TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

THESIS NO: T13/078

Assessing Travel Time Prediction Models for Mixed Traffic on a Two-Lane Highway: A Case Study of the Dhankhola-Bhaluwang (H01) Road Section

by

SANJAY LUITEL

A THESIS

SUBMITTED TO THE DEPARTMENT OF CIVIL ENGINEERING

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE IN TRANSPORTATION ENGINEERING

DEPARTMENT OF CIVIL ENGINEERING

LALITPUR, NEPAL

DECEMBER, 2023

COPYRIGHT

…………………………………..

Head

Department of Civil Engineering

Pulchowk Campus, Institute of Engineering

Lalitpur, Kathmandu

Nepal

TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

DEPARTMENT OF CIVIL ENGINEERING

The undersigned certify that they have read and recommended to Institute of Engineering for acceptance, a thesis entitled "**Assessing Travel Time Prediction Models for Mixed Traffic on a Two-Lane Highway: A Case Study of the Dhankhola-Bhaluwang (H01) Road Section**" submitted by **Sanjay Luitel** in partial fulfillment of the requirement for degree of Master of Science in Transportation Engineering.

…………………………………………………………

Supervisor: **Dr. Pradeep Kumar Shrestha**

Assistant Professor

Institute of Engineering

…………………………………………………………

Co- Supervisor: **Hemant Tiwari**

General Secretary, Society of Transport Engineers Nepal-SOTEN

…………………………………………………………

External Examiner: **Subhash Dhungel**

Traffic Engineer and Road Safety Expert

Chairman, Traffic & Transport Unlimited Solution

…………………………………………………………

Committee Chairperson: **Anil Marsani**

Coordinator, M.Sc. in Transportation Engineering

Department of Civil Engineering

Date…………………………

ABSTRACT

Over the years, substantial efforts have been dedicated to improving travel time prediction for corridors, using a multitude of variables. However, predicting travel time in corridors remains an inherently challenging task due to the intricate interplay of numerous factors, which are often difficult to comprehensively collect. This challenge is particularly pronounced in undivided roads, where corridor access is unrestricted, leading to a heightened presence and influence of various influencing factors.

The present study is focused on the development of a travel time prediction model for the Dhankhola-Bhaluwang road section, which is a two-lane, two -way undivided highway, for both directions. Using an extensive analysis of 72-hour datasets on vehicle travel times, sourced from traffic volume counts and speed surveys, this study evaluates the effectiveness of travel time prediction models, taking into account through traffic, opposing traffic, and the proportions of heavy vehicles in through traffic. Evaluation metrics derived from Random Forest Regression consistently outperform those of other regression models for both directions. Subsequently, Support Vector Regression, Decision Tree Regression, LASSO Regression, and Multiple Linear Regression follow in effectiveness.

Moreover, the study suggests travel time functions for both light and heavy vehicles, which are derived from traffic data generated by a microscopic traffic simulation model. The experiments conducted using the VISSIM model reveal that, apart from the volume of through traffic, the composition of traffic and the volume of opposing traffic play a substantial role in determining vehicle travel times. The travel time functions for both light and heavy vehicles have been modified and proposed for a two-lane, two-way undivided carriageway road, exhibiting high accuracy and a better relationship than standard BPR functions.

Keywords: Travel Time Prediction, Travel Time Function, VISSIM, Machine Learning, Regression, Two-Lane Two-Way Undivided Carriageway Road

ACKNOWLEDGEMENT

The successful completion of this research was made possible through the unwavering motivation, guidance, and assistance of numerous individuals. I would like to express my sincere gratitude to my supervisor, Dr. Pradeep Kumar Shrestha, and my co-supervisor, Hemant Tiwari, for their continuous support and invaluable guidance throughout the entire research period. I am equally thankful to our program coordinator, Mr. Anil Marsani, and Dr. Rojee Pradhananga for their helpful suggestions and support during the study.

Moreover, I extend my appreciation to the Society of Transport Engineers Nepal-SOTEN for providing the research grant that facilitated data collection. Additionally, my heartfelt thanks go to Traffic & Transport Unlimited Solution Pvt. Ltd. for providing the necessary cameras and radar gun for various surveys. Their contributions have been instrumental in the successful execution of this research.

