Urban Arterial Traffic Volume and Travel Time Estimation with Use of Data Driven Models

Chris Konstantinidis*, Panagiotis Fafoutellis* and Eleni I. Vlahogianni, Ph.D*.

National Technical University of Athens, 5 Iroon Polytechniou Str., Zografou Campus, GR 157 73, Athens, Greece

Abstract

This work aims to develop traffic prediction models, with a specific focus on traffic volume and travel time of an urban arterial. These models are based on Machine Learning algorithms, which find frequent application in the literature for various forecasting tasks. The traffic data that were utilized in the development process were sourced from a road section of Alexandras Av. in Athens. Finally, a comparison is drawn between Machine Learning models, the BPR Volume-Delay function, and a BPR-ML hybrid approach. According to the results, the models are capable of accurate predictions with an acceptable fit to the data. The findings reveal that Machine Learning methods exhibit superior forecasting capabilities.

Keywords: machine learning, traffic forecasting, traffic volume, travel time, decision trees, GBDT, BPR function

1. Introduction

In the modern highly congested urban environments, accommodating the growing traffic demand has proven to be a complex task on a short-term scale, as permanent interventions in transportation systems are time-consuming and involve significant costs. Therefore, optimal management and utilization of existing transportation networks are deemed necessary to address the challenge of rising demand.

Research has shown that predictive traffic models that can link the fundamental macroscopic traffic variables with travel times stand as pivotal tools for the optimal management of transportation systems, enhancing efficiency and leading to safer, more viable network level traffic conditions (Vlahogianni et al. 2014). Volume-Delay functions, which are mathematical expressions correlating travel time and traffic volume within a road link, stand out as a typical example. In particular, the Bureau of Public Roads (BPR) function, has been consistently utilized as a tool for traffic modeling due to its straightforward mathematical formula and computational efficiency. In a complex spatiotemporal traffic environment, it provides a simple theoretical framework for the relationship of traffic volume and travel time without involving many variables and necessitating excessive amounts of input data (Mtoi & Moses, 2014). Additionally, one notable advantage of the BPR function lies in its ability to be adjusted to locally observed traffic volume-travel time data of specific road links, by tuning its parameters appropriately (Mehbub Anwar et al., 2011).

The question is raised whether contemporary methods, such as Machine Learning prediction models, may provide more accurate predictions compared to conventional approaches. The absence of a calibration framework for the BPR function is apparent in the pertinent literature. Originally designed for segments of highways in developed nations, the function encounters challenges when applied to diverse road infrastructures with varying characteristics such as illegal parking, road geometry, vehicle size and percentage of heavy transport (Mehbub Anwar et al., 2011; Pan et al., 2022; Saberi & Figliozzi, 2011). As mentioned above, tailoring the parameters to the specific characteristics of a road segment has the potential to optimize the functions performance. However, this approach is constrained by the requirement for re-calibration for each unique road segment case (Pan et al., 2022).

The diverse characteristics of road infrastructures mentioned have a substantial influence on the intricate process of estimating the traffic variables necessary for applying the function (Mehbub Anwar et al., 2011). Accurately determining the traffic capacity (C) of a road link is a complex undertaking, particularly when faced with a lack of appropriate data (Petrik et al., 2014; Saberi & Figliozzi, 2011).

In the realm of transportation, a substantial portion of scientific prediction is dedicated to discovering the optimal prediction methodology through experimentation and comparative analysis. Lately, research has diverged from classical mathematical / analytical expressions to Machine Learning prediction models that appear less constrained by assumptions about data behavior and are flexible and resilient to noise and missing data.

Specifically, Shallow Machine Learning models pose as viable candidates for tackling the traffic forecasting problem (Manibardo et al., 2022). Drawing again on the literature, it is also worth noting that traffic state data that are closely related to the predictor variable, are frequently utilized as input feature for these models. Common examples include the use of travel time data of previous time intervals when predicting travel time (Qiu & Fan, 2021; Zhang & Haghani, 2015, Cheng et al., 2019), or traffic volume of adjacent road links when forecasting traffic volume (Yang et al., 2017; Ying et al., 2017). Additionally, a lack of traffic prediction at network level is a common denominator among the literature (Manibardo et al.; 2022, Zou et al., 2020), further emphasizing the challenges of data collection pertaining to certain traffic parameters, such as traffic volume. The framework proposed in this work aims to address those issues by formulating Shallow Machine Learning models that rely solely on easily accessible data. Also, a deliberate decision was made to explore the predictive capabilities of the models without incorporating data of temporal and spatial correlation like the ones mentioned above.

The primary objective of this study is to develop Machine Learning models that can predict traffic volume in an urban road segment, by utilizing travel time data as the only traffic parameter related feature. This approach is justified by the ease of access and abundance of travel time data, which can be collected via GPS devices, mobile apps, or other location-tracking services at a minimal cost. On the other hand, the collection of traffic volume data is more complex. It requires the installation of infrastructure such as traffic cameras and loop detectors, which involves logistical challenges and higher costs (Bae et al., 2017). Moreover, these data are usually managed by public authorities, making access more complicated. Therefore, the ability to infer prediction results between travel time and traffic volume could provide valuable insights for transportation system management. For the prediction tasks, Decision Trees (DT) and Gradient Boosting Decision Trees (GBDT) were used, two tree-based supervised Machine Learning algorithms. The models were trained on actual travel time data from Google Maps and traffic volume data from loop detectors installed on Alexandras Avenue in Athens.
Aside from its apparent utility, the task of travel time forecasting also assists in accomplishing the supplementary goal of this study, which is the comparison of contemporary Machine Learning methods with conventional approaches of correlating traffic variables, such as Volume-Delay Functions. This comparative analysis between historical and contemporary methodologies elucidates whether and to what extent predictive capabilities have advanced in the digital era. The Bureau of Public Roads (BPR) function is one of the most widely used Volume-Delay Functions and was employed as a tool for travel time prediction in this study. The modelling capabilities of the BPR function are restricted solely to road links, necessitating the assignment of values to its parameters for its practical application. Hence, an algorithm was developed using the least squares method to calibrate the function by minimizing prediction errors. The assumptions inherent in the BPR function present an opportunity for comparison with the non-parametric, Machine Learning models. This allows for an assessment of whether the premises of VDFs hold for the analysis of traffic flow phenomena. Lastly, a hybrid BPR-ML approach was explored to evaluate potential benefits of the integration of the functions modelling capabilities to previously mentioned Machine Learning models. The same traffic data were used in all the mentioned approaches.

2. Methodological approach

Machine Learning models have also been successfully used in combination with several analytical equations from traffic flow theory, e.g., Microscopic traffic flow models (Lighthill – Whitham – Richards - LWR model) and car following models (Lu et al. 2023). In this paper, one of the first attempts to model the volume-delay function is proposed. Other attempts include (Madadi & Homem de Almeida Correia, 2024; Bae et al., 2017). In the first example, a combination of a Graph Neural Network and a Genetic Algorithm is employed to address the User-Equilibrium Traffic Assignment Problem in road networks. The BPR function serves as the readout function for calculating total travel times, similar to performing a graph regression task. In the latter case, a Multilayer Perceptron Neural Network is used to achieve long-term traffic volume and truck percentage estimates. Consequently, adjusted volume-to-capacity ratios are proposed, by incorporating the predictions, and alongside a transformation of the BPR function into various curve-fitting models, travel time estimates in hypothetical traffic conditions of construction zones are achieved.

