IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 November 2023, accepted 15 November 2023, date of publication 20 November 2023,
date of current version 27 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3334388

RESEARCH ARTICLE

A Framework for Travel Speed Prediction Inclusive of Service Area Dwell Times

XU LUO[1,2], FUMIN ZOU[1,2], SIJIE LUO[3], AND FENG GUO[1,2]

[1]Fujian Key Laboratory for Automotive Electronics and Electric Drive, Fujian University of Technology, Fuzhou 350118, China
[2]Renewable Energy Technology Research Institute, Fujian University of Technology, Ningde 352000, China
[3]School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu 611731, China

Corresponding author: Fumin Zou (fmzou@fjut.edu.cn)

This work was supported in part by the Renewable Energy Technology Research Institution, Fujan University of Technology, Ningde, under Grant KY310338; in part by the 2020 Fujian Province ''Belt and Road'' Technology Innovation Platform under Grant 2020D002; in part by the Provincial Candidates for the Hundred, Thousand and Ten Thousand Talent of Fujian under Grant GY-Z19113; in part by the Patent Grant Project under Grant GY-Z18081, Grant GY-Z19099, and Grant GY-Z20074; in part by the Horizontal Projects under Grant GY-H-20077; in part by the Municipal Level Science and Technology Projects under Grant GY-Z-22006 and Grant GY-Z-220230; in part by the Fujian Provincial Department of Science and Technology Foreign Cooperation Project under Grant 2023I0024; and in part by the Open Fund Project under Grant KF-X19002 and Grant KF-19-22001.

ABSTRACT Accurate modeling of travel speeds is crucial for optimizing roadway management, yet traditional methods overlook a key factor the influence of vehicle dwell times in service areas. This oversight introduces bias into speed measurements, impairing their utility for fine-grained traffic monitoring. To address this problem, we propose an innovative framework that integrates machine learning prediction of service area dwell times into travel speed calculation. We focus on a 9.3 km segment of a major highway in Fujian Province, China that includes the Qingyunshan service area. A Gradient Boosting Decision Tree model identifies vehicles entering the service area, while a Bayesian Backpropagation Neural Network predicts their dwell time. By adjusting the overall travel times using these predicted dwell times, our approach recovers normal driving behavior outside service areas. Experiments on electronic toll collection data from over 17 million transactions validate the framework's effectiveness. The corrected travel speeds better reflect typical highway conditions and enable more precise assessment of traffic state across multiple time horizons. This study highlights the vital role of service area dwell time in travel speed modeling. Our solution provides a promising direction to enhance the fidelity of current prediction practices.

INDEX TERMS Travel speed, machine learning, service area, dwell time, traffic state.

I. INTRODUCTION

Travel speed is a critical metric in road traffic analysis [1], commonly used to quantify the average velocity of vehicles over specific roadway segments. This metric serves multiple functions: it can gauge a driver's behavior, such as speeding or slow driving [2], [3], thereby evaluating the safety of their actions [4], [5]. Additionally, it can indicate the overall state of traffic higher speeds often suggest smoother traffic flow, whereas lower speeds typically denote congestion [6], [7]. Thus, travel speed is an indispensable factor in the realm of transportation studies.

Accurate prediction of travel speed stands as a cornerstone for an array of applications in the realm of transportation studies [8], [9], as highlighted in the preceding paragraph. Traditional methods primarily focus on sections of roads equipped with Interval Speed Enforcement System (ISES), leaving vast stretches of roadways unmonitored. However, the proliferation of Electronic Toll Collection (ETC) gantries across highways offers an unprecedented opportunity to bridge this gap. By leveraging the ubiquity of ETC devices, we can glean insights into traffic conditions on hitherto unmonitored sections, facilitating a holistic management of the entire highway network rather than isolated segments [12]. This granularity in data not only advances road safety by offering drivers precise road conditions but also empowers highway authorities with actionable intelligence [10]. They can pinpoint and rectify anomalies in specific road sections, paving the way for enhanced traffic flow and highway service quality [11].

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

The field of travel speed prediction has a long history and has seen substantial advancements recently, largely propelled by innovations in big data and machine learning technologies [13], [14]. Generally, travel speed prediction methodologies fall into two primary categories [15]. The first involves traditional methods, which rely on empirical models, statistical analyses, and traffic flow theories. These approaches often demand extensive domain knowledge and might be less suited for complex or large-scale data sets. The second category encompasses machine learning and deep learning techniques, which have gained prominence owing to enhanced computational capabilities and burgeoning data volumes. These methodologies are adept at handling large and varied data sets and can capture complex, nonlinear relationships [16], [17]. Due to their ease of implementation and broad applicability, we have chosen these as our foundational tools for travel speed prediction.

However, travel speed prediction is not without its complexities [18]. One particular challenge that remains underexplored is the impact of service areas on travel speeds, especially on highways. These areas allow drivers to rest, eat, and refuel, potentially leading to extended stops and consequently affecting the speed measurements of vehicles that enter these areas. This is an important aspect that has been overlooked in previous studies. Accurately modeling the fluctuations in speed across space and time is crucial for various transportation applications, from route planning to infrastructure improvements. However, overlooking dwell times can obscure the true underlying patterns [19]. For instance, speeds may appear consistently low near a service area simply due to long dwell times, even if traffic is actually flowing smoothly outside the service zone. Without adjusting for dwell times, issues like congestion hotspots could be masked. Likewise, the timing and duration of speed reductions may be skewed if dwell times are not considered. By integrating dwell time estimation into speed predictions, we can effectively filter out this influence to reveal authentic spatiotemporal trends. The corrected speeds should thus better expose where and when traffic speeds change, enhancing monitoring and planning.

In light of this, our study introduces a tailored framework to address this often-overlooked factor. We centered our analysis on a designated road segment featuring a service area, gleaning transaction data for all transiting vehicles. In our methodology, a Gradient Boosting Decision Tree (GBDT) model discerns vehicles that enter the service area. Concurrently, a Bayesian Backpropagation Neural Network (Bayes-BPNN) predicts the duration of their dwell time. We then refine travel speed metrics: the dwell time is subtracted from the total travel time, yielding an adjusted travel duration. This novel speed calculation method offers a more accurate reflection of actual driving speeds, considering the influence of dwell times. Our empirical validation is rooted in data from the Qingyunshan service area in Fujian Province. The analysis delves deep into the recalibrated speeds and their implications for traffic condition evaluations. The outcomes not only underscore the efficacy of our approach but also spotlight the pivotal role of dwell time in shaping travel speed metrics.

The principal contributions of this paper are as follows:

1) Demonstrates the vital role of service area dwell time as a determinant of vehicle travel speed.
2) Provides a solution to recover authentic speeds by adjusting for estimated dwell times.
3) Enables more accurate assessment of traffic conditions using corrected speed metrics.
4) Introduces an innovative framework to significantly improve current speed prediction practices.

The rest of the paper is organized as follows: Section II reviews related studies on trip speed prediction and provides a comprehensive overview of related studies on service area approach and dwell time. Section III describes our methodology and model in detail. Section IV outlines our experimental design and execution. In Section V, we analyze the experimental results and provide an in-depth discussion. Finally, Section VI summarizes our study and suggests future research directions.

II. LITERATURE REVIEW

Travel speed prediction is crucial for understanding vehicle driving behavior and evaluating road traffic conditions. Prior studies have sought to enhance the accuracy of such predictions by considering a multitude of factors. For instance, Weng et al. [20] examined the relationship between the speeds of different vehicle types and the volume of data. Zahid [21] investigated the influence of various time domains on traffic speed prediction. Ambrose Wei [22] employed Pearson's correlation coefficient to demonstrate a strong relationship between the average speeds of vehicles in different sections of a service area. Malek et al. [23] focused on key factors such as road speed limits, gradients, and curve curvatures. Wang et al. [24] quantitatively assessed the impact of rainfall intensity and traffic flow on travel speed. Meanwhile, Xu et al. [25] studied the speed distribution of different vehicle models on identical road segments and used statistical characteristics to predict speeds within those segments.

While prior research has integrated various elements, such as sample size, vehicle attributes, time domains, road characteristics, and weather conditions, into their predictive models, there remains a noticeable gap: the neglect of dwell time in service areas as a contributing factor. This oversight is particularly significant given that dwell time is a key variable affecting travel speed.

The first step in acquiring dwell time data is to predict whether a vehicle will enter a service area. Historically, research in this domain has aimed to optimize both the scale and operational efficiency of service areas. Various studies have deployed diverse methods and focused on a range of influencing factors to make this prediction. For instance, Cui and Liu [26] conducted a quantitative analysis on the link between continuous vehicle travel time and the likelihood of

entering a service area. Wang and Tang [27] formulated a potential traffic model that accounts for cross-sectional traffic flow, location variables, and access characteristics to predict vehicle entry into highway service areas.