Finally, I want to express my gratitude to all the individuals who supported me directly and indirectly throughout the completion of this research.

Sanjay Luitel

078/MSTRE/013

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATION

AADT	Average Annual Daily Traffic
ADT	Average Daily Traffic
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
BPR	Bureau of Public Road
DNN	Deep Neural Network
DTR	Decision Tree Regression
GEH	Geoffrey E. Havers
HV	Heavy Vehicles
kmph, km/hr.	Kilometer Per Hour
LASSO	Least Absolute Shrinkage and Selection Operator
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
min	Minute
MLR	Multiple Linear Regression
MSE	Mean Square Error
PCU	Passenger Car Unit
R2 or R^2	Coefficient of Determination
RFR	Random Forest Regression
RMSE	Root Mean Square Eror
S.T.	Subjected To
SVR	Support Vector Regression
vpd	Vehicle Per Day

CHAPTER 1: INTRODUCTION

1.1 Background

Road infrastructure is a necessity that ensures accessibility and mobility and provides a path for overall economic development in the country. Transport infrastructures impact various dynamics of development hence the infrastructure is often called a backbone for the development of the country. Anticipating future demands and planning infrastructure accordingly is vital, but it also poses numerous challenges for planners. They must consider various factors, including travel times, when devising transport infrastructure plans. The travel duration along a corridor yields critical insights for transportation planning, aiding in the assessment of different transportation options' efficiency and potential system alterations. Users rightfully anticipate consistent and reliable travel times from transport infrastructure throughout its operational lifespan.

Travel time reliability, which measures the consistency of travel duration on a specific road section, is a crucial metric in highway transportation planning and management. It is typically gauged by comparing average travel time to durations during congested conditions. This metric significantly influences driver behavior and overall transportation efficiency. Having an accurate and reliable information about the duration of travel is crucial in order to recognize and evaluate any issues with the operation of highway facilities (Billings & Yang, 2006). It can also help in the estimation of travel demand, and the development of plans to improve mobility and reduce travel time for travel of people and goods. Accurate travel time prediction can also assist government agencies and transportation planners in evaluating the resilience of the road segment and guide future alternative route planning. It can also aid emergency responders and government entities in responding to natural disasters, congestion, and other emergencies by identifying critical areas that are more likely to be affected. Effective traffic management relies on the various factors that impact travel time and traffic flow. Various elements such as traffic, geometry, bottleneck section, bridges, curves, roadside environment etc. can affect the reliability of travel time on the road corridor.

The capability of correctly forecasting the future travel times of links in transportation networks is essential for numerous applications in intelligent transportation systems (ITS) (Oh, Byon, Jang, & Yeo, 2015). Traffic parameters, such as traffic volume, speed, and travel time, are vital for designing and planning road facilities. These parameters collectively impact the performance of traffic movement on the road section, making it necessary to collect data and establish relationships among them. This capability enables timely responses to dynamic traffic conditions and facilitates the implementation of effective policies, ultimately improving the travel experience for commuters and the flow of goods and services (Yang & Qian, 2019).

Over the past years, numerous studies have delved into establishing the relationship between travel time and its prediction. This prediction, along with its influencing variables, hinges on the type of roadway and its surrounding characteristics. Most research has concentrated on freeways or multilane divided highways, primarily considering traffic volume as the pivotal factor in travel time prediction. However, some researchers have used simulation techniques to evaluate the effects of heavy vehicles on other traffic streams.

Hence, this study sets out to develop a relation for predicting travel time using various models and to compare their accuracy specifically for the Dhankhola-Bhaluwang road section, a two-lane, two-way road with an undivided carriageway. Additionally, the study endeavors to construct a simulation model to assess the impact of mixed traffic on travel time functions in this particular section.

