Proceedings of 14th IOE Graduate Conference
Peer Reviewed
Year: 2023 Month: December Volume: 14
ISSN: 2350-8914 (Online), 2350-8906 (Print)

Evaluating Multiple Linear Regression for Travel Time Prediction in Mixed Traffic on a Two-Lane Undivided Highway

Sanjay Luitel [a], Pradeep Kumar Shrestha [b], Hemant Tiwari [c]

[a, b] *Department of Civil Engineering, Pulchowk Campus, Tribhuvan University, Nepal*
[c] *Society of Transport Engineers Nepal-SOTEN*
[a] sanjayluitel2074@gmail.com, [b] pradeep.shrestha@pcampus.edu.np, [c] hemu.ioe@gmail.com

Abstract

Over time, considerable efforts have been committed to enhancing the forecasting of travel time along the corridor, employing a wide range of factors. Nevertheless, forecasting travel times along these routes continues to be an inherently difficult task due to the intricate interactions among numerous variables that are frequently challenging to comprehensively gather. This complexity is especially evident on undivided roads, where access to the route is unrestricted, resulting in a heightened presence and impact of various influential factors on travel time. The present study is focused on the development of a travel time prediction model for both directions of the Dhankhola-Bhaluwang road section, which is a two-lane, two-way undivided rural highway section of the East-West Highway. Using an extensive analysis of 72-hour datasets on vehicle travel time, gathered from traffic volume counts survey, this study evaluates the effectiveness of a multiple linear regression model on predicting travel time, taking into account through traffic volume, opposing traffic volume and the percentage of heavy vehicles in through traffic. The regression results indicate a moderate relationship between the variation in travel time and the variation in independent variables. The statistical error test results demonstrate that the model is significant and has a high capability for accurate predictions.

Keywords

Travel Time Prediction Model, Travel Time, Multiple Linear Regression

1. Introduction

Travel time refers to the duration needed to journey from one end of the corridor to the opposite end. Corridor travel time holds significance in transportation planning, aiding in the evaluation of various transportation infrastructures and the assessment of potential alterations to the transportation infrastructure's effects. Accurate travel duration data is vital for identifying and appraising operational issues within highway facilities [1]. It serves as a crucial metric in highway transportation planning and management, exerting an influence on driver behavior and overall transportation efficiency. Efficient traffic management hinges on understanding the myriad factors affecting travel time and traffic flow. The precise real-time prediction of travel time, along with the reliability of these forecasts, is essential for implementing technologies and policies effectively [2]. The ability to accurately predict future travel times for links within transportation networks is vital for numerous applications within the field of Intelligent Transportation Systems (ITS) [3]. Traffic parameters, including but not limited to traffic volume, speed, and travel time, play a pivotal role in the design and planning of road facilities. Travel time is affected by a variety of factors including traffic volume, percentage of heavy vehicles, weather, time of day, road characteristics and side friction [4], [5].

Many research studies have used a variety of forecasting techniques, including both multiple linear and nonlinear approaches, to formulate equations and models for the prediction of travel time [4], [6]. In 2003, Zhang and Rice proposed a method to predict freeway travel times using a linear model in which the coefficients vary as smooth functions of the departure time. The result showed prediction errors range from about 8% at zero lag to 13% at a time lag of 30 min or more [7]. In 2017, Roh et al. conducted a study to investigate the impact of heavy vehicles (HVs) on traffic flows using real-time traffic data, analyzing the relationship between average speed, heavy vehicle (HV) ratio, flow rate, and the number of lanes. Their findings demonstrated that the average speed decreased as the flow rate and HV ratio increased, across six-lane, eight-lane, and four-lane highways [8]. Most of the studies were focused on freeways or multi-lane divided highways where the effect of oncoming vehicles no longer affects travel time.

In this study, a predictive model, multiple linear regression (MLR) is applied, to predict travel times for both directions on the Dhankhola-Bhaluwang road section, which is a two-lane, two-way undivided highway and part of the East-West highway (H01) in Nepal.