In general, Machine Learning algorithms iteratively improve their performance as they are exposed to more data, aiming to generalize insights and adapt to new, unseen situations without explicit programming. The mentioned algorithms fall within the domain of supervised Machine Learning, which means they are trained on datasets containing labeled data points. In this case, the input data, accompanied by corresponding desired outputs or labels, guide the algorithms training process. One key characteristic of Shallow Machine Learning models is the requirement of manual feature selection or engineering. Features are input variables or attributes by which patterns are extracted from data in order to make predictions or classifications. Lastly, model performance is evaluated by quantifying the deviation between the predicted outcomes and the actual values in the dataset. In this section, a basic description of the algorithms is presented.

Decision Trees stand as a fundamental algorithm in Shallow Machine Learning, adept at both classification and regression tasks. Employing a hierarchical structure, they partition the data by choosing the feature value that minimizes the Mean Squared Error (MSE) across subsets. The root node, which represents the entirety of the dataset, is positioned at the top. The MSE for a node is calculated as follows:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y})^2 \qquad (1)$$

where N is the number of samples in the node, y_i is the target value for the i_{th} sample, and $\hat{y}$ is the mean target value for all samples in the node.

The algorithm selects the feature and the corresponding threshold that minimizes the MSE across the two child nodes after the split, and this is quantified by the Gain:

$$\text{Gain} = MSE_{parent} - \left(\frac{N_{left}}{N_{parent}}\right) * MSE_{left} - \left(\frac{N_{right}}{N_{parent}}\right) * MSE_{right} \qquad (2)$$

where $\mathrm{N_{parent}}$ is the number of samples in the parent node, $\mathrm{N_{left}}$ and $\mathrm{N_{right}}$ are the number of samples in the left and right child nodes, and $\mathrm{MSE_{parent}}$, $\mathrm{MSE_{left}}$, $\mathrm{MSE_{right}}$ are the mean squared errors for the respective nodes.

This process leads to the creation of child nodes, with branches indicating potential values for a feature. The procedure iterates until it meets the termination criterion specified by the hyperparameters. In regression scenarios, the algorithm concludes at terminal nodes, which serve as the prediction values, determined as the average of the target values within that leaf node:

$$\hat{y}_{node} = \frac{1}{N_{node}} \sum_{i=1}^{N_{node}} y_i \qquad (3)$$

where $\hat{y}_{node}$ is the predicted target value for the node, and N_{node} is the number of samples in the node.

To effectively train a decision tree, it is crucial to define specific hyperparameters, such as:

- Maximum Depth (max_depth): The maximum depth of the tree, determining the total number of nodes from the root to the leaf node.
- Minimum Samples for Splitting (min_samples_split): The split of nodes is based on this parameter. If a node contains fewer samples than the specified value, it becomes a leaf node.
- Minimum Samples per Leaf (min_samples_leaf): Nodes with fewer samples than this parameter are merged with others.
- Maximum Number of Features (max_features): The maximum number of features the algorithm considers at each node to achieve the optimal split.
- Random State (random_state): The configuration of a decision tree involves randomness, especially when equivalent splits exist during its creation. Setting this parameter to a constant value ensures the repeatability of the structure and performance of the model.

While Decision Trees boast advantages, such as high interpretability, ease of visualization, and effectiveness in handling complex, non-linear problems, they encounter limitations. Decision Trees can be prone to overfitting the data, especially in complex scenarios. The top-down approach used by the algorithm may not always lead to the optimal decision tree structure, potentially affecting its performance. Additionally, Decision Trees exhibit instability, forming different tree structures with small changes in the data, which can impact the reliability of the model. In problems displaying linearity, capturing relationships between variables might be challenging for Decision Trees.

To overcome some of the aforementioned limitations, a pairing of the Decision Trees algorithm with ensemble learning methods, such as boosting, proves beneficial. One such application of this synergy is seen in the Gradient Boosting Decision Trees (GBDT) algorithm. At its core, this algorithm revolves around Decision Trees. The utilization of the Gradient Boosting technique enhances predictive accuracy by progressively introducing Decision Trees into the model. The algorithm initiates with a single Decision Tree. The loss function, denoted as $L(y, F(x))$, measures the difference between the true target values y and the current models predictions $F(x)$. The typical metric for gauging this loss is typically Mean Squared Error (MSE) in the context of regression problems.

$$L(y, F(x)) = \frac{1}{2}\left(y - F(x)\right)^2 \quad (4)$$

In each iteration, a new decision tree is trained to predict the negative gradient of the loss function.

$$-\frac{dL}{dF(x)} = y - F(x) \quad (5)$$

$$h_t(x) = argmin_h \sum_{i=1}^{N} L(y_i, F_{t-1}(x_i) + h(x_i)) \quad (6)$$

where t is the iteration, N is the number of samples, $F_{t-1}(x_i)$ is the current ensemble's prediction for sample i, and $h_t(x)$ is the prediction of the new tree.

Subsequent iterations involve the sequential integration of new decision trees, each aiming to refine its adaptation to the data. The additional trees aim to capture patterns that were not identified by their predecessors and minimize the loss function. This iterative process continues until the predefined terminal criteria are met, as dictated by the hyperparameters. The ultimate outcome of a GBDT model materializes by aggregating the predictions from all individual decision trees.

The final prediction is obtained by aggregating the predictions of all weak learners.

$$F(x) = \sum_{t=1}^{T} a_t * h_t(x) \quad (7)$$

Where T is the total number of weak learners (trees) in the ensemble, a_t is a weight assigned to each learner called learning rate. The learning rate modulates the impact of each new tree on the final result ensuring a nuanced and calibrated influence.

Some common hyperparameters of GBDT are presented below:

- Number of estimators (n_estimators): The number of individual decision trees to be included in the ensemble.
- Learning rate: Adjusts the contribution of each decision tree to the final result.
- Maximum depth (max_depth): The maximum depth of the tree, defining the total number of nodes from the root node to the leaf node.

Gradient Boosting offers superior performance and accuracy when compared to other Shallow Machine Learning algorithms. Particularly effective in handling complex, non-linear problems, it excels in mitigating data noise by assigning weights to correctly classified data, enhancing overall robustness. Moreover, it provides tools for assessing the importance of features, contributing to a better understanding of the data. Despite its high-performance capabilities, GBDT remains relatively easy to understand, sharing interpretability akin to decision trees. However, it is not without its challenges. The algorithm tends to exhibit overfitting, especially when dealing with deep tree structures and numerous iterations. Achieving satisfactory performance requires proper tuning of hyperparameters, a crucial aspect that demands careful optimization. GBDT is characterized by high computational costs, particularly noticeable in cases of large datasets and deep Decision Trees. Despite offering valuable insights that contribute to their interpretability as models, GBDTs are inherently more complex compared to individual Decision Trees due to their ensemble nature.