In a similar vein, Liu et al. [28] considered an array of factors such as the distance between service areas, cross-sectional traffic flow, and attributes of the service areas themselves, including the number of parking spaces, restaurant size, and the regional economy. Liu [29] introduced an elasticity factor method aimed at predicting the current drive-in rate for service areas. Additionally, Ou [30] employed an enhanced case-based inference model, which incorporated traffic volume, roadway network levels, the spacing between neighboring service areas, and urban characteristics to forecast vehicle entry into service areas.

Most existing research on dwell time prediction concentrates on public transportation systems. For example, in the realm of high-speed rail, Hou et al. [31] developed a delay recovery model and employed a gradient-enhanced tree algorithm for accurate train dwell time prediction. In the subway sector, Arlovski et al. [32] applied logarithmic transformations to dwell time data to account for its nonlinearity. In the context of bus transportation, Huo [33] crafted a prediction model based on nearest neighbor and support vector machine techniques, while Isukapati et al. [34] utilized a Bayesian hierarchical approach to derive dwell time distributions from historical data.

However, these models, tailored for public transportation, may not readily apply to passenger vehicles, which are primarily privately owned and thus exhibit different dwell characteristics. A thorough review of the literature reveals a noticeable gap: no existing studies have explored the dwell time characteristics of passenger vehicles in service areas. Our research aims to fill this gap by being the first to investigate the dwell time features of passenger vehicles in service areas. Additionally, we provide a qualitative analysis of how dwell time in these areas impacts travel speed.

III. METHODOLOGY

Travel speed prediction is critical for evaluating road traffic conditions. To enhance the accuracy of these predictions, this study investigates the impact of dwell time in service areas on travel speed. One challenge we faced is the disparate data systems for Electronic Toll Collection (ETC) and service areas; the current ETC data fields alone do not indicate whether a vehicle has entered a service area, nor can the service area data confirm if a vehicle is part of the ETC system. To overcome this limitation, we merged the two datasets, creating a labeled set that can facilitate the use of supervised machine learning algorithms for both training and prediction of unknown instances. This fusion of datasets allows us to better understand the traffic characteristics of vehicles entering service areas and their subsequent impact on travel speed.

A. PROBLEM DEFINITION

1) SECTION

A large number of ETC gantries are laid along the highway network, each ETC gantry is regarded as a node, and a zone is defined between two nodes, i.e., between two gantries $Sec = \{Q, Dis\}, Q = (Node1, Node2)$, Where Node1 is the starting node, Node2 is the end node, and Q represents a zone consisting of two nodes, Node1 and Node2, Dis is the mileage of the zone.

2) TRAVEL TIME

From the time the vehicle enters the zone to the time the vehicle leaves the zone, the time the vehicle passes through the entire zone is defined as the travel time T_{traj},which can be calculated from the transaction time recorded by the ETC front and rear gantries: $T_{traj} = t_{Node2} - t_{Node1}$

3) DWELL TIME

From the time the vehicle enters the service area to the time the vehicle leaves the service area, the time the vehicle passes in the service area is defined as the dwell time T_{dwell}

4) TRAVELLING TIME

The vehicle in the non-service area section is in the driving state, the time the vehicle is in the driving state, defined as the travelling time T_{trav}

5) DEFINITION OF TRAVEL SPEED PROBLEM

In the past, travel speed took into account the time the vehicle spent in the entire zone, which is obviously fine for most roads. However, this is not the case for roads containing service areas, where the vehicle is likely not to be in a driving state when it enters the service area. Therefore, if the dwell time in the service area is also taken into account, there is a great error. We define a new formula for calculating travel speed, which is determined by Equation 1.

$$T_{traj} = T_{trav} + T_{dwell} \tag{1}$$

where,T_{trav} is the time spent in the driving state and T_{dwell} is the dwell time in the service area. Hence, the travel speed for a single vehicle is defined as:

$$V_{ij} = \frac{Dis}{T^{ij}_{trav}} = \frac{Dis}{T^{ij}_{traj} - T^{ij}_{dwell}} \tag{2}$$

where *Dis* is the mileage of the road segment, i represents the vehicle type, and j represents the vehicle serial number.

B. TRAVEL SPEED PREDICTION FRAMEWORK

The framework for travel speed prediction in this study is depicted in Figure 1 and consists of three main stages, in addition to the data fusion process previously described. First, a feature engineering process is conducted using the fused dataset, which includes attributes such as vehicle type, time domain, road characteristics, and continuous driving time. Second, a supervised machine learning algorithm is employed to construct a predictive model for vehicles

entering the service area. The output of this model, namely the identification of vehicles entering the service area, serves as input for a subsequent dwell time prediction model. Finally, the corrected travel speed is calculated using our newly-defined methods, as outlined in equations 1 and 2.

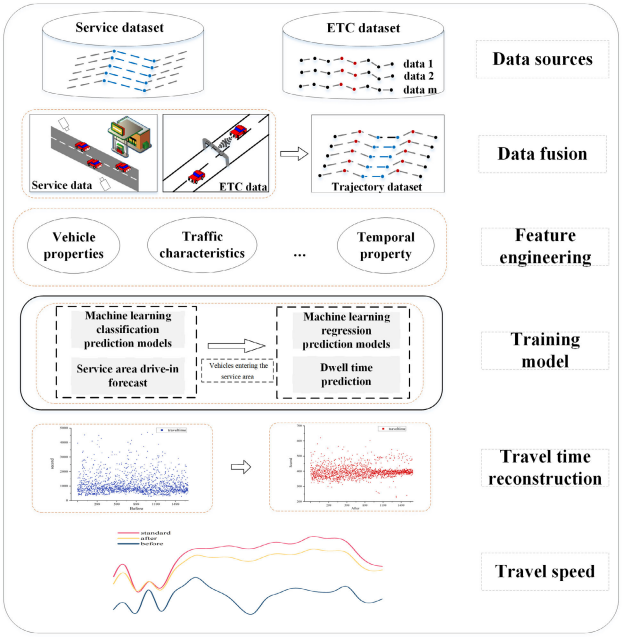


FIGURE 1. Travel speed prediction framework inclusive of service area dwell times.

C. SERVICE AREA DRIVE-IN RECOGNITION

The behavior of highway vehicles driving into service areas is a typical binary classification problem, and both traditional machine learning methods and deep learning methods can effectively solve the binary classification problem. However, traditional machine learning methods have the advantages of high interpretability, mature technology, and fast training speed, while deep learning methods require more computational resources in addition to weak interpretability. As the most dominant machine learning method in tabular data, gradient-boosting decision tree (GBDT) has the advantages of high interpretability and fast training speed. It also has a decision manifold with approximate hyperplane boundaries, which is very effective for tabular data.

It is an iterative decision tree algorithm [35], which is based on the categorical regression tree algorithm. GBDT is based on the idea of boosting iteration, except the first decision tree is generated by the original predictor, the goal of each iteration is to minimize the loss function of the current learner, i.e., the loss function continuously decreases along its gradient direction, and the final residual converges to 0 through continuous iterations. the algorithm can be regarded as an additive model composed of M trees, and its corresponding equation is as follows. As shown in equation 3.

$$F(x,\omega)=\sum_{m=1}^{M}\alpha_m h_m(x,\omega_m)=\sum_{m=1}^{M} f_m(x,\omega_m) \tag{3}$$

where x is the input sample, ω is the model parameter, h is the classification regression tree, and α is the weight of each tree. $F(x, w)$ represents the GBDT model, which is a sum of M trees. Each tree, $h_m(x, \omega_m)$ or $f_m(x, \omega_m)$, has its own model parameter ω_m optimized during training.

The weights α_m are not arbitrarily assigned; they are learned during the training process. In the gradient boosting framework, each tree's weight, α_m, is determined by solving the following optimization problem:

$$\alpha_m=\arg\min_{\alpha}\sum_{i=1}^{n} L\left(y_i, F_{m-1}(x_i)+\alpha h_m(x_i)\right) \tag{4}$$

Here, $L(y_i, F_{m-1}(x_i)+\alpha h_m(x_i))$ is the loss function, and the goal is to find the α_m that minimizes this loss. This process is done iteratively for each tree in the ensemble.

Thus, for a given training data set $T=\{(x_1,y_1),\cdots(x_n,y_n)\}$ containing n samples, $x_i\in X$, X is the input space, $y_i\in Y$ and Y is the output space. Its GBDT modeling process is as follows, using supervised learning to give an estimation function $\hat{f}(x)$ for the true function $f:x\to y$ and minimizing the loss function $L=(y,\hat{f}(x)))$ to improve the accuracy of prediction, as shown in equation 5.

$$\hat{f}(x)=\arg\min_{f(x)} L(y,\hat{f}(x)) \tag{5}$$

Equation 5 minimizes the expected loss form as shown in equation 6:

$$\hat{f}(x)=\arg\min_{f(x)} E_x\left[E_y[L(y,\hat{f}(x))]\mid x\right] \tag{6}$$

To specify the target problem, the search space is now restricted by the parameter θ, as shown in equation 7.

$$\hat{\theta}=\arg\min_{\theta} E_x\left[E_y[L(y,\hat{f}(x,\theta))]\mid x\right] \tag{7}$$

So far, no specific formal assumptions have been made about the estimated and true functions, and in most cases there is no closed form solution to the problem described by the above equation, so recursive numerical process optimization is usually used.