2. Research Objective and Scope

The objective of the current research is to develop a multiple linear regression model for predicting travel time using the relationship among through traffic volume, opposing traffic volume, the proportion of heavy vehicles in through traffic, and travel time on a heterogeneous condition highway for both directions. The research also aims to evaluate the accuracy of the multiple linear regression model. Performance evaluation of the models is conducted using several performance indices, such as MAE, MSE, R^2, MAPE, and RMSE, to assess accuracy and reliability.

3. Study Area

This research is focused on the Dhankhola-Bhaluwang road section (see Figure 1), a part of the East-West highway (H01) covering a distance of 12.9 kilometers. This roadway extends from the eastern point in Dhankhola (located at 82°48’7.36"E, 27°47’3.90"N) to its western terminus in Bhaluwang (at 82°44’52.76"E, 27°50’19.05"N), running from east to west. Situated in the western Terai region of Nepal, the road exhibits undulating terrain in certain areas and serves as a vital connection between Kapilvastu and Dang. The road consists of a two-lane, two-way undivided carriageway surfaced with blacktop but lacks both a shoulder and a footpath.

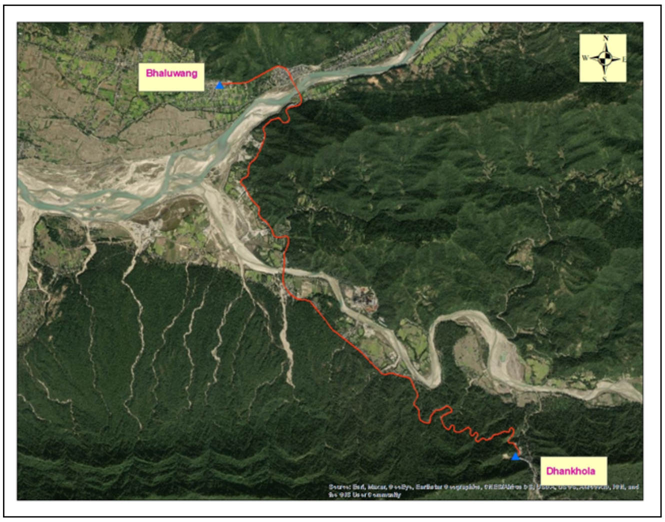

Figure 1: Alignment of Dhankhola-Bhaluwang Road Section

4. Methodology

The following approach and methodology were adopted to meet the objectives of the study. The series of activities taken in the methodology to accomplish the study can be divided into three steps as shown in Figure 2.

4.1 Literature Review

In the course of conducting this research, a thorough examination of various papers pertaining to the study's subject matter was undertaken. The primary objectives of this literature review were to identify potential variables that could be associated with the prediction of travel time. For this study, three independent variables were selected to predict travel time: through traffic volume [9], opposing traffic volume [10] and percentage of heavy vehicles [11]. Furthermore, standard data sheets were prepared for the collection of traffic volume data and travel time data.

4.2 Data Collection and Extraction

The data collection process began with the preparation of datasheets for both traffic volume surveys and travel time data. A videographic recording survey was conducted to capture video footage of traffic volume, enabling the collection of relevant traffic data. Vehicle movements within corridors were monitored over a three-day period, spanning from Monday, June 12th, to Wednesday, June 14th, 2023, using cameras.