An alternative approach for the prediction of travel time proposed by the modeling framework of this study is the BPR function. Travel time stands as a fundamental consideration when determining a commute within a road system, intricately tied to the traffic flow of a road segment. To describe this phenomenon, Volume-Delay Functions (VDFs) have been developed. One widely known function is the Bureau of Public Roads Function (BPR), developed by the Bureau of Public Road in the 1960s as part of the Highway Capacity Manual. This mathematical function describes the congestion level of road segments expressed as a time delay based on changes in traffic volume.

$$T = T_{ff}\left(1 + \alpha\left(\frac{V}{C}\right)^{\beta}\right) \quad (8)$$

In Equation (8), parameter α of the BPR function expresses the ratio of travel time under free-flow conditions to the corresponding time under maximum traffic capacity conditions. Meanwhile, parameter β expresses the rate of change from free-flow conditions to congested conditions. Higher values of the β parameter imply shorter time delays under low traffic volume (Mahdi et al., 2022; Pan et al., 2022). This results in the delayed recognition of traffic congestion. Typically, the parameters α and β are assigned values of 0.15 and 4, respectively. However, practical applications of the function, as indicated by Mahdi et al. (2022), involve a diverse set of parameter values

The variable T_{ff} represents the travel time of the road segment under free-flow conditions. Free-flow conditions are characterized by low traffic volume and speed determined solely by driver preferences, defined speed limits, and prevailing road conditions. The term of traffic capacity (C) has been presented in a previous section.

Fig. 3 illustrates the relationship between travel time and traffic volume as described by the BPR mathematical function.

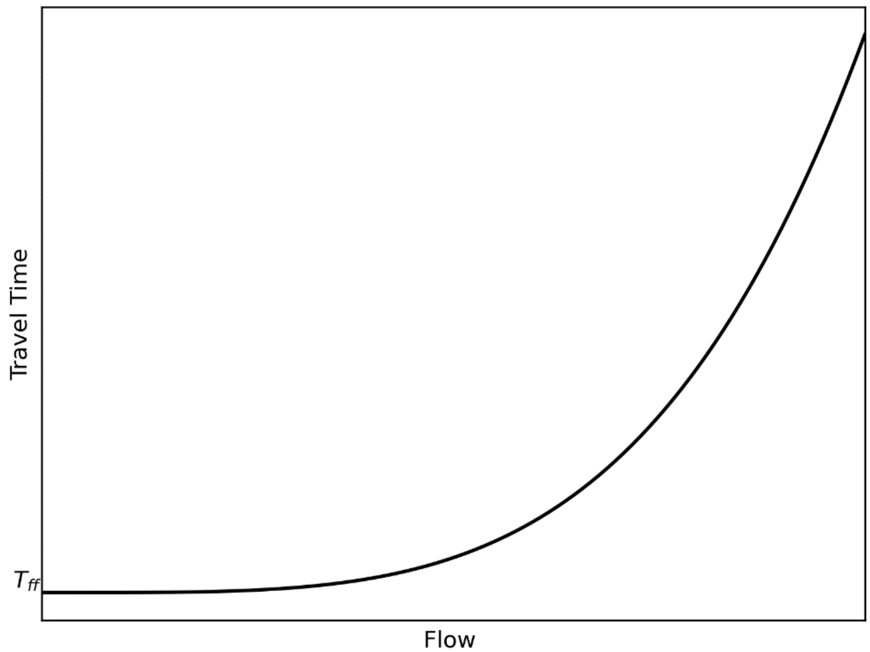

Fig. 3. BPR travel time-flow diagram

As mentioned in earlier sections, BPR function requires a calibration process for parameters α and β. For this purpose, the least squares method is proposed, a widely used technique for fitting functions. Pythons SciPy library employs the Levenberg-Marquardt algorithm to adjust function parameters and minimize the sum of squared residuals. After defining the objective function *S*, initial values need to be assigned to the parameters. The algorithm iteratively calculates *S* for the entire dataset, adjusting parameter values until changes become insignificant. The objective function *S* is given by:

$$S(\alpha, \beta) = \sum_{i=1}^{n} \left(T_{obs,i} - T \right) \tag{9}$$

where n = number of data points, $T_{obs,i}$ = observed travel time values, T = travel time values calculated with BPR

The described process ensures a best-case scenario for the prediction outcome of the BPR function, particularly beneficial when uncertainty exists regarding specific function variables. For instance, in this study, only an approximate estimate of road traffic capacity (C) could be derived from the available data. Notably, fitting the function with various road traffic capacity values revealed that the prediction error remained consistent, unlike the values of parameters α and β. However, the primary focus lies in the functions predictive capability rather than achieving precise parameter estimates for the road link under consideration.

3. Implementation and results

3.1 Data collection and preprocessing

The models were developed using traffic volume and travel time data obtained from two selected locations on Alexandras Avenue in central Athens. These locations are equipped with loop detectors, spaced at an approximate distance of 2.3 km. The first sensor, labeled MS423, is situated at geographical coordinates: (37.98706952, 23.75875395), while the second sensor, MS407, at: (37.9914926, 23.73333264).

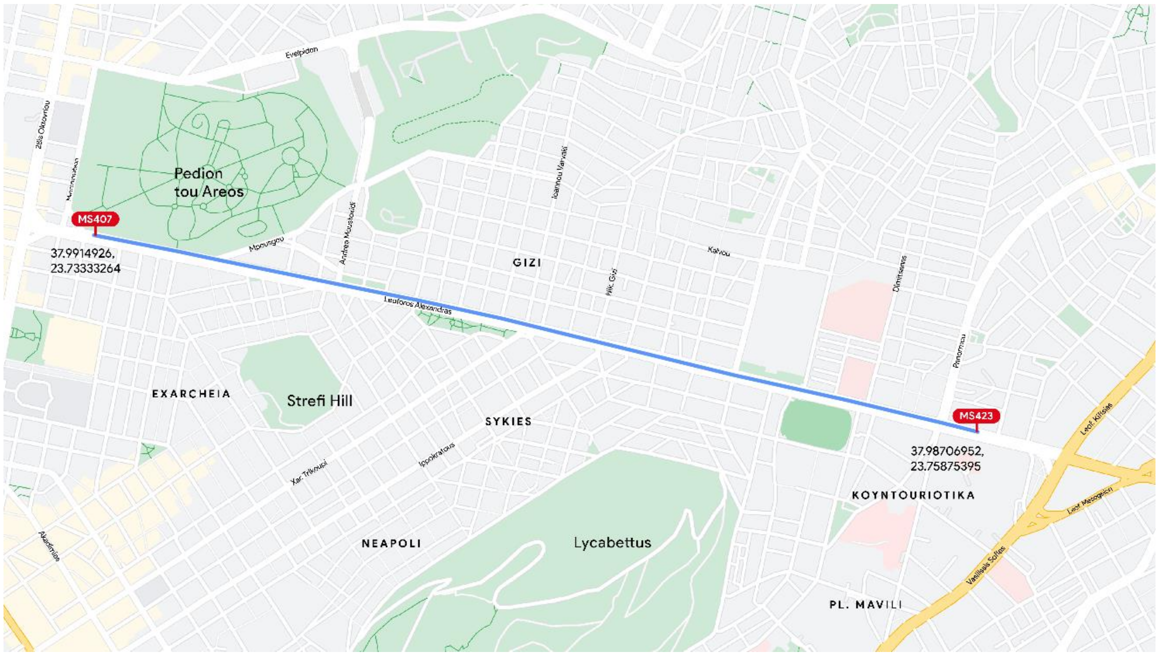

Fig. 4. Geographical representation of Alexandras Avenue road segment.