The loss functions used to perform optimization are squared loss and exponential loss, such loss functions are relatively simple and can be optimized by using the general Boosting algorithm. But for general loss functions, it is difficult if the usual optimization methods are used, and for this problem, Freidman [36] used the value of the loss function in the negative gradient direction of the current model, as shown in equation 8.

$$r_{m,i}=-\left[\frac{\partial L\left(y_i,F(x_i)\right)}{\partial F(x)}\right]_{F(x)=F_{m-1}(x)} \tag{8}$$

Optimization is continuously performed along the gradient direction of the gradient loss function to achieve improved prediction model performance.

GBDT is an algorithm for recursively solving the predictive model, and at each stage m of the solution, one can start

with an imperfect model $F_m(x)$ and then obtain a better model by adding an estimator $h(x)$ to $F_m(x)$, as shown in equation 9.

$$F_{m+1}(x) = F_m(x) + h(x) \tag{9}$$

According to the empirical principle of minimization of risk, it is obtained that:

$$F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma) \tag{10}$$

$$F_m(x) = F_{m-1}(x) + \arg\min_{h} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + h_m(x_i)) \tag{11}$$

The gradient descent method is then used on the loss function minimization problem to update the model according to equations 12 and 13.

$$F_m(x) = F_{m-1}(x) - \gamma_m \sum_{i=1}^{n} \nabla_{F_{m-1}} L(y_i, F_{m-1}(x_i)) \tag{12}$$

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L\left(y_i, F_{m-1}(x_i) - \nabla_{F_{m-1}} L(y_i, F_{m-1}(x_i))\right) \tag{13}$$

The above solution methods are subject to overfitting and overprediction in some cases, we generally use regularization technique to reduce the overfitting effect by controlling the fitting process, so the update rule of the above algorithm is modified as follows.

$$F_m(x) = F_{m-1}(x) + \nu \cdot \gamma_m h_m(x), \quad 0 < \nu < 1 \tag{14}$$

where ν is the learning rate, and a smaller learning rate significantly improves the generalization ability of the model, but increases the number of iterations at the same time.

D. ESTIMATED SERVICE AREA DWELL TIME

In order to eliminate the effect of dwell time, after the identification of vehicles entering the service area, it is necessary to further predict the dwell time of these vehicles in the service area, with the aim of obtaining information on the dwell time of vehicles in the service area. Back Propagation Neural Network (BPNN) is a multilayer feed-forward neural network, which consists of an input layer, an output layer and at least one hidden layer. Unlike polynomial regression, which assumes a specific polynomial relationship between input features and output, BPNN can model the complex nonlinear relationship between input features and target variables, and can effectively capture the complex interactions between features. Therefore, BPNNs are well suited for supervised learning problems that model input-output relationships. On the other hand, deep reinforcement learning is commonly used for decision and control tasks, where agents learn to perform operations in the environment to achieve specific goals while maximizing the cumulative reward. Compared with the DRL algorithm, BPNN has the advantages of relatively simple implementation and training, lower computational cost, and higher computational efficiency.

The main feature of this network is that the signal is transmitted forward and the error is propagated backward [37]. In forward transmission, the input signal is processed from the input layer through the implicit layer layer by layer until the output layer. The neuron state in each layer only affects the neuron state in the next layer, and if the desired output is not obtained in the output layer, it is transferred to back propagation, and the network weights and thresholds are adjusted according to the prediction error, so that the predicted output of the BP neural network continuously approximates the desired output. Therefore, BP networks can be regarded as tools for solving function approximation [38].

However, as with most deep learning or neural network methods, the number of hidden layers and neuron nodes is generally unknown. To address this problem, we first set the network to have two hidden layers, and the number of neuron nodes is determined by Bayesian optimization techniques. The hyperparameter optimization method is described in detail in the section IV-E, and this subsection focuses on the computational steps of the network.

The forward propagation process is only used to calculate the output of the network and does not adjust the parameters of the network, which is deduced as follows.

$$z^{[l]} = W^{[l]} a^{[l-1]} + b^{[l]} \tag{15}$$

$$a^{[l]} = g^{[l]}\left(z^{[l]}\right) \tag{16}$$

Vectorized implementation:

$$Z^{[l]} = W^{[l]} \cdot A^{[l-1]} + b^{[l]} \tag{17}$$

$$A^{[l]} = g^{[l]}\left(Z^{[l]}\right) \tag{18}$$

where $a^{[l]}$ represents the output of the node in layer l and $a^{[l-1]}$ is the output of the node in the previous layer and the input of the node in the current layer. $A^{[l]}$ represents the output of layer l, and $b^{[l]}$ represents the bias of layer l.

Cost function:

$$J(w, b) = \frac{1}{m} \sum_{i=1}^{m} J\left(w, b; x^{(i)}, y^{(i)}\right) \tag{19}$$

The backward conduction process is used for the adjustment of network weights and thresholds during training, using the gradient descent algorithm to find the minimal value of the cost function so that the error between the expected value and the output is as small as possible. Thus, the core process is as follows.

$$w_{ij} = w_{ij} - \alpha \frac{\partial J}{\partial w_{ij}} \tag{20}$$

$$b_{ij} = b_{ij} - \alpha \frac{\partial J}{\partial b_{ij}} \tag{21}$$

Vectorized implementation:

$$w^{(l)} = w^{(l)} - \alpha \frac{\partial J}{\partial w^{(l)}} \tag{22}$$

$$b^{(l)} = b^{(l)} - \alpha \frac{\partial J}{\partial b^{(l)}} \tag{23}$$

where α is the learning rate, $\frac{\partial J}{\partial w^{(l)}}$ is the partial derivative with respect to ω, $\frac{\partial J}{\partial b^{(l)}}$ is the partial derivative with respect to b.

Therefore, calculating the partial derivatives with respect to ω and b allows iterative updating, which completes the updating of the weights and the adjustment of the parameters, thereby realizing the whole algorithm.

E. DRIVING BEHAVIOR REHABILITATION

After completing the above stages of computation, we will recover the driving behavior of this part of the vehicles driving into the service area, for which new formulas have been given in the problem definition subsection. If the vehicle does not enter the service area, i.e., $T_{dwell} = 0$, then Equations 1 and 2 are updated as follows.

$$T_{trav}^{ij} = T_{traj}^{ij} - T_{dwell}^{ij} = t_{Node2}^{ij} - t_{Node1}^{ij}, \quad T_{dwell}^{ij} = 0 \tag{24}$$

$$V_{ij} = \frac{Dis}{T_{trav}^{ij}} = \frac{Dis}{T_{traj}^{ij} - T_{dwell}^{ij}} = \frac{Dis}{T_{traj}^{ij}}, \quad T_{dwell}^{ij} = 0 \tag{25}$$

In terms of theoretical analysis, our proposed travel speed calculation formula has stronger applicability. For vehicles that have entered the service area, we eliminate the effect of their service area dwell time by reconstructing the travel time; for vehicles that have not entered the service area, the vehicle travel time is equivalent to the driving time, which does not affect the travel speed calculation.

IV. CASE STUDY

A. DATA PRESENTATION

The experimental dataset used in this study is supplied by Fujian Highway Information Technology Co., Ltd. It comprises a fusion of the Electronic Toll Collection (ETC) gantry transaction dataset and the Qingyunshan service area dataset. The ETC dataset contains approximately 17 million transactions recorded from September 3 to September 5, 2020. This dataset includes various fields such as license plate numbers, trip IDs, transaction times, gantry numbers, and vehicle types, as detailed in Table 1. On the other hand, the Qingyunshan service area dataset captures vehicle information from September 3 to September 12, 2020. It includes fields like license plate numbers, vehicle types, timestamps for entry and exit, and more, amounting to a total of 38,675 data points. Some of this data is presented in Table 2. It should be noted that due to data sensitivity concerns, all privacy-sensitive information has been anonymized.

The service areas and road sections involved in this study are shown in Figure 2, and we used three anchor points to emphasize the study area. The front and rear anchor points represent the distribution locations of the gantry nodes, and the road section between them is the study road section, totaling 9.328 km. the middle anchor point is used to characterize the location of the Qingyunshan service area, which is located within the road section. On average, this segment sees a daily traffic flow of 3,000 vehicles, experiencing peak hours between 9:00 a.m. and 11:00 a.m. The vast majority of these vehicles are passenger cars. Notably, 17% of the entire vehicle count opts to enter the service area. On this road section, vehicles typically travel at an average speed of 89 km/h, translating to an average travel duration of about 479 seconds.