1.2 Problem Statement

Comprehending the relationship between traffic volume and travel time data is pivotal for enhancing the accuracy of travel time forecasts and for future planning (D'Angelo, Al-Deek, & C., 1999). This correlation sheds light on traffic performance across various scenarios. The fluctuation of travel time in relation to traffic volume introduces unreliability. Precisely and consistently estimation trip durations in the near future holds paramount importance for efficiently managing traffic and planning routes in Intelligent Transportation Systems. In Nepal, diverse traffic conditions pose additional challenges for both traffic engineers and planners. The coexistence of light and heavy vehicles on the road leads to a blend of traffic types. This is evident in disparities in size, speed, maneuverability, and performance capabilities. These differences in performance between these types of vehicles can have an impact on the overall flow of traffic (Afandizadeh, Nasiri, & Mirzahossein, 2021). Another factor potentially influencing travel time prediction lies in the characteristics of opposing traffic on a two-lane, two-way undivided carriageway, an area that has seen limited study in globally.

Hence, considering that most of the roads/highways in Nepal are two-lane, two-way with an undivided carriageway, this study attempts to identify measurable variables that have the potential to impact travel time. The study then applies various regression techniques to enhance the accuracy of prediction. Additionally, the research assesses the traditional travel time function and, through the use of simulation and regression techniques, modifies the travel time function to achieve better accuracy, especially for two-lane, two-way undivided carriageway roads.

1.3 Objectives of Study

The primary objective of the study is to develop travel time prediction models for Nepalese traffic conditions. The specific objectives are enlisted as below:

- To assess the relationship among through traffic volume, opposing traffic volume, percentage of heavy vehicles in through traffic in travel time prediction using various regression techniques on a two-lane, two-way undivided carriageway road section under mixed traffic condition.
- To refine travel time functions for both light and heavy vehicles, employing simulation and regression techniques.

1.4 Scope of Study

The scope of the study includes the following points:

1. **Data Collection: -** Collection of traffic volume, travel time and speed data on the road section.
    - **Traffic Volume: -** Carry out a classified directional traffic volume count survey on two locations of the road section for 3 days using a video graphics survey. The time interval of traffic volume data is 30 minutes interval.

- **Travel time: -** Collection of travel time data of vehicles excluding motorcycles and non-motorized vehicles by viewing the video footage. This exclusion is necessary due to limitations in detecting these vehicles through video footage. The data includes arrival time, departure time, and overall travel duration.
- **Speed Study: -** Carry out a speed survey in various elements of the road section such as curves, straight segments, bridges and grade sections.

2. **Various prediction models: -** Determining the relationship among travel time, through traffic volume, opposing volume and proportion of heavy vehicle using a 48-hour dataset by different models and validating the models with a 24-hour remaining unused dataset. Evaluating and comparing the performance of various models such as multiple linear regression, LASSO regression model, decision tree regression, random forest regression and support vector regression.
3. **Modify Travel Time Functions:** Develop a simulation model in VISSIM, and refine previously established travel time functions for both light and heavy vehicles using simulation results and regression techniques to enhance the accuracy of the functions.

1.5 Significance of Study

The study aims to develop travel time prediction models and assess various models for their effectiveness in predicting travel time. Additionally, the study focuses on creating a simulation model specifically designed for mixed traffic conditions. The outcomes of the study have the potential to significantly improve transportation planning efforts by reducing congestion and improving travel time. One of the key benefits of the study is its ability to predict future congestion on the road section. By developing a reliable travel time prediction model, researchers can anticipate when congestion is likely to occur in the coming years. This information is crucial for planning alternative measures, such as adding lanes or implementing traffic management strategies, to mitigate congestion and maintain smooth traffic flow. The study's simulation model allows for the visualization of future scenarios with traffic growth, facilitating proactive decision-making and infrastructure development. Understanding this relationship is valuable for traffic improvement initiatives. For example, the findings can be used to determine the most suitable times for heavy vehicles to enter the road section, thereby optimizing traffic flow and reducing congestion. This information benefits not only transportation planners but also individual companies, travelers, and suppliers who can identify the most opportune times for movement on the highway.

Moreover, the study findings and methodologies have broader applicability beyond the specific road section being studied. Similar road sections and transportation networks can benefit from the insights gained in travel time prediction and system evaluation. By sharing the knowledge and approaches developed in this study, transportation professionals can make informed decisions and implement effective measures to optimize travel time and enhance overall transportation efficiency.

1.6 Limitation of Study

The study aims to examine variables to the fullest extent, but due to limitations in resources, it cannot delve deeper into the following enlisted points.