Hourly traffic volume data for 80 days were collected from the databases of the National Technical University of Athens (NTUA) for the specified loop detectors. For each month from December 2022 to March 2023, 20-day intervals were selected, during which the detectors recorded data without any malfunctions. Table 1 provides the precise dates of the utilized recordings. For the examined road segment of Alexandras Av. between the two detectors, travel time data (duration) were sourced from Google Maps databases.

Table 1
Dates of traffic data recordings.

December 1, 2022 to December 21, 2022
January 6, 2023 to January 26, 2023
February 1, 2023 to February 21, 2023
March 6, 2023 to March 26, 2023

Errors and missing values were eliminated from the aforementioned databases, pertaining to travel time and traffic volume. Rows across all databases containing absent or incorrect values were entirely disregarded. Subsequently, the two databases were merged into one, encompassing all necessary travel time and volume data, appropriately aligned by hour and date.
In order to gain a better understanding between the traffic volume and travel time data the correlation coefficient was calculated. A high correlation between two variables suggests the potential for one to serve as a predictor for the other. The Pearson correlation coefficient ranges from -1 to 1. Perfect positive linear relationship and perfect negative linear relationship are signified by values 1 and -1 respectively, and value 0 implies no linear relationship. Fig. 4 visualizes the correlation between travel time and traffic volume for each loop detector, while Fig. 5 offers a more detailed perspective on their values within the database. The diagrams revealed a good correlation between the variables, thus enabling the process to proceed to the stage of value prediction.

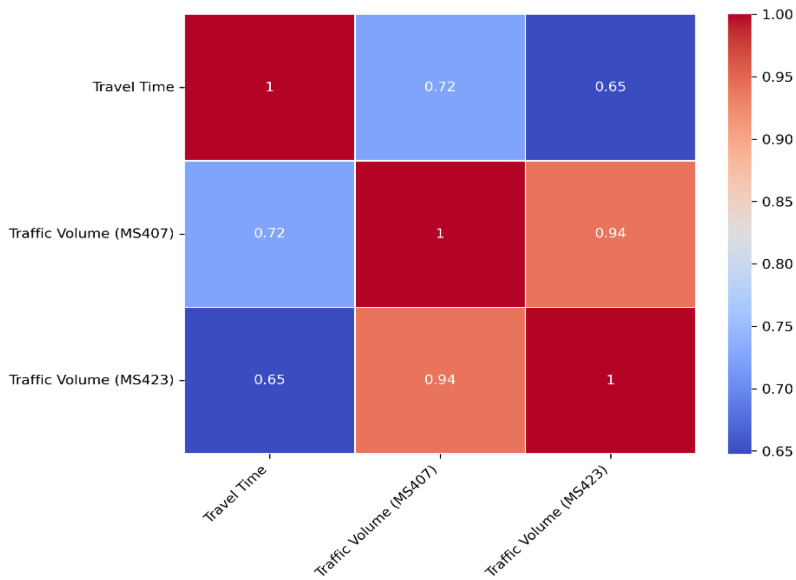

Fig. 5. Correlation heatmap of traffic volume and travel time data.

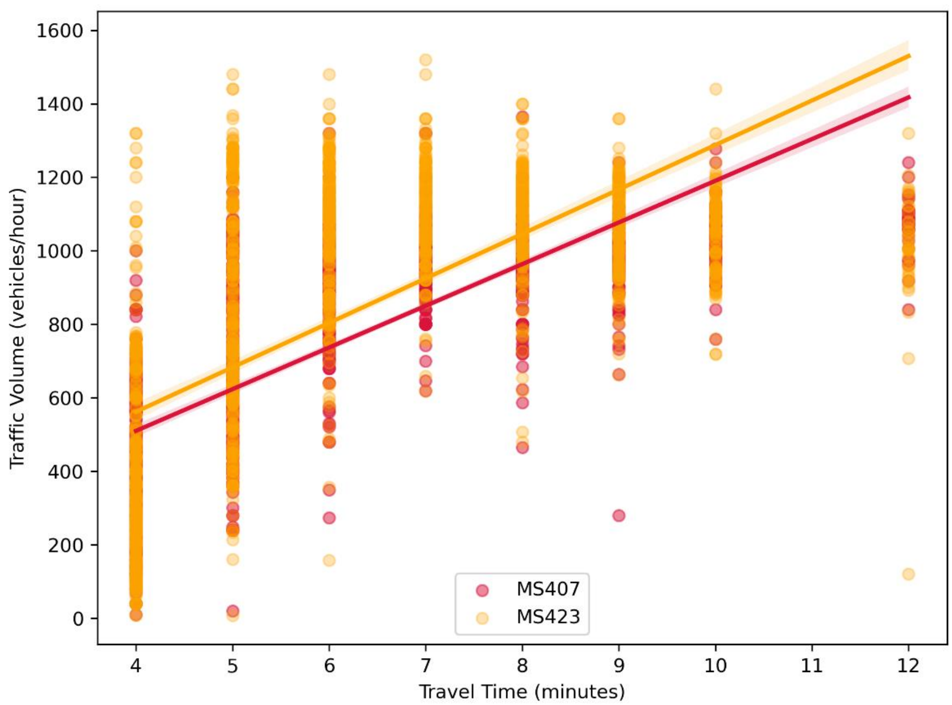

Fig. 6. Correlation scatterplot of traffic volume and travel time data.

3.2 Model development

Next, to facilitate the identification of patterns in the data, a subset of relevant input features was selected based on which the algorithms categorized the data. The goal of feature selection in Machine Learning is to optimize the fit to the data of the problem by identifying and discarding features that provide unnecessary information. Consequently, the necessary features were generated, deemed capable of describing the relationship between travel time and traffic volume, resulting in an acceptable prediction error in each case. Finally, these feature values were added to the database. In Table 2 the features that were utilized in the models are presented.

Table 2
Input features of ML models.

Weekday	The value of 1 is assigned for weekday and 0 for weekends
Rush_hour	The value of 1 is assigned for morning peak hours (8:00 - 11:00), 2 for afternoon peak hours (15:00 - 19:00), and 0 for all other hours of the day.
W+R2	Combination of the two features above. For weekdays, values of 1 or 2 are assigned depending on the peak hour. For the weekend, it takes the value 0 regardless of the observation time.
min_max	The difference between the maximum and minimum travel time estimates from Google Maps database (duration_max - duration_min).
diff	The difference of two consecutive travel time data points.

The goal of this study is to forecast travel time and traffic volume using Machine Learning algorithms, specifically Decision Trees and GBDT. The development process of the algorithms was largely identical across both prediction scenarios. Traffic volume and travel time were not only the target variables for the models but were also utilized as features to predict one another. The exact features that were used in each model will be presented subsequently.