FIGURE 2. Study area of the QingYunShan service area.

B. DATA PREPROCESSING

Because the ETC transaction data and the service area data come from different systems, we fused these datasets using the unique identifier of the license plate number. This fusion enables us to tag the vehicles entering the service area, thereby creating a comprehensive training dataset. However, the data fusion process introduces some inconsistencies, including null values, outliers, and duplicate entries. Strategies for handling these issues are discussed in the following section.

Most null values in our dataset originate from the service area data, as illustrated in Table 3. The data is collected via image recognition, which can be compromised by external factors like weather conditions, lighting, and physical obstructions. Statistical analysis revealed that the recognition rate at the entrance of the service area is around 69%, while at the exit, it improves to approximately 97%. This discrepancy indicates that nearly one-third of entrance data is missing. While the exit data enables us to identify which vehicles have entered the service area, the absence of corresponding entrance data poses a significant challenge for subsequent dwell time prediction. To address this, we opted to remove entries with missing values.

The majority of anomalous data originates from the ETC gantry transaction records, as detailed in Table 4. For this subset of data, the transaction time recorded at the front gantry often lags behind that of the subsequent gantry, resulting in a negative travel time. Such instances are clearly inconsistent with actual travel patterns. The reasons for these anomalies are multifaceted; they may arise from non-compliance with traffic regulations by the drivers, or be due to erroneous readings from the gantry equipment interacting with on-board vehicle systems on adjacent roads. To maintain the integrity of our dataset, we have elected to remove these anomalous entries.

TABLE 1. Partial ETC transaction data table fields.

Entime	Tradeime	Flagid	Obuplate	Vehclass
2020-09-02 23:39:56	2020-09-03 00:10:28	340827	***	14
2020-09-02 23:43:31	2020-09-03 00:00:37	340825	***	1
2020-09-02 23:43:37	2020-09-03 00:01:38	340825	***	1
2020-09-02 23:50:14	2020-09-03 00:12:16	340827	***	11

TABLE 2. Partial service area data table fields.

entrance/exit	License plate	Vehicle Class	Capture time
entrance	***	Minibuses	2020-09-03 00:00:11
entrance	***	Minibuses	2020-09-03 00:02:00
exit	***	Minibuses	2020-09-03 00:00:57
exit	***	Minibuses	2020-09-03 00:07:48

TABLE 3. Service area partially without license plate data.

entrance/exit	License plate	Vehicle Class	Capture time
entrance	nan	Minibuses	2020-09-03 00:25:38
entrance	nan	Minibuses	2020-09-05 23:37:16
entrance	nan	Minibuses	2020-09-05 23:43:42
entrance	nan	Minibuses	2020-09-05 23:55:42

Duplicate entries primarily arise during the fusion of the two datasets. To maintain data integrity, we retain only one instance of each duplicate and remove all subsequent repetitions.

C. EXPLORATORY ANALYSIS

According to China's vehicle type classification standards, it can be divided into four categories: passenger vehicles, freight vehicles, special operation vehicles, and other vehicles, of which the first three categories are divided into six categories according to different rules. The ETC data set, including four model categories, involves sixteen specific types of vehicles. Among them, type I vehicles (passenger vehicles) accounted for nearly 90%, type II vehicles (freight vehicles) accounted for almost 10%, type III (special operation vehicles) and type IV (other vehicles) vehicles accounted for almost nothing, specific models accounted for the distribution of Figure 3. can be seen, there is a severe imbalance between the models, increasing the difficulty of analysis of a smaller number of models.

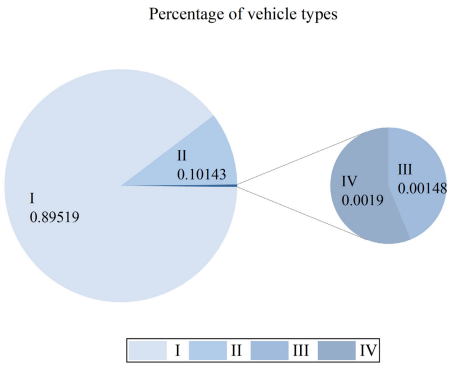


FIGURE 3. Analysis of the composition of traffic flow.

To better understand the patterns of vehicles entering the service area while in-transit on the highway, we undertook a detailed count of all vehicles. Of these, 17.11% chose to enter the service area. We then categorized the vehicles into three predominant types found on the highway: private cars, passenger vehicles, and freight vehicles. Our analysis revealed that 18.23% of private cars, 4% of passenger vehicles, and 10.31% of freight vehicles opted to enter the service area. A visual representation of these findings can be viewed in Figure 4.

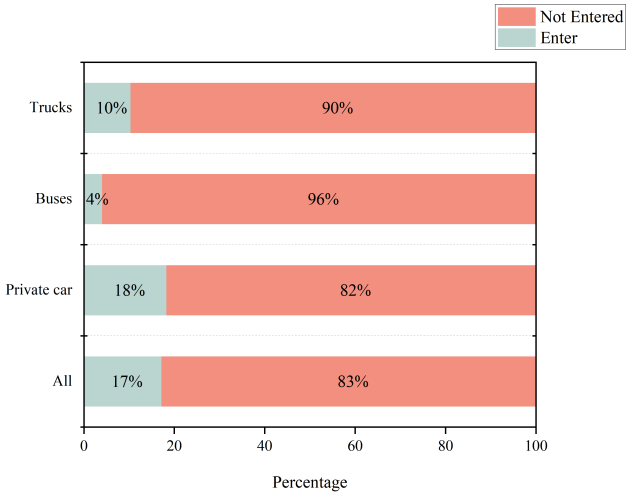


FIGURE 4. Analysis of vehicle types entering the service area.

The percentage of each vehicle type entering the service area determines what percentage of vehicles exhibit dwell time behavior. By constructing a probability density function (PDF) of dwell time for these vehicles, it is possible to identify and infer dwell time patterns, especially when sufficient historical data is available. The combined use of the entry percentage and dwell time distribution helps provide a clearer understanding of travel speeds between highway zones containing service areas.

For this study, we focused on the three predominant categories of highway vehicles: private, passenger, and freight. Using Kernel Density Estimation (KDE), we generated their respective PDFs for dwell times. To confirm the robustness of these PDFs, we employed quantile-quantile (Q-Q) plots, comparing our data's distribution against the standard normal distribution. As visualized in Figure 5, dwell times across all vehicle types exhibit a positive skew. This suggests that while prolonged dwell times occur, they are infrequent. This aligns with the understanding that traffic-related human behavior is generally consistent, but exceptions do arise.

TABLE 4. Partial abnormal data table fields.

Obuplate	Vehclass	Pretime	Nextime	Trajtime
***	0	2020-09-03 04:07:45	2020-09-03 04:02:07	-338.0
***	0	2020-09-03 19:40:51	2020-09-03 19:35:12	-339.0
***	0	2020-09-03 23:34:13	2020-09-03 22:56:01	-2292.0
***	0	2020-09-04 15:58:55	2020-09-04 15:53:13	-342.0

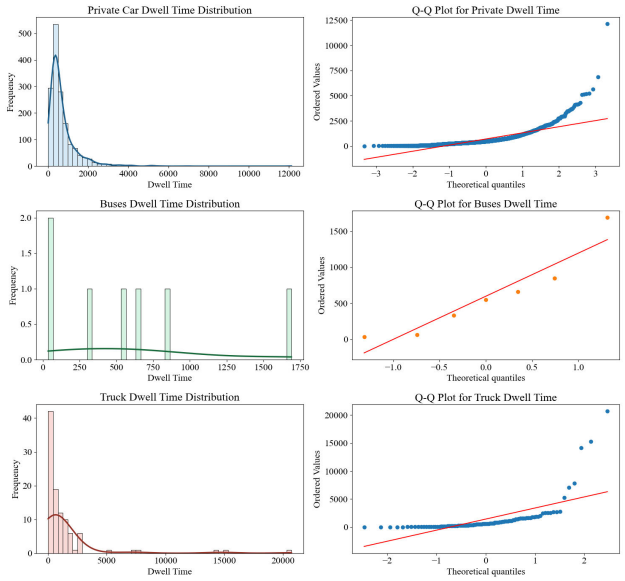


FIGURE 5. PDF based on KDE of dwell time for different classes of vehicles.