- The road section in the study was limited to 12.9 km as it is challenging to track vehicles by video footage for longer distances. The highway being studied contains numerous access roads and settlements. The longer the length of the road, the more access roads there will be, which may negatively impact the accuracy of the results.
- The travel time of motorcycles and non-motorized vehicles was not considered due to the difficulty in accurately detecting these vehicles through video footage.
- Three-day traffic volume data were collected although some models such as decision tree regression, support vector regression, random forest regression.
- The arrival and departure times were only used to calculate travel time, so any system or personal delays were ignored, which may affect the accuracy of the travel time. Efforts are made to minimize this error by removing unreliable data
- The side friction, which could potentially influence travel time prediction, was not measured or taken into consideration during the development of regression models.
- The accuracy of the VISSIM model may be affected by the absence of an access road and its associated traffic properties. However, the model's performance was validated by comparing its results to real-time data collected in the field.

- The effect of visibility in various time period was not considered which may have impact on travel time.

CHAPTER 2: LITERATURE REVIEW

2.1 Travel Time Prediction Models

Various models and methodologies have been adopted to predict travel time over the past two decades. Some definitions of such models are described below:

2.1.1 Linear regression

Linear regression is a method used to predict the value of one variable based on the value of another. The variable being predicted is known as the dependent variable and the variable used to make the prediction is called the independent variable. The analysis involves determining the coefficients of a linear equation that best predicts the value of the dependent variable. It finds the best fit line or surface by minimizing the difference between the predicted and actual values. There are tools available that use the "least squares" method to find the best fit line for a given set of data. The goal is to estimate the value of the dependent variable based on the independent variable (Oh, Byon, Jang, & Yeo, 2015) (Zhang & Rice, 2003) (Liu, Wang, Yang, & Zhang, 2017) (Sharma, Singh, & Upteti, 2021).

2.1.2 Lasso regression

Short for Least Absolute Shrinkage and Selection Operator, is a type of linear regression that adds a regularization term to the cost function. Tibshirani in 1996 introduced the LASSO, an ingenious technique for variable selection in regression. It achieves this by minimizing the residual sum of squares while ensuring that the sum of the absolute values of the coefficients remains below a specific constant. LASSO is widely recognized as an effective sparse regression method, effectively regularizing the parameter α with a sparse assumption (Tibshirani, 1996). The regularization term is a penalty term that helps to prevent overfitting, which is a common problem in traditional linear regression. The lasso method encourages the coefficient estimates of the independent variables to be exactly equal to zero. By forcing some of the coefficients to be equal to zero, lasso regression can be used for feature selection, which can be useful when you have a large number of independent variables and only a few of them are important for predicting the outcome (Liu, Wang, Yang, & Zhang, 2017).

2.1.3 Decision Tree Regression

A decision tree, originating from the field of machine learning, serves as an efficient tool for tackling classification and regression problems. In contrast to other classification methods that collectively utilize a group of features or bands to make a single classification decision, the decision tree operates through a multi-stage or hierarchical decision process resembling a tree structure. This tree consists of a root node (containing all the data), a series of internal nodes (referred to as splits), and a set of terminal nodes (referred to as leaves). At each node within the decision tree structure, a binary decision is made, which effectively separates either one specific class or a subset of classes from the remaining ones. The process typically involves traversing down the tree until reaching a leaf node, following a top-down approach (Xu, Watanachaturaporn, Varshney, & Arora, 2005).

2.1.4 Random Forest Regression

Random Forest is a supervised learning algorithm that is based on the ensemble learning method and many Decision Trees. Random Forest is a Bagging technique, so all calculations are run in parallel and there is no interaction between the Decision Trees when building them. RF can be used to solve both Classification and Regression tasks (Liu, Wang, Yang, & Zhang, 2017) (Sharma, Singh, & Upteti, 2021).

2.1.5 Support Vector Regression

Supervised Machine Learning Models with associated learning algorithms that analyze data for classification and regression analysis are known as Support Vector Regression. SVR is built based on the concept of a Support Vector Machine or SVM. It is one of the popular Machine Learning models that can be used in classification problems or assigning classes when the data is not linearly separable (Wu, Ho, & Lee, 2004).