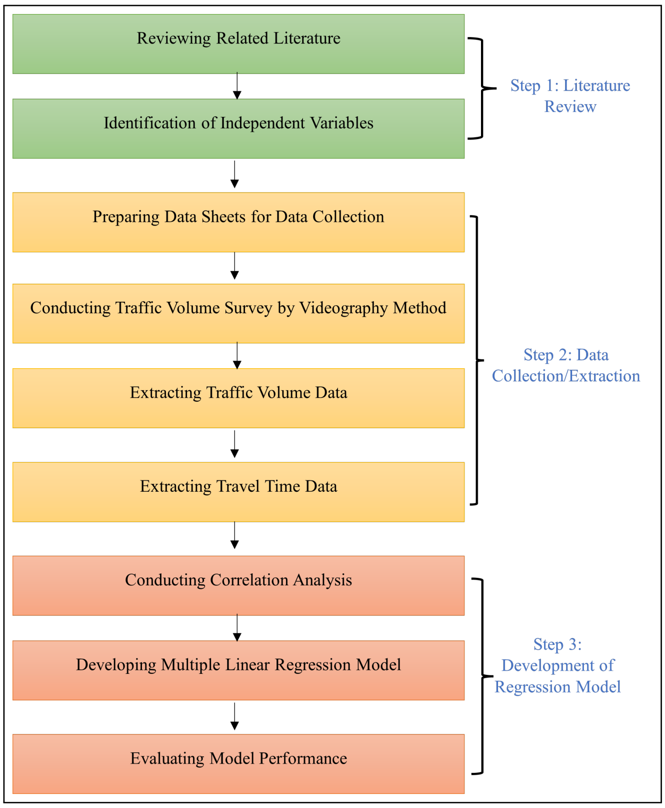

Figure 2: Methodological Framework of the Study

Specifically, two cameras were strategically placed: one at the entry point and another at the exit to document the directional flow of traffic. The schematic representation of camera placements for video recording is illustrated in Figure 3.

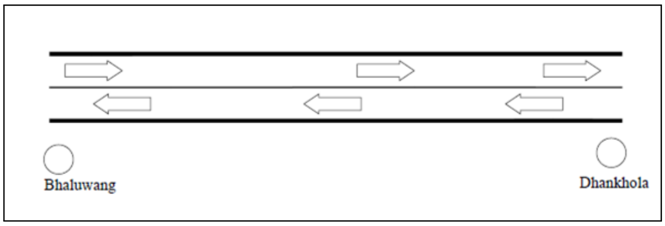

Figure 3: Typical Layout for camera's position for recording video footages

Furthermore, traffic volume counts were conducted manually at two stations by reviewing the recorded video footage. The classified and directional traffic volume data were extracted at 30-minute intervals, with only motorized traffic taken into account for the volume count. To assess the impact of each classified vehicle in mixed traffic conditions, the counts of various vehicle types were transformed into standardized PCU units [12]. This transformation was achieved by multiplying the total count of each vehicle type by its corresponding PCU factor, as indicated in Table 1.This process was applied in both directions to determine the volumes of through traffic and opposing traffic. This implies that the through traffic volume in one direction represents the opposing traffic volume in the other direction, and vice versa. Similarly, the percentage of heavy vehicles in traffic is also determined by dividing the number of heavy vehicles (denoted by * in Table 1) by the total number of vehicles. This was also adopted for both directions.

Table 1: Vehicle Equivalency Factors (Source : [12])

SN	Vehicle Type	PCU Equivalency Factor
1	Car	1.0
2	Heavy Truck*	3.0
3	Light Truck*	1.5
4	Multi Axle Truck*	4.0
5	Tractor (farm tractor)*	1.5
6	Bus*	3.0
7	Minibus*	2.5
8	Microbus	1.5
9	Utility	1.0
10	Four Wheel Drive	1.0
11	Motorcycle	0.5
12	Three wheeler (Auto Rickshaw)	0.75
13	Power Tiller	1.5

Furthermore, the task involved noting the arrival and departure times of the vehicles (excluding motorcycles and non-motorized vehicles) within the corridor by viewing recorded video footage of two locations. This process was accomplished through a thorough examination of recorded video footage captured from two locations. The relevant data was then discerned by visually identifying each vehicle present in the footage. Subsequently, travel times were calculated using arithmetic operations involving the recorded arrival and departure times shown in Eq.1.

$$\text{Travel Time} = \text{Departure Time} - \text{Arrival Time} \quad (1)$$

A total of 2882 vehicle travel time samples were collected for Dhankhola to Bhaluwang direction while the number of sample travel time data for the Bhaluwang to Dhankhola direction was 2887. Then travel times data were averaged at 30-minute intervals to match with other variables.