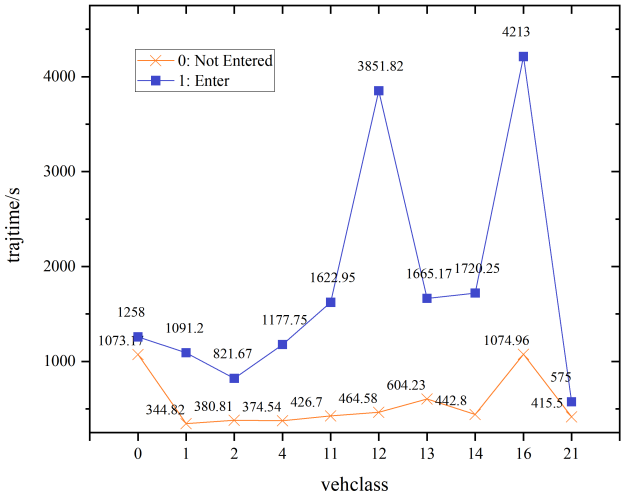


FIGURE 6. Travel time for similar type of vehicles entering and not entering the service area.

Based on the principles of travel speed measurement, travel time serves as the determining factor affecting interval speed. To analyze this, we compare the travel times of various vehicle types, both those entering and not entering the service areas within the same road segment. In Figure 6, a '0' denotes vehicles that do not enter the service area, while a '1' indicates those that do. The figure reveals that, for any given vehicle type, those entering the service area generally exhibit longer travel times compared to those that do not. Moreover, freight vehicles (types 11, 12, 13, 14, and 16) typically have longer travel times than passenger vehicles (types 0, 1, 2, and 4). This is primarily attributed to the heavier loads and correspondingly slower speeds of freight vehicles.

Figure 7 illustrates the travel speeds of various vehicle types, segregating those that enter service areas from those that do not. The numbers "0" and "1" color-coded on the graph denote vehicles not entering and entering the service areas, respectively. The vertical axis categorizes the different vehicle types. The left-hand side of the graph represents vehicles bypassing service areas, while the right-hand side represents those that enter. A clear pattern emerges: regardless of vehicle type, those that avoid service areas travel at significantly higher speeds than those that enter. Specifically, the average travel speed for vehicles entering service areas typically falls below 60 km/h, while vehicles bypassing these areas average speeds greater than 60 km/h. Additionally, passenger vehicles (types 0, 1, 2, and 4) generally travel faster than freight vehicles (types 11, 12, 13, 14, and 16).

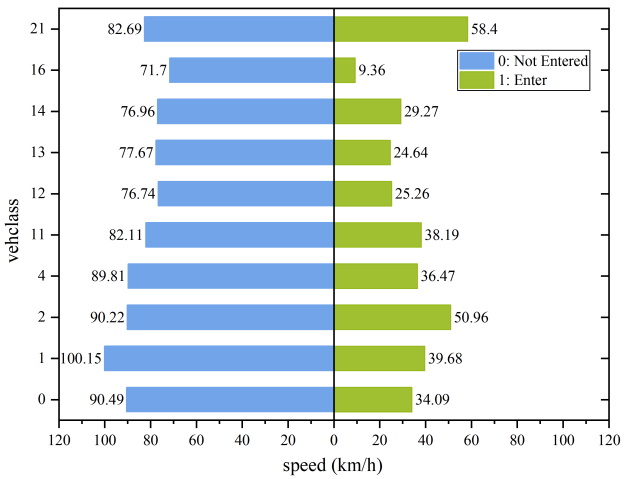


FIGURE 7. Travel speed of similar type of vehicles entering and not entering the service area.

D. COMPETITIVE ALGORITHM

In this study, we take a two-step approach to model the driving behavior of vehicles entering service areas and to estimate their realistic driving speeds. First, we construct a vehicle identification model focused on determining whether or not a vehicle will enter a service area. This is a binary classification task. For this purpose, we employ a variety of popular classification algorithms, including XGBoost, LightGBM,

SVM, Naive Bayes, KNN, and AdaBoost, among others. It's worth noting that we used the default parameters for these algorithms without hyperparameter tuning.

Second, we build a dwell time prediction model aimed at estimating the length of time vehicles will spend in the service area. This task is a regression prediction problem. To address it, we use state-of-the-art learning models specifically designed for tabular data, including TabNet, XGBoost, LightGBM, as well as ResNet (Residual Neural Network) and polynomial regression models. While this summary briefly outlines several models, it is not an exhaustive list.

TabNet [39] is a neural network architecture designed for tabular data. It was proposed by Google researchers in 2019 for regression tasks and traditional categorization tasks.

LightGBM [40] is a gradient boosting framework proposed by Microsoft researchers in 2017, using a histogram optimization algorithm with efficient computational efficiency.

XGBoost [41] is suitable for a wide range of predictive modeling and machine learning problems involving tabular datasets, including classification, regression etc. It is widely used by data scientists.

ResNet [42] is a convolutional neural network architecture. It is widely used for regression predict. due to its ability to train very deep networks accurately.

AdaBoost (Adaptive Boosting) is an ensemble machine learning algorithm known for its ability to boost weak learners into strong predictive models.

Polynomial regression is a form of linear regression that models the relationship between independent variables x and a dependent variable y using an nth degree polynomial.

The evaluation metrics for the classification task include accuracy, recall, F1_Score, Roc_score, and the evaluation metrics for the regression task include Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination R^2. The indicators chosen are the most popular ones in the field to which they belong, so we will not go into details.

E. HYPERPARAMETER TUNING

Hyperparameter optimization is crucial for enhancing the performance of machine learning models. Various methods such as grid search, random search, evolutionary algorithms, and Bayesian optimization are widely used, each with its own merits and limitations. Grid Search: Exhaustive and computationally expensive, it is less suitable when dealing with a large parameter search space and numerous hyperparameters. Random Search and Bayesian Optimization: These methods offer computational efficiency at the potential expense of some accuracy, as they don't evaluate all possible hyperparameter combinations. Evolutionary Algorithms: Though effective, these are complex to implement and computationally demanding. Considering the trade-offs between computational cost, accuracy, and complexity, we opted for Bayesian optimization in this study. This method strikes a balance between computational efficiency and model accuracy, making it well-suited for our needs. The optimal hyperparameter configuration obtained through Bayesian optimization is detailed in Table 5.

TABLE 5. Optimal hyper-parameterized configuration.

hyperparameterization	search spaces	optimum value
Learning rate	(0.01, 0.25)	0.25
Number of neurons in first hidden layers	(6, 18)	9
Number of neurons in second hidden layers	(6, 18)	8
Number of Iterations	(200, 500)	300

V. EXPERIMENTAL RESULTS AND ANALYSIS

A. SERVICE AREA DRIVE-IN RECOGNITION RESULTS

In order to validate our constructed service area drive-in vehicle recognition model GBDT, we evaluated a variety of the most popular classification algorithms using the service area dataset. The selected methods include both the integrated methods XGBoost, LightGBM, AdaBoost, as well as the clustering algorithm KNN and the probabilistic learning algorithm Naive Bayes. These models were chosen based on their widespread use, diversity of approaches, and ability to handle tabular data. This extensive comparison provides a comprehensive benchmark, and the performance of each classifier is shown in Table 6, where the best results under different metrics are shown in bold.

TABLE 6. Comparison of recognition models for vehicles driving into service areas.

Model	Accuracy	Recall	F1_score	Roc-Auc
LightGBM	0.978521	0.938	0.938939	0.96259
XGBoost	0.977465	**0.942**	0.936382	0.963521
Adaboost	0.973944	0.918	0.925403	0.951949
SVM	0.972183	0.902	0.91947	0.944589
KNN	0.969718	0.896	0.912424	0.940735
Naive_bayes	0.965141	0.848	0.895459	0.919085
GBDT	**0.980634**	**0.942**	**0.944835**	**0.965444**

The experimental results indicate that our constructed GBDT model achieves excellent performance across all evaluation metrics compared to the other methods. Specifically, it attained over 98% accuracy, 94% recall, 94% F1-score and 96% Roc-Auc score. The GBDT model and other decision tree-based algorithms generally surpass the other approaches. This aligns with existing research showing decision trees are well-suited for tabular data classification due to their hierarchical structure and ability to capture interactions. The poorer performance of algorithms like SVM, KNN, and Naive Bayes suggests they may have more difficulty learning complex feature relationships in this scenario.

B. SERVICE AREA DWELL TIME PREDICTION RESULTS

To obtain the dwell time of vehicles entering the service area, we evaluated and compared several state-of-the-art machine learning and deep learning models, including TabNet, ResNet, XGBoost, LightGBM, Random Forest and Polynomial Regression. These models were selected based on their proven performance for regression tasks, ability to

handle tabular data, and diversity of algorithms. We utilized the most popular regression performance metrics - RMSE, MAE, and Coefficient of Determination R^2. RMSE indicates model prediction accuracy, MAE measures error magnitude, and R^2 represents the degree of fit.

TABLE 7. Comparison of prediction models for vehicle dwell time in service areas.