Its worth noting that during the model development phase, various features and their combinations were tested to achieve optimal predictive performance. Travel time prediction models exhibited acceptable prediction errors and demonstrated good fit to the data with minimal experimentation. In the context of traffic volume prediction, all tested feature combinations failed to yield the desired outcome. Notably, the normalization of traffic volume data in the database significantly contributed to reducing prediction errors. Using the logarithm of traffic volume data resulted in decreased variance in the database. This normalization technique is commonly applied in Machine Learning models, particularly when certain algorithms are sensitive to the scale of the data points. In order to evaluate the models prediction accuracy, the predicted traffic volume values were converted back to their original scale. To further mitigate the error, a statistical analysis of the database was conducted. The nearly identical traffic volume values of loop detectors MS407 and MS423 were averaged and grouped according to their corresponding travel time values found in the database. Next, the statistical measures for traffic volume values, specified in Table 3, were calculated.

Table 3

Traffic volume statistics per travel time class.

Travel Time (minutes)	Count	Median	Mean	Mode	Standard Deviation
4	496	280.75	334.17	120	213.05
5	295	737.50	780.12	560	263.85
6	279	1031.00	1000.73	1080	153.00
7	249	1063.00	1054.40	1080	82.15
8	220	1057.75	1036.74	1000	96.47
9	177	1041.50	1032.03	1080	83.24
10	96	1040.00	1042.26	1040	69.13
12	36	1045.25	1034.42	1136.50	119.58

Subsequently, the z-score, a statistical metric indicating the number of standard deviations a value deviates from the mean, was computed for the entire dataset. For each travel time value, the count of observations deviating by more than 1, 2, and 3 standard deviations was documented. Lastly, the percentage of these deviations, relative to the total observations within the respective class, was calculated. Also, in Fig. 6, the count of distinct travel time values for each hour of the day is presented.

Table 4

Z-scores of travel time classes.

Travel Time (minutes)	$\|z\|>1$	$\|z\|>1$ (%)	$\|z\|>2$	$\|z\|>2$ (%)	$\|z\|>3$	$\|z\|>3$ (%)
4	157	0.317	18	0.036	6	0.012
5	120	0.407	5	0.017	0	0.000
6	72	0.258	10	0.036	5	0.018
7	45	0.181	10	0.040	4	0.016
8	34	0.155	13	0.059	5	0.023
9	32	0.181	7	0.040	3	0.017

10	30	0.313	6	0.063	1	0.010
12	3	0.083	2	0.056	1	0.028

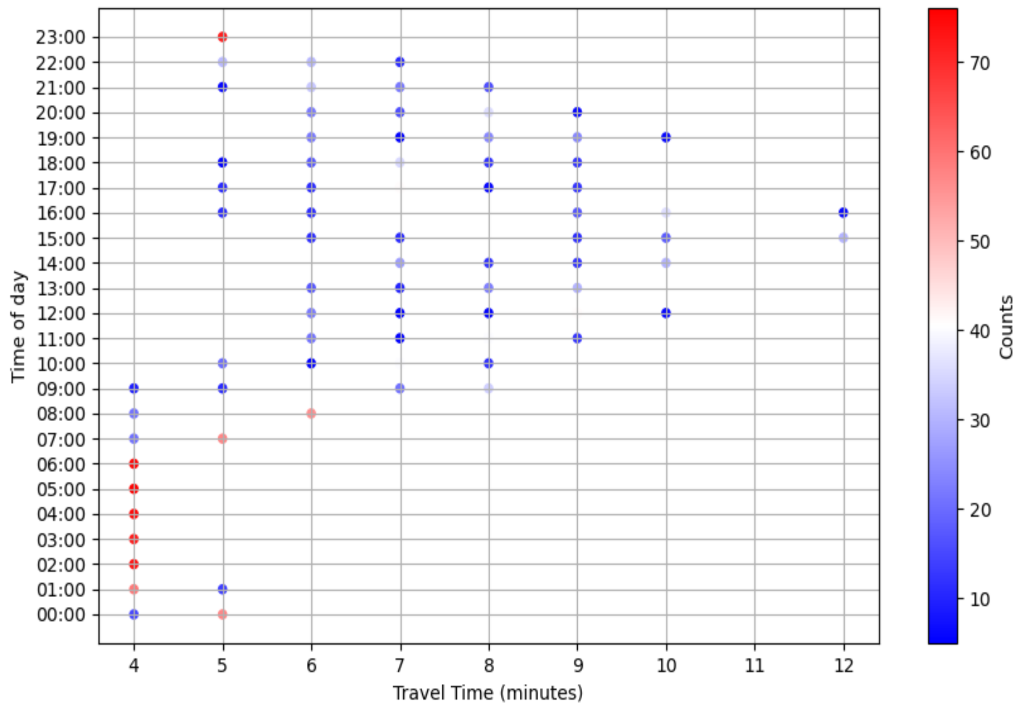

Fig. 7. Count of distinct travel time values for each hour of the day.

The analysis indicated that the models encountered difficulties when dealing with traffic volume values that correspond to low travel time duration. The calculation of the z-score revealed that especially for travel times of 4 and 5 minutes a larger percentage of traffic volume values exceed one standard deviation ($|z| > 1$) from the mean, despite the high standard deviation in these instances. To address this challenge, all records during the midnight hours from 00:00-3:00, characterized exclusively by travel times of 4 and 5 minutes, were excluded from the database. The removal of values was implemented to ensure a reasonable prediction error (MAPE <20%) and a satisfactory fit to the data ($R2 > 0.6$), avoiding an overly aggressive reduction of the original database. Coincidentally, predictions for midnight hours are of the least practical utility comparatively, at least in the case of traffic volume.

During the preparation process, a random split of the samples was performed, allocating 80% for model training (train set) and 20% for testing (test set). The hyperparameters for each model and prediction scenario are presented in Table 5.

Table 5
Hyper-parameters of ML models.

Hyper-parameters	Decision Trees	
	Travel Time prediction	Traffic Volume prediction
max_depth	8	10

min_samples_split	6	2
min_samples_leaf	3	2
	GBDT	
	Travel Time prediction	Traffic Volume prediction
n_estimators	100	100
max_depth	8	8

Next, using the same database, the prediction of travel time values for the same road segment was carried out using the BPR function. The mathematical formula of the function (8) was presented in a previous section. For its application, the calibration of the functions parameters was necessary. Given that the BPR function pertains to a specific road link, the approach adopted involved using the average traffic volume of both detectors. An algorithm was developed, utilizing the least squares method, to fit the function to the existing data. The initial values of α and β were set at the common values of 0.15 and 4 respectively.

Subsequently, free-flow travel time (T_{ff}) and road traffic capacity (C) were approximately estimated based on the existing data. The minimum observed travel time value in the data (Tff = 4 minutes) was designated as the free-flow travel time. The theoretical traffic capacity of the road segment probably surpasses the maximum observed traffic volume value in the data. Practical factors, such as traffic signals, the presence of a dedicated bus lane, and illegal parking in the right lane, significantly influence the actual road capacity. The exact calculation of this variable was deemed infeasible with the available data. It was also assumed that, for certain time intervals, demand exceeded the actual traffic capacity of the road (V/C > 1). Various traffic capacity values between the mean and maximum traffic volume values observed in the database were tested for the function. The actual road traffic capacity value was selected as 1050 vehicles, corresponding roughly to an average travel time of 10 minutes in the database. In each case, different traffic capacity values did not seem to significantly impact the prediction results. The optimization logic of the algorithm is to generate the smallest possible error, irrespective of whether the parameter values fall within "logical" bounds. With each new initial traffic capacity value, the algorithm provided different values for the parameters α and β, yet the final prediction result remained consistent. The significance of estimating a realistic traffic capacity value lies in finding suitable values for the parameters α and β for the application of the function to Alexandras Av., which is not the aim of this work. After establishing the terms of the BPR function, the final values of parameters α and β were devised by minimizing the error between the predicted values of the function and the actual travel time values. For the evaluation of the model, the same error and goodness-of-fit metrics were applied. Finally, a hybrid approach was carried out by combining the predictions of BPR and Machine Learning models. The travel time predictions generated by these models were averaged and evaluated for accuracy.