4.3 Model Development

The three-day data, collected at 30-minute intervals, was organized into four columns: travel time in minutes, the percentage of heavy vehicles in through traffic, through traffic volume in PCU/hr, and opposing traffic volume in PCU/hr. There are a total of 144 rows for both directions, and separate analyses were conducted for each direction. Then correlation analysis was done to determine to measure and strength of the relationships between the variables. The dataset was then separated into two files: one containing a two-day dataset for creating a model and the other containing a third-day dataset for validating the model. The multiple linear regression model was fitted to the data. Using least squared regression, the best-fitted line was determined using Regression tools within Excel.

Additionally, the model's output was examined using various performance indices, including mean absolute error (MAE), mean square error (MSE), mean absolute percentage error (MAPE), root mean square error (RMSE), coefficient of correlation (R), and coefficient of determination (R^2) by using Eq. 2, Eq. 3, Eq. 4, Eq. 5, Eq. 6 and Eq. 7 [13], respectively.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|OV_i - PV_i| \quad (2)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(OV_i - PV_i)^2 \quad (3)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{(OV_i - PV_i)^2}{OV_i} \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(OV_i - PV_i)^2}{n}} \quad (5)$$

$$R = \frac{\sum_{i=1}^{n}(OV_i - \text{OV\_mean})(PV_i - \text{PV\_mean})}{\sqrt{\sum_{i=1}^{n}(OV_i - \text{OV\_mean})^2 \sum_{i=1}^{n}(PV_i - \text{PV\_mean})^2}} \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(OV_i - PV_i)^2}{\sum_{i=1}^{n}(OV_i - \text{PV\_mean})^2} \quad (7)$$

Where, OV_i, PV_i and n represent observed value, predicted value and total number of samples respectively.

5. Description of Database

In this study, a total of 144 data sets were gathered for both directions. The analysis of the data was carried out based on Four parameters. Three parameters were used as input parameters in the following order: through traffic volume, opposing traffic volume, percentage of heavy vehicles in through traffic, and travel time as output parameters. The statistical characteristics such as mean, standard deviation (S.D), minimum and maximum value are expressed for individual variables shown in Table 2. Table 2 illustrates that the adopted dataset covers a wide range of parameters, and the standard deviation (S.D.) represents the distribution of the dataset around the mean values.

For the Dhankhola to Bhaluwang route, the mean travel time is 26.11 minutes, with a standard deviation of 2.27 minutes, ranging from a minimum of 20.53 minutes to a maximum of 31.43 minutes. The through volume ranges from 26 PCU/Hr to 206 PCU/Hr, with a mean of 105.30 PCU/Hr, and the percentage of heavy vehicles in through traffic averages at 0.45 Decimal Percentage.

In contrast, the Bhaluwang to Dhankhola route has a mean travel time of 28.40 minutes, with a standard deviation of 3.17 minutes, ranging from 19.15 minutes to 35.87 minutes. The through volume varies between 22 PCU/Hr and 287 PCU/Hr, with a mean of 128.69 PCU/Hr, and the percentage of heavy vehicles in through traffic averages at 0.35 Decimal Percentage.

Table 2: Statistical Characteristics of Data

Variable	Unit	Mean	Std Dev	Min	Max
Dhankhola to Bhaluwang					
Travel Time	Minutes	26.11	2.27	20.53	31.43
Through Volume	PCU/Hr	105.30	35.80	26.00	206.00
Opposing Volume	PCU/Hr	128.69	61.61	22.00	287.00
Percentages of Heavy Vehicles in Through Traffic	Decimal Percentage	0.45	0.18	0.15	0.97
Bhaluwang to Dhankhola					
Travel Time	Minutes	28.40	3.17	19.15	35.87
Through Volume	PCU/Hr	128.69	61.61	22.00	287.00
Opposing Volume	PCU/Hr	105.30	35.80	26.00	206.00
Percentages of Heavy Vehicles in Through Traffic	Decimal Percentage	0.35	0.15	0.17	0.88