Model	RMSE	MAE	R^2
TabNet	12.2395	137.6143	0.9297
ResNet	**8.1103**	44.9034	0.9986
Polynomial	60.60101488	32.68036024	0.9942044
Random Forest	57.30136392	34.52576929	0.994818344
XGBoost	57.34238705	35.17253746	0.994810922
LightGBM	202.7045684	69.84250201	0.935156612
Bayes-BPNN	32.72423485	**25.39864210**	**0.999338639**

The experimental results in Table 7 demonstrate the strengths of deep learning, especially ResNet, for modeling the complex relationships between features that determine dwell time patterns. However, the proposed Bayes-BPNN model attained the lowest MAE of 25.3986421 and highest R^2 of 0.999338639, indicating it most accurately captured the nuances of private vehicle dwell time characteristics. The traditional machine learning models like Random Forest, Xgboost, and Polynomial regression also performed reasonably well. However, they are more difficult to map nonlinear relationships between features in the context of applications to large-scale datasets.

C. SINGLE-VEHICLE TRAFFIC STATUS DETECTION

The new method of travel speed calculation proposed in this paper takes into account the dwell time of vehicles in the service area, with the aim of restoring the normal driving behavior of vehicles entering the service area for accurate road traffic state monitoring. We categorized by vehicle type and analyzed the travel time and speed of eight common vehicle types - four passenger car types (0, 1, 2, 4) and four van types (11, 12, 13, 14). The numerical results represent the average across all vehicles of the same type due to the large sample sizes.

We conducted experiments from two perspectives, travel time and travel speed, respectively, and took significance analysis and visualization of the experimental results. Figure 8 shows the visualization of the experimental results regarding travel time, with $p\_value = 0.004$ at a set significance level of $\alpha = 0.05$. This indicates that under the assumption that there is no significant difference between the pre-correction and post-correction travel times, the probability of obtaining such a significant difference is only 0.004. In other words, we can reject this hypothesis, i.e., we can consider that the post-correction travel times are significantly different from the pre-correction travel times.

Figure 9 shows the visualization of the experimental results of the pre-corrected and post-corrected travel speeds. To further objectively illustrate the scientific validity of the

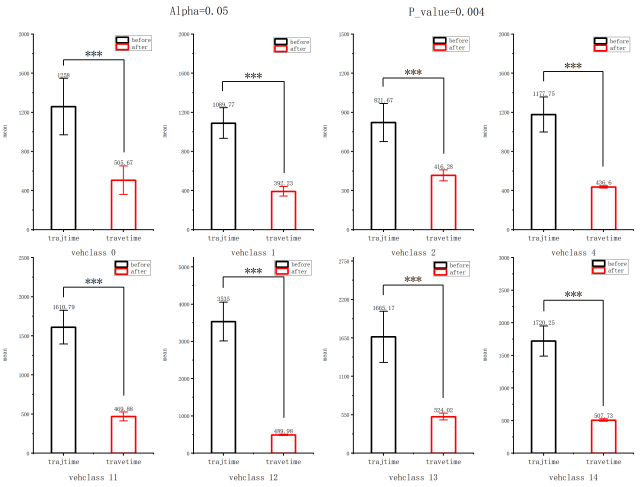


FIGURE 8. Comparison of corrected travel times for the same type of vehicle.

corrected travel speeds, we use the travel speeds of the same type of vehicles that did not enter the service area as a standard reference. From the experimental results, it can be seen that the corrected travel speed is closer to the travel speed without driving into the service area, for the same type of vehicle. Secondly, the corrected travel speed is more in line with the realistic scenario of the highway. Specifically, the travel speed of vehclass1 before correction is about 40km/h which is a low speed driving behavior, and the travel speed after correction is about 86km/h which is a normal driving behavior. Finally, although the post-correction travel speeds are closer to the travel speeds of the vehicles not entering the service area, the post-correction travel speeds are all lower than the travel speeds of the vehicles not entering the service area. The reason for this may be that the vehicle entering the service area decelerates early before entering the service area and accelerates slowly after leaving the service area, and this process reduces the vehicle's travel speed.

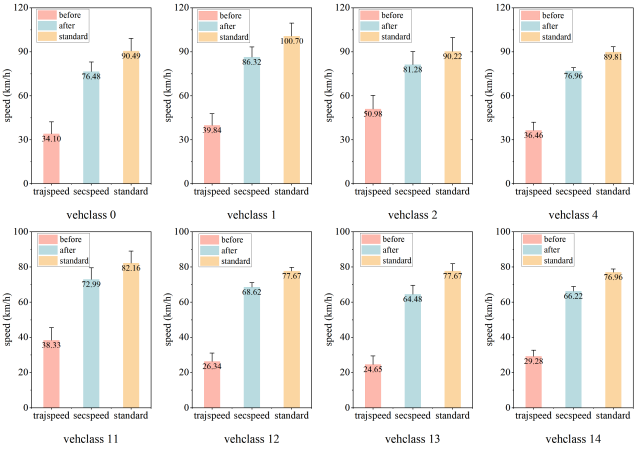


FIGURE 9. Comparison of corrected travel speeds for the same type of vehicle.

D. SHORT-TIME ROAD TRAFFIC STATUS MONITORING

Accurate measurement of travel speeds is crucial for effective road traffic monitoring. To gain a more nuanced understanding of road traffic conditions, we conducted further analyses on both travel times and speeds, utilizing a one-hour observation window. Since road traffic status is influenced by all vehicles in transit, we calculated the average travel time and speed across all vehicles. These averages serve as representative metrics for the road's overall transit time and speed. The distribution of travel times is depicted in Figure 10.

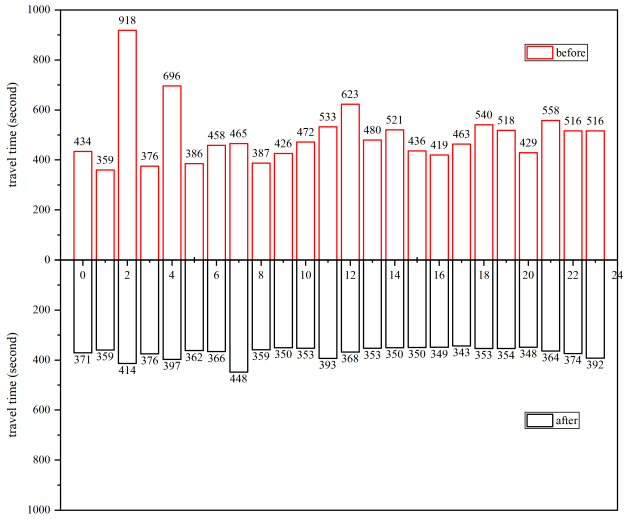


FIGURE 10. Comparison of hourly corrected passing times for the road segment.

The pre-correction pass times for the given road segment predominantly exceed 400 seconds, averaging around 497 seconds, and exhibit significant fluctuations. Post-correction times, however, gravitate around 350 seconds and average 368 seconds, demonstrating a smoother pattern. This lends credence to the notion that the corrected passing times are more reflective of authentic highway conditions. For instance, at 2:00 a.m., the pre-correction passing time was 918 seconds, which dropped dramatically to 414 seconds post-correction–a reduction of 504 seconds. Excluding the time slots of 1:00 a.m. and 3:00 a.m., where no vehicles entered the service area and thus no corrections were made, the passing times were universally reduced.

The corrected passing speeds are visualized in Figure 11. The vertical axis represents time, and the horizontal axis displays the range of passing speeds. A connecting line between two data points serves as a visual cue for the magnitude of change–the longer the line, the greater the difference. Apart from the aforementioned time slots of 1:00 a.m. and 3:00 a.m., there is an observed increase in passing speed across all time frames. The unadjusted speeds primarily ranged between 80km/h and 95km/h, with a mean value of 87.7km/h. In contrast, corrected speeds mostly fell between 90km/h and 100km/h, averaging 94km/h. Given that the road's speed limits are 110km/h for cars and 100km/h for other vehicles, these corrected values align more closely with actual traffic patterns on the highway.

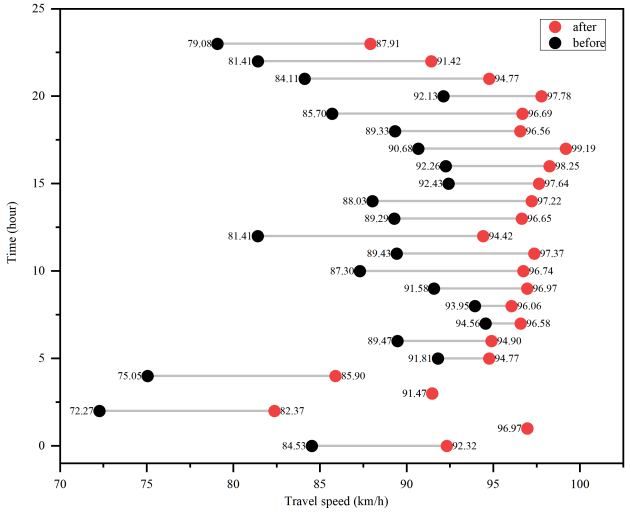


FIGURE 11. Short-time road traffic status - travel speed comparison.