3.3 Results

Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) were utilized to evaluate the prediction accuracy, while R^2 was employed to evaluate the models fit to the data. Tables 6 showcases performance metrics and the importance level of each feature in the Machine Learning models designed for predicting traffic volume. Subsequent to these tables, diagrams depicting predicted versus actual values (Fig. 7, Fig. 8) are provided for both loop detectors, initially with the Decision Trees model and then with the GBDT model.

Table 6

Performance metrics of ML models for traffic volume prediction.

Metric	Decision Trees	GBDT
R^2	0.79	0.80

MAE	95.88	95.88
MAPE	14.55%	14.55%
Feature importance		
duration (min)	0.8056	0.8092
min_max	0.0960	0.0889
Weekday	0.0733	0.0796
Rush hour	0.0129	0.0059
W+R2	0.0017	0.0037
diff	0.0105	0.0127

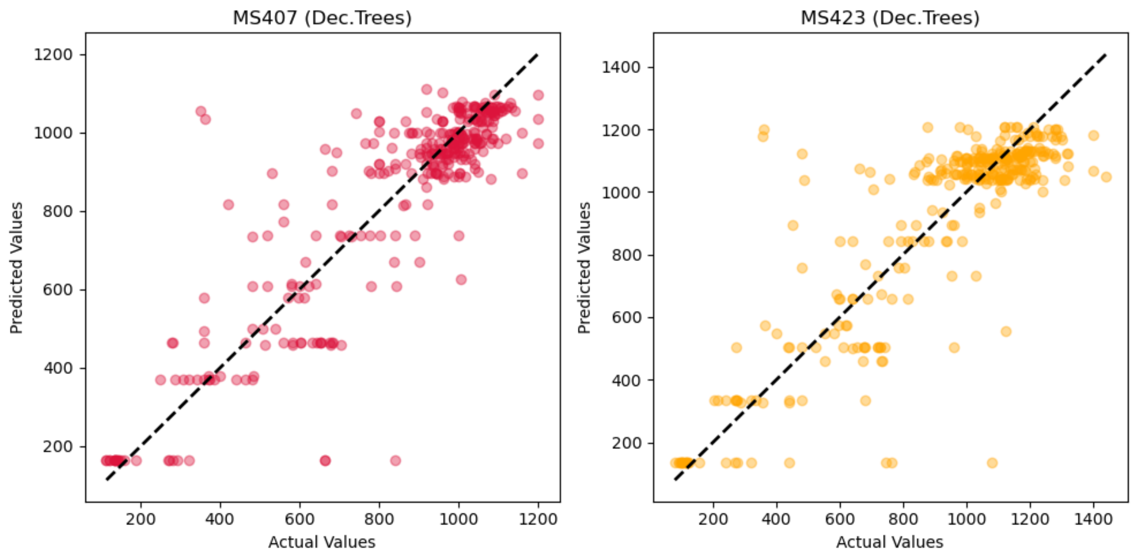

Fig. 8. Predicted versus actual traffic volume values for both detector loops (Decision Trees).

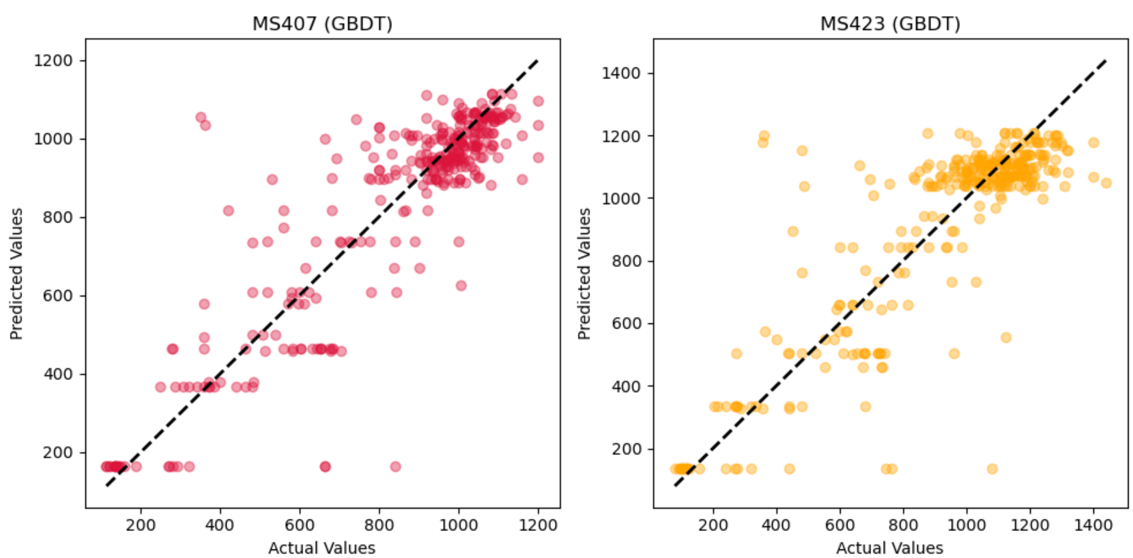

Fig. 9. Predicted versus actual traffic volume values for both detector loops (GBDT).

Table 7 presents the results for predicting travel time using both ML and BPR-based models. These are visually complemented by Figures 9, 10 and 11, providing a graphical representation of the findings, similar to the approach utilized for traffic volume.

Table 7
Performance metrics of trained models for travel time prediction.

Metric	Decision Trees	GBDT	BPR	BPR-DT	BPR-GBDT
R^2	0.82	0.76	0.48	0.68	0.66
MAE	0.69	0.69	1.08	0.86	0.89
MAPE	9.75%	9.75%	16.78%	12.87%	13.42%
	Feature importance				
MS423	0.1021	0.0942			
MS407	0.6768	0.6865			
Weekday	0.1299	0.1018			
W+R2	0.0912	0.1175			

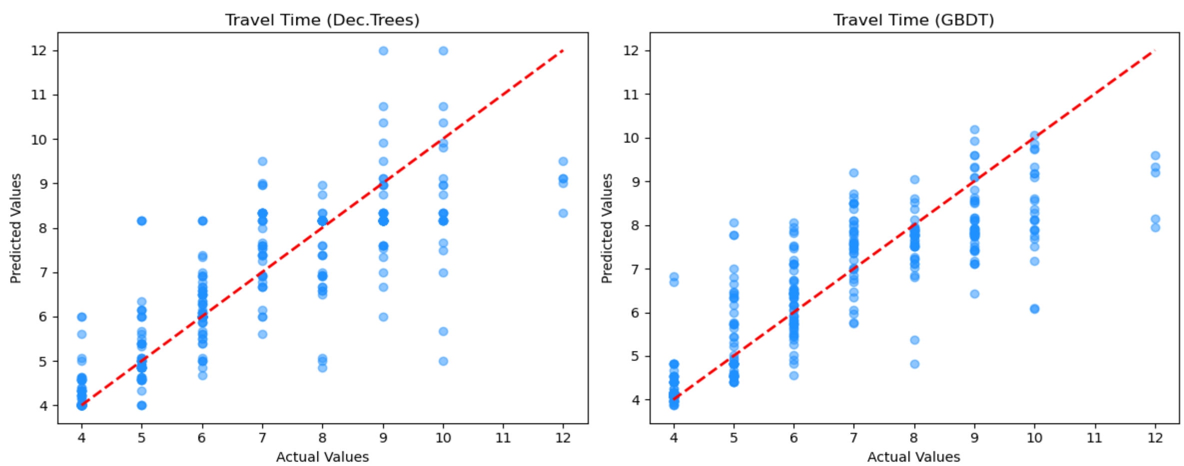

Fig. 10. Predicted versus actual travel time values of ML models.