Table 3: Pearson's correlation coefficient between two variables

Correlation Matrix	Travel Time	Through Volume	Opposing Volume	Proportional of Heavy Vehicle
Bhaluwang to Dhankhola				
Travel Time	1.00	0.55	0.48	0.43
Through Volume	0.55	1.00	0.24	0.30
Opposing Volume	0.48	0.24	1.00	0.31
Percentages of Heavy Vehicles in Through Traffic	0.43	0.30	0.31	1.00
Dhankhola to Bhaluwang				
Travel Time	1.00	0.47	0.55	0.43
Through Volume	0.47	1.00	0.25	0.10
Opposing Volume	0.55	0.25	1.00	0.15
Percentages of Heavy Vehicles in Through Traffic	0.43	0.10	0.15	1.00

6. Result and Discussion

6.1 Correlation Analysis

Correlation analysis is a method used to examine the relationships between different variables. In this research, Pearson's correlation coefficient is employed to understand the associations among the variables. The Pearson correlation coefficient ρ_{xy} is calculated by dividing the covariance of two variables (COV (X, Y)) by the product of their individual standard deviations σ_x and σ_y [14].

The range of Pearson's correlation coefficient ρ_{xy} lies within [-1,1]. The high values of ρ_{xy} denote that the parameters are related and have a high effect on the output parameter while σ_x = -1 means that variables are not directly correlated i.e., the relationship between the variables is inverse. Also, if σ_x= 0, it represents parameters X and Y are linearly independent since Pearson's correlation coefficient is capable of showing only a linear relationship. Table 3 displays the relationships between the adopted variables for two directions of the road: from Bhaluwang to Dhankhola and from Dhankhola to Bhaluwang. In both directions, travel Time exhibits a moderately strong positive correlation with through volume, opposing volume, and the percentage of heavy vehicles in through traffic. This implies that as these independent variable increases, travel time tends to increase as well. Therefore, all three independent variables can be used to predict the travel time of the selected corridor.

6.2 Multiple Linear Regression

Multiple Linear regression is carried out to develop a prediction equation that allows the estimation of the value of travel time base on the value of through traffic, opposing traffic volume and proportional of heavy vehicles in through traffic. The models from multiple linear regression for Dhankhola to Bhaluwang direction and Bhaluwang to Dhankhola direction are shown in Eq. 8 and Eq. 9 respectively.

$$\begin{aligned}\text{Travel Time (min)} =&19.25 + 0.026 \times Q_T + 0.019 \times Q_O \\ &+2.803 \times \rho_T\end{aligned} \tag{8}$$

$$\begin{aligned}\text{Travel Time (min)} =&16.85 + 0.035 \times Q_T + 0.039 \times Q_O \\ &+5.97 \times \rho_T\end{aligned} \tag{9}$$

Where Q_T indicates through traffic volume (pcu/hr), Q_O indicates opposing traffic volume (pcu/hr), and ρ_T indicates the percentage of heavy vehicles in through traffic (decimal).

In the analysis of travel times between Dhankhola and Bhaluwang in two directions, statistical metrics were employed to assess model performance shown in Table 4. The R-squared values for both directions were moderate. This suggests a moderate level of predictive power, but there is still a significant amount of unexplained variance. The Mean Absolute Error (MAE) ranged from 0.95 to 1.52 minutes, with the Dhankhola to Bhaluwang direction showing a lower error. Meanwhile, the Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) suggested more significant errors in the Bhaluwang to Dhankhola direction. The Mean Absolute Percentage Error (MAPE) ranged from 3.68% to 5.58%, with the Dhankhola to Bhaluwang direction having a lower percentage error. A lower MAPE percentage suggests that, on average, the predictions are very close to the actual values, indicating a high level of precision and reliability in the model's predictions. The F-statistics indicated the overall

Table 4: Evaluation Metrics of Multiple Linear Regression Model

Direction	R^2	MAE	MSE	RMSE	MAPE	F-Statistics
Dhankhola to Bhaluwang	0.44	0.95	1.32	1.15	3.68%	24.32
Bhaluwang to Dhankhola	0.47	1.52	3.67	1.92	5.58%	27.39

significance of the regression models, with both directions showing reasonably high values greater than F-critical (2.7) at a significance level of 0.05. These metrics collectively indicate the predictive accuracy and model fit for the travel time data, with some variations between the two travel directions.