E. LONG-TERM ROAD TRAFFIC STATUS MONITORING

The long-term road traffic status is equally important as it helps the traffic management authorities to formulate relevant improvement measures to enhance the travel service level of the road segment. We analyzed the traffic status of the road segment for one day on a one-day basis. Again, we have considered the passing time and passing speed. The passing status of all vehicles not entering the service area on that day was used as a standard reference to compare the passing status before and after correction. Analyze and compare the significant difference before and after correction, the symbols are indicated as follows, ns: no significant difference, *: significant difference, the more * means the more significant difference.

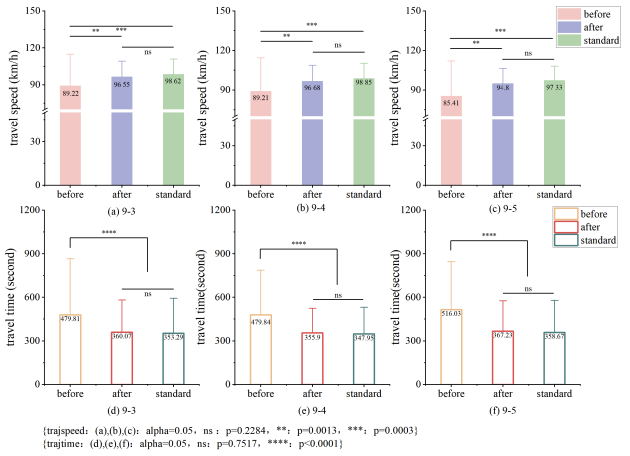


FIGURE 12. Long-time road traffic status - comparison after correction.

Figure 12 (a) (b) (c) shows the passing speeds on September 3, September 4, and September 5, respectively, and it can be seen that there is a very significant difference

between the passing speeds before correction and the passing speeds after correction, $p_{value} = 0.0013$; there is no significant difference between the corrected passing speeds and the passing speeds of the vehicles that did not drive into the service area, $p_{value} = 0.2284$; the pre-correction passing speed is very significantly different from the passing speed of vehicles not driving into the service area, $p_{value} = 0.0003$.

Similarly, Figure 12 (d) (e) (f) shows the passing times for the three days from September 3rd to September 5th. There is a highly significant difference between the pre-reconstruction travel time and the post-reconstruction passing time for vehicles that did not drive into the service area, $p_{value} < 0.0001$; the post-reconstruction passing time is not significantly different from the standard passing time, $p_{value} = 0.7517$. The experimental results strongly indicate that the roadway passing states obtained from the corrected travel times and travel speeds are more accurate, which may be an interesting finding.

VI. CONCLUSION

This paper introduces a novel framework to enhance travel speed modeling by accounting for the significant influence of service area dwell times, an overlooked factor in previous research. By integrating machine learning prediction of dwell times into travel speed calculations, our approach provides more accurate measurements that reflect normal driving behaviors, as validated experimentally. The recalibrated speeds enable precise short-term and long-term assessment of traffic conditions across multiple time horizons.

The study offers three key contributions - highlighting the importance of dwell time for travel speed analysis, demonstrating the ability to recover authentic speeds after adjusting for dwell time, and presenting an innovative solution to address limitations in existing methods. The proposed framework significantly advances the capability to model speeds on roadways with service areas.

Furthermore, this research establishes a crucial connection between service area dwell times, travel speed metrics, and comprehensive traffic condition evaluations. As dwell times escalate, the resulting diminished speeds are primarily due to extended stops, not genuine congestion. Overlooking dwell times can inadvertently lead to an underestimation of the roadway's true capacity. By recognizing and accounting for dwell times, our methodology facilitates a more discerning assessment of traffic states and congestion severity based on speed indicators. This, in turn, paves the way for a more robust appraisal of highway service quality and the formulation of impactful traffic management policies.

While this study focuses on a specific road segment, future research should validate the framework more extensively across diverse locations and conditions. Overall, by considering service area dwell times, this pioneering study opens a promising path to enrich travel speed prediction practices and provides a springboard for further enhancements. The framework offers a viable direction to help formulate effective traffic management strategies that account for the nuances of driver behaviors.

REFERENCES

[1] S. Tak, J.-D. Lee, J. Song, and S. Kim, "Development of AI-based vehicle detection and tracking system for C-ITS application," *J. Adv. Transp.*, vol. 2021, Aug. 2021, Art. no. e4438861, doi: 10.1155/2021/4438861.

[2] X. G. Jiang et al., "Systematic review on anti-speeding countermeasure research," *China J. Highway Transp.*, vol. 33, no. 3, pp. 1–31, 2020, doi: 10.19721/j.cnki.1001-7372.2020.03.001.

[3] T. Chen, N. N. Sze, S. Saxena, A. R. Pinjari, C. R. Bhat, and L. Bai, "Evaluation of penalty and enforcement strategies to combat speeding offences among professional drivers: A Hong Kong stated preference experiment," *Accident Anal. Prevention*, vol. 135, Feb. 2020, Art. no. 105366, doi: 10.1016/j.aap.2019.105366.

[4] X. Xiong, Q. Liu, Y. Shen, Y. Cai, and L. Chen, "Study on risk model of highway traffic accidents based on LSTM BF," *China Saf. Sci. J.*, vol. 32, no. 5, pp. 170–176, 2022, doi: 10.16265/j.cnki.issn1003-3033.2022.05.1602.

[5] Y.-Y. Choi, D.-K. Kim, S.-Y. Kho, and P. Y. Park, "Safety evaluation of four safety countermeasures on freeways in South Korea," *J. Transp. Saf. Secur.*, vol. 12, no. 7, pp. 945–958, Aug. 2020, doi: 10.1080/19439962.2018.1564945.

[6] Q. R. Li, E. Q. Hao, L. Chen, Z. G. Fan, and W. W. Yang, "Urban traffic state recognition based on genetic algorithm optimized support vector machine," *J. Chongqing Jiaotong Univ., Natural Sci.*, vol. 39, no. 1, pp. 1–5 and 13, 2020.

[7] Y. C. Zeng et al., "Traffic prediction and congestion control based on directed graph convolution neural network," *China J. Highway Transp.*, vol. 34, no. 12, pp. 239–248, 2021, doi: 10.19721/j.cnki.1001-7372.2021.12.018.

[8] H. Li and S. Xiong, "Time-varying weight coefficients determination based on fuzzy soft set in combined prediction model for travel time," *Expert Syst. Appl.*, vol. 189, Mar. 2022, Art. no. 115998, doi: 10.1016/j.eswa.2021.115998.

[9] P. Yong, Z. Xin, S. Qiankun, and X. Zhonghua, "A combined predicting model for expressway travel time based on EMD-GRU," *Appl. Math. Mech.*, vol. 42, no. 4, pp. 405–412, 2021, doi: 10.21656/1000-0887.410165.

[10] S. Luo, F. Zou, C. Zhang, J. Tian, F. Guo, and L. Liao, "Multi-view travel time prediction based on electronic toll collection data," *Entropy*, vol. 24, no. 8, p. 1050, Jul. 2022, doi: 10.3390/e24081050.

[11] S. L. Luo et al., "Active traffic guidance method for recurrent congestion points," *J. Transp. Inf. Saf.*, vol. 39, no. 5, pp. 68–75, 2021.

[12] B. Tian, X. M. Zhao, Z. G. Xu, W. Miao, and Z. Yu-Qin, "NRT-V2X: Adaptive data dissemination protocol for traffic efficiency of connected and automated highways," *China J. Highway Transp.*, vol. 32, no. 6, pp. 293–307, 2019, doi: 10.19721/j.cnki.1001-7372.2019.06.029.

[13] L.-L. Chen, A. Zhang, and X.-G. Lou, "Cross-subject driver status detection from physiological signals based on hybrid feature selection and transfer learning," *Expert Syst. Appl.*, vol. 137, pp. 266–280, Dec. 2019, doi: 10.1016/j.eswa.2019.02.005.

[14] Y. Li, M. Abdel-Aty, J. Yuan, Z. Cheng, and J. Lu, "Analyzing traffic violation behavior at urban intersections: A spatio-temporal kernel density estimation approach using automated enforcement system data," *Accident Anal. Prevention*, vol. 141, Jun. 2020, Art. no. 105509, doi: 10.1016/j.aap.2020.105509.