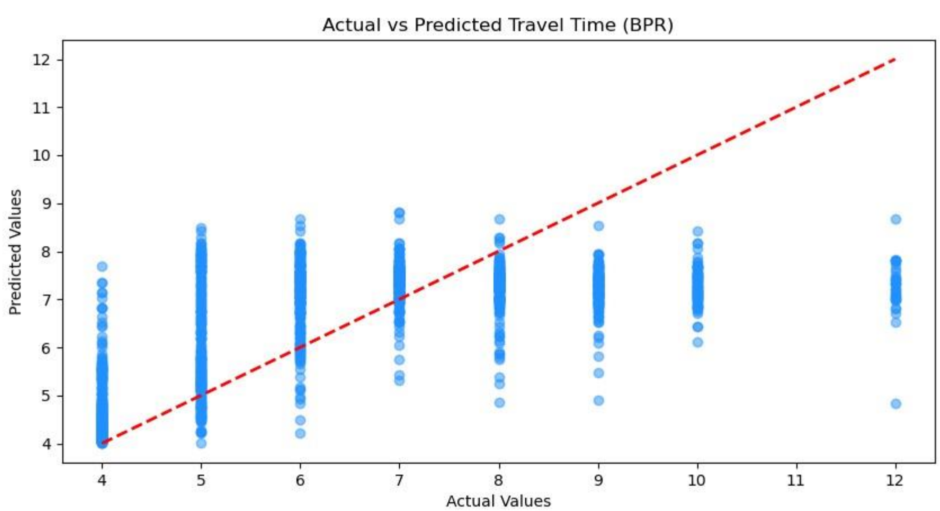

Fig. 11. Predicted versus actual travel time values (BPR)

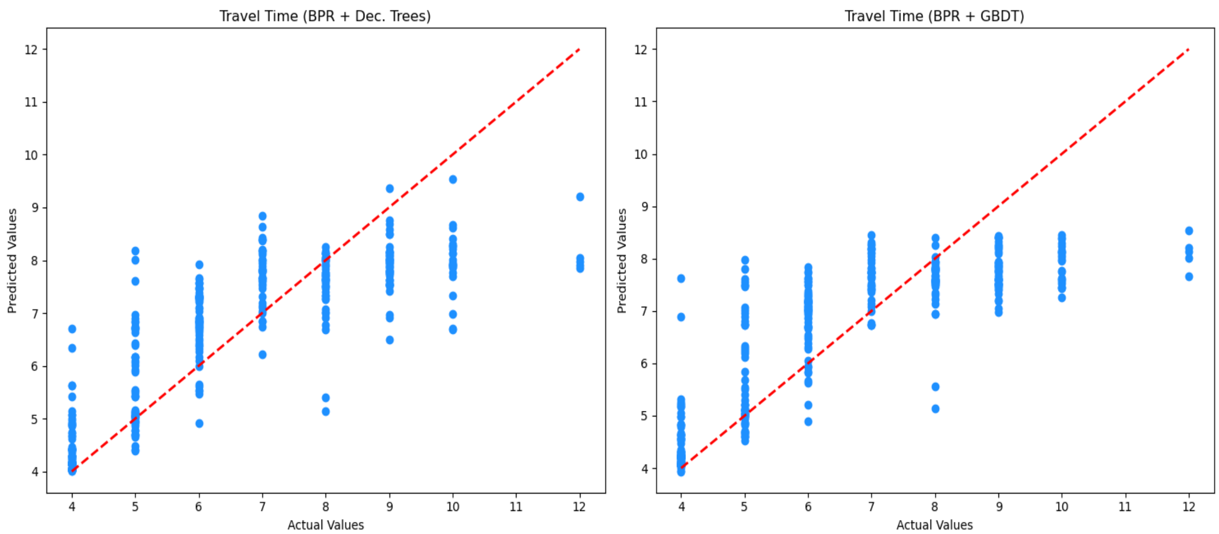

Fig. 12. Predicted versus actual travel time values for BPR-ML hybrid models.

The two Machine Learning models appear to exhibit similar performance, both in terms of prediction accuracy and fit to the data, regardless of the prediction variable. Finally, as expected, travel time proved to be the feature with the greatest importance on traffic volume prediction. Conversely, in the case of travel time, the volume data of loop detector MS407 emerged as the most influential variable.

The BPR function also demonstrated an acceptable MAPE in predicting travel time, albeit higher compared to Machine Learning methods. The R^2 metric indicates that the BPR function struggles to accurately capture the patterns observed in the data. This is further confirmed by Fig. 15. Particularly noticeable is the functions inability to predict travel time accurately at the maximum values of the database (9-12 minutes). By combining the travel time predictions of ML and BPR models, a small improvement is attained, although Machine Learning models still exhibit superior accuracy.

4. Conclusions

After summarizing the findings of this work, several key conclusions emerge. The Machine Learning models developed, consistently demonstrated impressive predictive capabilities for the targeted traffic variables in every scenario, despite their straightforward and simplistic design. Conversely, the BPR function-based prediction methods proved to be less effective, exhibiting lower accuracy in predictions and struggling to fit to the available data. Even with the application of the least squares optimization method, the travel time predictions using the BPR function still incurred greater errors compared to the Machine Learning approaches.

In light of the above, it can be deduced that employing comparable methods might prove effective in estimating traffic volume data, which is often more difficult to collect. In the specific case of this study, forecasting traffic volume values for a road link was achieved solely with knowledge of easily accessible travel time data. This underscores the notion that simple models with a concise set of variables can serve as valuable tools, though they may not universally represent the optimal strategy for every traffic prediction scenario.

Delving into the process of developing Shallow Machine Learning models, it became evident that data quality, quantity, and format played a pivotal role in ensuring high performance. In the context of traffic volume prediction, the models encountered a challenge due to the low variance in travel time values, which was effectively addressed through normalization and the removal of outliers.