6.3 Travel Time Model Validation

Linear regression analysis was conducted to assess the significance of both observed data (which were not used in the model's development) and predicted data (obtained using the developed model). Figure 4 illustrates the results of the linear regression analysis of observed and predicted data, revealing an R^2 value of 0.4494, which is closely aligned with the model's R^2 value of 0.44. This suggests that approximately 44.94% of the variability in forecasting travel time can be accounted for by the observed travel time during validation for the Dhankhola to Bhaluwang direction. Similar results are depicted in Figure 5, which shows an R^2 value of 0.4573, also close to the model's R^2 of 0.47, indicating that around 45.7% of the variability in forecasting travel time can be explained by the observed travel time during validation for Bhaluwang to Dhankhola direction.

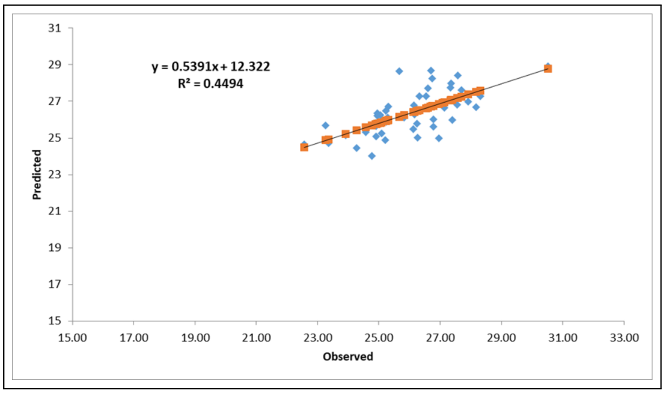

Figure 4: Validation of prediction model for Dhankhola to Bhaluwang direction

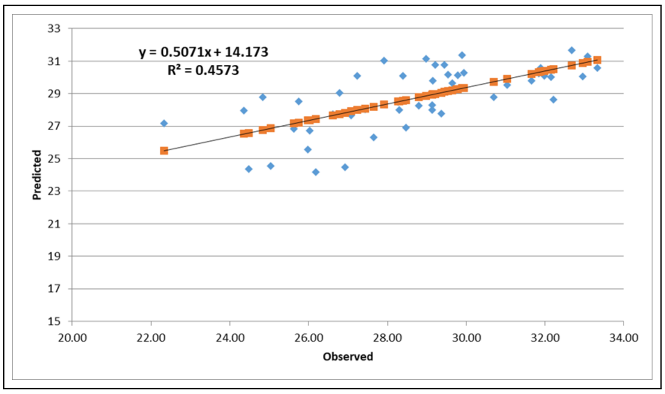

Figure 5: Validation of prediction model for Bhaluwang to Dhankhola direction

7. Conclusion and Recommendation

The objective of this research was to develop a travel time prediction model using the multiple linear regression technique for both directions on the Dhankhola-Bhaluwang road section, a two-lane, two-way undivided highway. In this study, a comprehensive analysis of three-day datasets on vehicle travel times was conducted using data gathered from a traffic volume count survey. The developed travel time prediction model illustrates the correlation between travel time and independent variables: through traffic volume, opposing traffic volume, and the percentage of heavy vehicles in through traffic. Pearson's correlation coefficient between dependent and independent variables indicates a moderate relationship between these variables. The multiple regression model demonstrates a moderate relationship in the variance of travel time and the independent variables, with an R-squared (R_2) value of 0.44 for the Dhankhola to Bhaluwang direction and an R_2 value of 0.471 for the Bhaluwang to Dhankhola direction. For the Dhankhola-Bhaluwang direction, the Mean Absolute Error (MAE) was 0.95 minutes, Mean Squared Error (MSE) was 1.32 minutes_2, Root Mean Squared Error (RMSE) was 1.15 minutes, and Mean Absolute Percentage Error (MAPE) was 3.68%. In the Bhaluwang to Dhankhola direction, the MAE was 1.52 minutes, MSE was 3.67 minutes_2, RMSE was 1.92 minutes, and MAPE was 5.58%. A lower MAPE percentage suggests that, on average, the predictions are very close to the actual values, indicating a high level of precision and reliability in the model's predictions. This study demonstrates the potential of the multiple linear regression model for accurately predicting travel times.