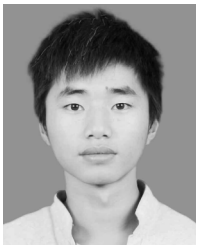
[15] Z. Zhou, Z. Yang, Y. Zhang, Y. Huang, H. Chen, and Z. Yu, "A comprehensive study of speed prediction in transportation system: From vehicle to traffic," *iScience*, vol. 25, no. 3, Mar. 2022, Art. no. 103909, doi: 10.1016/j.isci.2022.103909.

[16] F. Zou, Q. Ren, J. Tian, F. Guo, S. Huang, L. Liao, and J. Wu, "Expressway speed prediction based on electronic toll collection data," *Electronics*, vol. 11, no. 10, p. 1613, May 2022, doi: 10.3390/electronics11101613.

[17] X. Zhang, L. Lauber, H. Liu, J. Shi, M. Xie, and Y. Pan, "Travel time prediction of urban public transportation based on detection of single routes," *PLoS ONE*, vol. 17, no. 1, Jan. 2022, Art. no. e0262535, doi: 10.1371/journal.pone.0262535.

[18] M. Abdollahi, T. Khaleghi, and K. Yang, "An integrated feature learning approach using deep learning for travel time prediction," *Expert Syst. Appl.*, vol. 139, Jan. 2020, Art. no. 112864, doi: 10.1016/j.eswa.2019.112864.

[19] China ITS Journal. (Aug. 17, 2022). *Highway Tolling Big Data Wisdom Integrated Application Platform Construction Program*. Accessed: Sep. 23, 2022. [Online]. Available: https://mp.weixin.qq.com/s/rgA5bitqJfBqDacL48KJlw
[20] Q. Wei, "ETC data based identification of vehicles entering the service area and traffic flow prediction," Chongqing Univ., Chongqing, China, Tech. Rep., 2021, doi: 10.27670/d.cnki.gcqdu.2021.000334.
[21] J. C. Weng et al., "A model of travel speed calculation for freeway road segments with ETC transaction data," *J. Transp. Inf. Saf.*, vol. 35, no. 5, pp. 76–82, 2017.
[22] J. Xu et al., "Research on speed distribution and vehicle type classification of mountain expressway based on electronic toll collection data," *J. Transp. Syst. Eng. Inf. Technol.*, vol. 22, no. 5, pp. 75–84, 2022, doi: 10.16097/j.cnki.1009-6744.2022.05.008.
[23] Y. N. Malek, M. Najib, M. Bakhouya, and M. Essaaidi, "Multivariate deep learning approach for electric vehicle speed forecasting," *Big Data Mining Anal.*, vol. 4, no. 1, pp. 56–64, Mar. 2021, doi: 10.26599/BDMA.2020.9020027.
[24] P. Wang, Y. Zhang, S. Wang, L. Li, and X. Li, "Forecasting travel speed in the rainfall days to develop suitable variable speed limits control strategy for less driving risk," *J. Adv. Transp.*, vol. 2021, May 2021, Art. no. e6639559, doi: 10.1155/2021/6639559.
[25] M. Zahid, Y. Chen, A. Jamal, and C. Z. Mamadou, "Freeway short-term travel speed prediction based on data collection time-horizons: A fast forest quantile regression approach," *Sustainability*, vol. 12, no. 2, p. 646, Jan. 2020, doi: 10.3390/su12020646.
[26] H. J. Cui and K. J. Liu, "A new determine method on pause rate in expressway service area based on vehicle continuous travel time," *J. Hebei Univ. Technol.*, vol. 37, no. 6, pp. 100–104, 2008.
[27] J. W. Wang and Y. Tang, "Transportation potential calculation model of pause rate in expressway service area," *China J. Highway Transp.*, vol. 5, no. 93, pp. 109–114, 2008, doi: 10.19721/j.cnki.1001-7372.2008.05.020.
[28] J. Liu, X. Y. Shen, and H. X. Liu, "Prediction model for percentage of expressway traffic entering rest area based on BP neural network," *Highway*, vol. 6, no. 6, pp. 164–168, 2012.
[29] Y. F. Liu, J. Chen, and S. T. Liu, "Study on the entering rate research and forecasting of Shanxi province expressway service area," *Highway Eng.*, vol. 38, no. 4, pp. 210–213, 2013.
[30] Z. L. Ou, "Research on the optimal scale of expressway service area based on A-CBR," *J. China Foreign Highway*, vol. 36, no. 6, pp. 304–308, 2016, doi: 10.14048/j.issn.1671-2579.2016.06.070.
[31] Y. Hou, C. Wen, P. Huang, L. Fu, and C. Jiang, "Delay recovery model for high-speed trains with compressed train dwell time and running time," *Railway Eng. Sci.*, vol. 28, no. 4, pp. 424–434, Dec. 2020, doi: 10.1007/s40534-020-00225-8.
[32] M. Volovski, E. S. Ieronymaki, C. Cao, and J. P. O'Loughlin, "Subway station dwell time prediction and user-induced delay," *Transportmetrica A, Transp. Sci.*, vol. 17, no. 4, pp. 521–539, Dec. 2021, doi: 10.1080/23249935.2020.1798555.
[33] H. Huo, C. J. Zheng, and J. X. Shen, "Prediction of bus dwell time using support vector machine based on near neighbors," *J. Transp. Eng. Inf.*, vol. 19, no. 3, pp. 59–66, 2021, doi: 10.19961/j.cnki.1672-4747.2020.11.001.
[34] I. K. Isukapati, C. Igoe, E. Bronstein, V. Parimi, and S. F. Smith, "Hierarchical Bayesian framework for bus dwell time prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 3068–3077, May 2021, doi: 10.1109/TITS.2020.2979390.
[35] H. Gao, "Study on driving mileage prediction of pure electric vehicles based on machine learning," Beijing Jiaotong Univ., Beijing, China, 2018.
[36] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," *Ann. Statist.*, vol. 29, no. 5, Oct. 2001, doi: 10.1214/aos/1013203451.
[37] F. H. Ma and Z. Q. Rao, "Comparison and analysis of results from professional experts of different specialty by AHP," *Highway*, vol. 62, no. 6, pp. 192–196, 2017.
[38] W. P. Li et al., "BPNN-HDMR nonlinear metamodeling technique and its application," *J. Hunan Univ., Natural Sci.*, vol. 41, no. 5, pp. 32–38, 2014.
[39] S. Ö. Arik and T. Pfister, "TabNet: Attentive interpretable tabular learning [J/OL]," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 8, pp. 6679–6687, doi: 10.1609/AAAI.v35i8.16826.
[40] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 3146–3154.
[41] T. Chen and C. Guestrin, "XGBoost: Extreme gradient boosting," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 785–794.
[42] S. Targ, D. Almeida, and K. Lyman, "Resnet in resnet: Generalizing residual architectures," 2016, *arXiv:1603.08029*.



XU LUO received the B.E. degree from the Chongqing Institute of Humanities and Technology, in 2021. He is currently pursuing the master's degree in transportation engineering with the Fujian University of Technology. His research interests include machine learning, deep learning, data mining, and vehicle-road collaboration.

FUMIN ZOU received the Ph.D. degree from Central South University, Changsha, China, in 2009.

He is currently a Professor and the Dean of the College of Electrical and Electronics and Physics, Fujian University of Technology, an Adjunct Professor, and a Doctoral Supervisor of the College of Computer and Big Data, Fuzhou University. He has presided over and participated in more than 50 provincial and ministerial research projects, published more than 100 academic papers, including 50 SCI/EI papers, applied for more than 100 national invention patents, of which more than 50 national invention patents and three U.S. invention patents have been authorized. He has presided over the completion of the National Natural Fund, Provincial Outstanding Youth Fund, Provincial Science and Technology Major Project, and other scientific research projects, and he has won more than ten awards, such as Fujian May Fourth Youth Medal, the China Industry-University-Research Cooperative Innovation Achievement Award, the Fujian Provincial Natural Science Award, the Fujian Provincial Science and Technology Progress Award, the Fujian Provincial Teaching Achievement Award, and other awards. His main research interests include intelligent driving, vehicle-circuit coordination, big data technology, artificial intelligence technology, ETC big data mining, traffic information, control engineering, and its applications.

SIJIE LUO received the M.S. degree in transportation engineering from the Fujian University of Technology, Fujian, China, in 2021. He is currently pursuing the Ph.D. degree in information and communication engineering with the University of Electronic Science and Technology, Chengdu, China, where he has published more than ten SCI/EI/CSCD-searchable scientific and technical papers or conference papers. His research interests include trajectory data mining and machine vision.

FENG GUO received the M.S. degree in computer science from Tsinghua University, in 2013, and the Ph.D. degree in computer science from Fuzhou University, in 2023. He is currently teaching with the Fujian University of Technology. His research interests include machine learning, deep learning, vehicle-circuit coordination, and intelligent driving.