It remains unclear whether the models employed in this study can replicate a similar predictive performance in different road segments. The application of analogous methodologies and models to different road sections could provide clarity on this matter. Moreover, the intricacies introduced by the geometry and spatial correlations between road links elevate the complexity of predicting traffic conditions at network level. While existing literature predominantly focuses on the less complex task of forecasting traffic parameters for individual road segments, emphasizing the development of models capable of comprehending and integrating the spatial dependencies among road links may facilitate more precise and consistent network-wide predictions

Expanding the scope of traffic forecasting by introducing additional variables could be a fruitful endeavor. Integrating weather data, accident statistics, impact of traffic signals, and other factors into Machine Learning models, coupled with established traffic parameters, holds the potential to enhance predictive capabilities. The models developed in this work could be refined by incorporating more extensive and diverse datasets, enabling additional feature combinations. In conjunction with the above, exploring the application of Deep Learning approaches also stands as a promising direction.

Lastly, the establishment of a comprehensive calibration framework for the BPR function, coupled with a rigorous definition of its terms and underlying assumptions, would significantly amplify its effectiveness as tools for traffic analysis. Only through this process could the capabilities of Volume-Delay Functions be integrated for the generation of Physics-Informed Machine Learning models. With technology advancing swiftly, it is evident that capitalizing on the ever-expanding wealth of information for the development of innovative traffic engineering applications should be central to both research endeavors and the management of transportation systems.

References

Barka, R. E., & Politis, I. (2021). Fitting Volume Delay Functions under interrupted and uninterrupted flow conditions at Greek urban roads. *European Transport - Trasporti Europei*, *83*. https://doi.org/10.48295/ET.2021.83.7

Barros, J., Araujo, M., & Rossetti, R. J. F. (2015). Short-term real-time traffic prediction methods: A survey. *2015 International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2015*, 132–139. https://doi.org/10.1109/MTITS.2015.7223248

Behrooz, H., & Hayeri, Y. M. (2022). Machine Learning Applications in Surface Transportation Systems: A Literature Review. *Applied Sciences (Switzerland)*, *12*(18). https://doi.org/10.3390/app12189156

Cheng, J., Li, G., & Chen, X. (2019). Research on travel time prediction model of freeway based on gradient boosting decision tree. *IEEE Access*, *7*, 7466–7480. https://doi.org/10.1109/ACCESS.2018.2886549

IEEE Staff. (2018). *2018 International Conference on Information, Communication, Engineering and Technology (ICICET)*. IEEE.

Janković, S., Uzelac, A., Zdravković, S., Mladenović, D., Mladenović, S., & Andrijanić, I. (n.d.). TRAFFIC VOLUMES PREDICTION USING BIG DATA ANALYTICS METHODS. *International Journal for Traffic and Transport Engineering*, *2021*(2), 184–198. https://doi.org/10.7708/ijtte2021.11(2).01

Laña, I., Sanchez-Medina, J. J., Vlahogianni, E. I., & Ser, J. Del. (2021). From data to actions in intelligent transportation systems: A prescription of functional requirements for model actionability. *Sensors (Switzerland)*, *21*(4), 1–34. https://doi.org/10.3390/s21041121

Liu, Y., Zhang, N., Luo, X., & Yang, M. (2021). Traffic Flow Forecasting Analysis based on Two Methods. *Journal of Physics: Conference Series*, *1861*(1). https://doi.org/10.1088/1742-6596/1861/1/012042

Liu, Z., Li, Z., Wu, K., & Li, M. (2018). Urban Traffic Prediction from Mobility Data Using Deep Learning. *IEEE Network*, *32*(4), 40–46. https://doi.org/10.1109/MNET.2018.1700411

Maerivoet, S., & De Moor, B. (2005). *Transportation Planning and Traffic Flow Models*.

Mahdi, M. B., Alrawi, A. K., & Leong, L. V. (2022). Compatibility between delay functions and highway capacity manual on Iraqi highways. *Open Engineering*, *12*(1), 359–372. https://doi.org/10.1515/eng-2022-0022

Manibardo, E. L., Lana, I., & Ser, J. Del. (2022). Deep Learning for Road Traffic Forecasting: Does it Make a Difference? *IEEE Transactions on Intelligent Transportation Systems*, *23*(7), 6164–6188. https://doi.org/10.1109/TITS.2021.3083957

Marshall, N. L. (2018). Forecasting the impossible: The status quo of estimating traffic flows with static traffic assignment and the future of dynamic traffic assignment. *Research in Transportation Business and Management*, *29*, 85–92. https://doi.org/10.1016/j.rtbm.2018.06.002

Mehbub Anwar, A. H. M., Fujiwara, A., & Zhang, J. (n.d.). *Estimation of BPR Function for Dhaka City by Incorporating the Influence of Illegal On-Street Parking and Street-Occupancy*.

Pan, Y., Guo, J., Chen, Y., Abbasi, M., List, G., & Zhou, X. (n.d.). *A review of volume-delay functions: Connecting theoretical fundamental, practical deployment and emerging applications A review of volume-delay functions: Connecting theoretical*. https://ssrn.com/abstract=4143355

Petrik, O., Moura, F., & De Abreu E Silva, J. (2014). The influence of the volume-delay function on uncertainty assessment for a four-step model. *Advances in Intelligent Systems and Computing*, *262*, 293–306. https://doi.org/10.1007/978-3-319-04630-3_22

Qiu, B., & Fan, W. (2021). Machine learning based short-term travel time prediction: Numerical results and comparative analyses. *Sustainability (Switzerland)*, *13*(13). https://doi.org/10.3390/su13137454

Saberi, M., & Student, G. (n.d.). *A Study of Freeway Volume-to-Capacity Ratio Based Travel Time Approximations Using Archived 1 Loop Detector Data 2 3*.

Suh, S., & John Kim, T. (1990). A HIGHWAY CAPACITY FUNCTION IN KOREA: MEASUREMENT AND CALIBRATION. In *Tmnspn. Res.-A* (Vol. 3).

TRB. (2011). *Dynamic Traffic Assignment: A Primer*. www.TRB.org

Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, *43*, 3–19. https://doi.org/10.1016/j.trc.2014.01.005

Wang, Y., Zhang, D., Liu, Y., Dai, B., & Lee, L. H. (2019). Enhancing transportation systems via deep learning: A survey. In *Transportation Research Part C: Emerging Technologies* (Vol. 99, pp. 144–163). Elsevier Ltd. https://doi.org/10.1016/j.trc.2018.12.004

Wu, X. B. (n.d.). *Characterization and calibration of volume-to-capacity ratio in volume-delay functions on freeways based on a queue analysis approach Food supply chain management View project DTALite 2.0 strategic and semi-dynamic transportation flow assignment and simulation View project*. https://www.researchgate.net/publication/347967287

Yang, S., Wu, J., Du, Y., He, Y., & Chen, X. (2017). Ensemble Learning for Short-Term Traffic Prediction Based on Gradient Boosting Machine. *Journal of Sensors*, *2017*. https://doi.org/10.1155/2017/7074143

Yuan, T., Da Rocha Neto, W., Rothenberg, C. E., Obraczka, K., Barakat, C., & Turletti, T. (2022). Machine learning for next-generation intelligent transportation systems: A survey. *Transactions on Emerging Telecommunications Technologies*, *33*(4). https://doi.org/10.1002/ett.4427

Zhang, Y., & Haghani, A. (2015). A gradient boosting method to improve travel time prediction. *Transportation Research Part C: Emerging Technologies*, *58*, 308–324. https://doi.org/10.1016/j.trc.2015.02.019