Incorporating side friction characteristics, such as roadside parking, pedestrian involvement, non-motorized activity, and access road conditions, may lead to more accurate travel time predictions, especially for unrestricted two-lane, two-way undivided roads. Furthermore, incorporating weather and time-of-day characteristics can further improve travel time predictions.

Acknowledgments

The study team express their gratitude to the Society of Transport Engineers Nepal-SOTEN for providing the research grant that facilitated data collection. The authors are equally thankful to all individuals who supported directly and indirectly throughout the completion of this research.

References

[1] Daniel Billings and Jiann-Shiou Yang. Application of the arima models to urban roadway travel time prediction-a case study. In *2006 IEEE International Conference on*

Systems, Man and Cybernetics, volume 3, pages 2529–2534. IEEE, 2006.

[2] Shuguan Yang and Sean Qian. Understanding and predicting travel time with spatio-temporal features of network traffic flow, weather and incidents. *IEEE Intelligent Transportation Systems Magazine*, 11(3):12–28, 2019.

[3] Simon Oh, Young-Ji Byon, Kitae Jang, and Hwasoo Yeo. Short-term travel-time prediction on highway: a review of the data-driven approach. *Transport Reviews*, 35(1):4–32, 2015.

[4] Matthew P D'Angelo, Haitham M Al-Deek, and Morgan C Wang. Travel-time prediction for freeway corridors. *Transportation Research Record*, 1676(1):184–191, 1999.

[5] Shahriar Afandizadeh, Nastaran Nasiri, and Hamid Mirzahossein. Evaluating the effect of heavy vehicles on traffic of two-way two-lane road in mountainous area of iran as a developing country. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, 46(4):3387–3398, 2022.

[6] Pranjal Sharma, Prashant Singh, and Plash Upreti. A knowledge based travel time prediction using regression technique. *International Journal of Innovative Science and Research Technology*, 6(7):2456–2165, 2021.

[7] Xiaoyan Zhang and John A Rice. Short-term travel time prediction. *Transportation Research Part C: Emerging Technologies*, 11(3-4):187–210, 2003.

[8] Chang-Gyun Roh, Bum-Jin Park, and Jonghoon Kim. Impact of heavy vehicles on highway traffic flows: case study in the seoul metropolitan area. *Journal of Transportation Engineering, Part A: Systems*, 143(9):05017008, 2017.

[9] Huizhao Tu, JWC Van Lint, and Henk J Van Zuylen. Impact of traffic flow on travel time variability of freeway corridors. *Transportation Research Record*, 1993(1):59–66, 2007.

[10] Gowri Asaithambi and Gugulothu Shravani. Overtaking behaviour of vehicles on undivided roads in non-lane based mixed traffic conditions. *Journal of traffic and transportation engineering (English edition)*, 4(3):252–261, 2017.

[11] Umama Ahmed, Alexander Drakopoulos, and ManWo Ng. Impact of heavy vehicles on freeway operating characteristics under congested conditions. *Transportation research record*, 2396(1):28–37, 2013.

[12] Highway management information system (HMIS).

[13] InDataLabs. [Online].

[14] Richard Taylor. Interpretation of the correlation coefficient: a basic review. *Journal of diagnostic medical sonography*, 6(1):35–39, 1990.