influence factors from the driver's perspective," *Sustainable Cities and Society*, vol. 61, Article ID 102307, 2020.

[34] L. Yang, Z. Feng, X. Zhao, K. Jiang, and Z. Huang, "Analysis of the factors affecting drivers' queue-jumping behaviors in China," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 72, pp. 96–109, 2020.

[35] P. Obst, K. Armstrong, S. Smith, and T. Banks, "Age and gender comparisons of driving while sleepy: behaviours and risk perceptions," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 14, no. 6, pp. 539–542, 2011.

[36] S.-M. Castritius, X.-Y. Lu, C. Bernhard, M. Liebherr, P. Schubert, and H. Hecht, "Public acceptance of semi-automated truck platoon driving. A comparison between Germany and California," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 74, pp. 361–374, 2020.

[37] F. Freuli, G. De Cet, M. Gastaldi et al., "Cross-cultural perspective of driving style in young adults: psychometric evaluation through the analysis of the multidimensional driving style inventory," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 73, pp. 425–432, 2020.

[38] W. Han and J. Zhao, "Driver behaviour and traffic accident involvement among professional urban bus drivers in China," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 74, pp. 184–197, 2020.

[39] B. Guo, L. Jin, D. Sun, J. Shi, and F. Wang, "Establishment of the characteristic evaluation index system of secondary task driving and analyzing its importance," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 64, pp. 308–317, 2019.

[40] S. Wang, S. Liu, J. Zhang et al., "A new method of diesel fuel brands identification: SMOTE oversampling combined with XGBoost ensemble learning," *Fuel*, vol. 282, Article ID 118848, 2020.

[41] M. A. Jun and J. C. P. Cheng, "Selection of target LEED credits based on project information and climatic factors using data mining techniques," *Advanced Engineering Informatics*, vol. 32, pp. 224–236, 2017.

[42] J. Nobre and R. F. Neves, "Combining principal component analysis, discrete wavelet transform and XGBoost to trade in the financial markets," *Expert Systems with Applications*, vol. 125, pp. 181–194, 2019.

[43] J. Ma, Y. Ding, J. C. P. Cheng et al., "Identification of high impact factors of air quality on a national scale using big data and machine learning techniques," *Journal of Cleaner Production*, vol. 244, Article ID 118955, 2020.

[44] E. Kamel, S. Sheikh, and X. Huang, "Data-driven predictive models for residential building energy use based on the segregation of heating and cooling days," *Energy*, vol. 206, Article ID 118045, 2020.

[45] X. Yan, J. He, C. Zhang, Z. Liu, B. Qiao, and H. Zhang, "Single-vehicle crash severity outcome prediction and determinant extraction using tree-based and other non-parametric models," *Accident Analysis & Prevention*, vol. 153, Article ID 106034, 2021.

[46] J. A. DeSimone, P. D. Harms, and A. J. DeSimone, "Predicting marital happiness and stability from newlywed interactions," *Journal of Organizational Behavior*, vol. 36, pp. 171–181, 2015.

[47] G. Dorantes Argandar, F. Tortosa Gil, and J. Ferrero Berlanga, "Measuring situations that stress Mexicans while driving," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 37, pp. 154–161, 2016.

[48] M. L. Hall-Lord, A. Skoogh, R. Ballangrud, A. Nordin, and C. Bååth, "The Swedish version of the TeamSTEPPS teamwork perceptions questionnaire (T-TPQ): a validation study," *Journal of Multidisciplinary Healthcare*, vol. 13, pp. 829–837, 2020.

[49] B. L. De Castro, F. J. Gracia, I. Tomás, and J. M. Peiró, "The safety culture enactment questionnaire (SCEQ): theoretical model and empirical validation," *Accident Analysis & Prevention*, vol. 103, pp. 44–55, 2017.

[50] S. P. Washington, M. G. Karlaftis, and F. L. Mannering, *Statistical and Econometric Methods for Transnportation Data Analysis*, Chapman and Hall/CRC, Boca Raton, FL, USA, Second edition, 2010.

[51] D. Liu, Q. Deng, Z. Ren et al., "Variation trends and principal component analysis of nitrogen oxide emissions from motor vehicles in Wuhan city from 2012 to 2017," *Science of The Total Environment*, vol. 704, Article ID 134987, 2020.

[52] J. Zhang, C.-m. Wang, L. Liu et al., "Investigation of carbon dioxide emission in China by primary component analysis," *Science of The Total Environment*, vol. 472, pp. 239–247, 2014.

[53] D. Voukantsis, K. Karatzas, J. Kukkonen, T. Räsänen, A. Karppinen, and M. Kolehmainen, "Intercomparison of air quality data using principal component analysis, and forecasting of PM10 and PM2.5 concentrations using artificial neural networks, in Thessaloniki and Helsinki," *Science of The Total Environment*, vol. 409, no. 7, pp. 1266–1276, 2011.

[54] P. Bühlmann and T. Hothorn, "Boosting algorithms: regularization, prediction and model fitting," *Statistical Science*, vol. 22, no. 4, pp. 477–505, 2007.

[55] J. Xu, M. Saleh, and M. Hatzopoulou, "A machine learning approach capturing the effects of driving behaviour and driver characteristics on trip-level emissions," *Atmospheric Environment*, vol. 224, Article ID 117311, 2020.

[56] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, 2002.

[57] V. Svetnik, T. Wang, C. Tong, A. Liaw, R. P. Sheridan, and Q. Song, "Boosting: an ensemble learning tool for compound classification and QSAR modeling," *Journal of Chemical Information and Modeling*, vol. 45, no. 3, pp. 786–799, 2005.

[58] E. Štrumbelj and I. Kononenko, "An efficient explanation of individual classifications using game theory," *Journal of Machine Learning Research*, vol. 11, pp. 1–18, 2010.

[59] A. Redelmeier, M. Jullum, and K. Aas, "Explaining predictive models with mixed features using shapley values and conditional inference trees," *Lecture Notes in Computer Science*, vol. 12279, pp. 117–137, 2020.

[60] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems*, vol. 2017, pp. 4766–4775, 2017.

[61] S. M. Lundberg, G. G. Erion, and S.-I. Lee, "Consistent individualized feature attribution for tree ensembles," 2018, http://arxiv.org/abs/1802.03888.

[62] M. Zhou, H. Jin, and W. Wang, "A review of vehicle fuel consumption models to evaluate eco-driving and eco-routing," *Transportation Research Part D: Transport and Environment*, vol. 49, no. 5, pp. 203–218, 2016.

[63] J. Díaz-Ramirez, N. Giraldo-Peralta, D. Flórez-Ceron et al., "Eco-driving key factors that influence fuel consumption in heavy-truck fleets: a Colombian case," *Transportation Research Part D: Transport and Environment*, vol. 56, pp. 258–270, 2017.

[64] E. Ericsson, "Independent driving pattern factors and their influence on fuel-use and exhaust emission factors," *Transportation Research Part D: Transport and Environment*, vol. 6, no. 5, pp. 325–345, 2001.

[65] J. Gao, H. Chen, K. Dave et al., "Analysis of driving behaviours of truck drivers using motorway tests," *Proceedings of the*

Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, vol. 234, no. 13, pp. 3103–3111, 2020.

[66] K. Ahn and H. Rakha, "The effects of route choice decisions on vehicle energy consumption and emissions," *Transportation Research Part D: Transport and Environment*, vol. 13, no. 3, pp. 151–167, 2008.