2.2 Related Studies on Travel Time Prediction Model

Many methods for travel time prediction have been developed over the years. It seems that many researchers spend significant time on this topic. D’Angelo et al. (1999) examined the utilization of a nonlinear time series model for forecasting traffic parameters along an 18 km stretch of freeway in Orlando, Florida. The travel time data were derived from speed data, which were collected from inductive loop detectors on the study road section. The issue was approached by considering both single-variable and multiple-variable predictions of travel times within the corridor. Single-variable prediction involved using speed time-series data to forecast travel times along the freeway corridor. For multivariable prediction, schemes were developed utilizing speed, occupancy, and volume data. The result showed that the single-variable model was superior to the multivariable prediction schemes (D'Angelo, Al-Deek, & C., 1999). Chien & Kuchipudi (2003) discussed the results and accuracy generated by different prediction models. Their study placed a significant emphasis on modeling real and historical data to predict travel times. The Kalman filtering algorithm was employed for this purpose, given its importance in consistently updating the state variable with new observations. The findings indicated that, particularly during peak hours, using historical path-based data for travel-time prediction is more effective than link-based data, attributed to the smaller travel-time variance and larger sample size associated with the former (Chien & Kuchipudi, 2003).

Gao & Zhang (2017) explored the relationship between travel time and traffic flow on the road in the static state that inflow equals outflow. The results concluded the driving time increases first and then decreases when flow on the road increases monotonously in a static state (Gao & Zhang, 2017). Billings & Yang (2006) centered on employing time series models to address the arterial time prediction challenge for a highway. The model's validation included both residual analysis and the portmanteau lack-of-fit test. The authors concluded that the method can be easily modified and applied to short-term arterial travel time prediction for urban areas (Billings & Yang, 2006). Oh et al. in 2014 carried out a comprehensive review study focusing on various kinds of literature, including modern techniques proposed recently, related to travel time and traffic condition prediction that is based on data-driven approaches. These approaches were analyzed for their strengths, potential weaknesses and performances from five main perspectives that are prediction range, accuracy, efficiency, applicability and robustness (Oh, Byon, Jang, & Yeo, 2015). Zhang & Rice (2003) 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 (Zhang & Rice, 2003). Liu et al. (2017) established a series of long short-term memory neural networks with deep neural layers (LSTM-DNN) using 16 settings of hyperparameters and investigates their performance on a 90-day travel time dataset. Then the model was tested along with linear models such as linear regression, Ridge and Lasso Regression, ARIMA and DNN models under 10 sets of sliding windows and predicting horizon via the same dataset (Liu, Wang, Yang, & Zhang, 2017).

Wu et al. (2004) applied support vector regression (SVR) for travel-time prediction and compared its results to other baseline travel-time prediction methods using real highway traffic data. The study revealed that the SVR predictor can significantly reduce both relative mean errors and root-mean-squared errors of predicted travel time (Wu, Ho, & Lee, 2004). Hongsuk Yi et al. (2017) suggested a deep learning neural network based on TensorFlow for the prediction of traffic flow conditions, using real-time traffic data. The model was trained by a deep learning algorithm, which uses real traffic data aggregated every five minutes. Results demonstrated that the model's accuracy rate is around 99% (Yi, Jung, & Bae, 2017). Sharma et al. (2021) proposed a technique in which every cab service provider can give exact trip duration to their customers taking into consideration the factors such as traffic, time and day of pickup. The authors used Random Forest Regressor and Linear Regression to predict the travel time. The results revealed that the Random Forest Regressor is more accurate than Linear Regression (Sharma, Singh, & Upteti, 2021). Fan & Gurmu (2015) used historical average, Kalman filtering and artificial neural network to develop and compare dynamic travel time prediction model. Results revealed that the ANN outperformed the other two models in both aspects: overall prediction accuracy and robustness (Fan & Gurmu, 2015). Murni et al. (2020) proposed the k-Nearest Neighbors Regression method with Tensorflow to construct an estimation model. The results revealed that the proposed model gives the Mean Absolute Error of 2.196078, a Root Mean Square Error of 2.977036294 and accuracy rate 88.1819% (Murni, et al., 2020). The summary of literature review related on travel time prediction models is shown in Table 2-1.

Table 2-1: Summary of Literature Review

Authors	Model Type	Variable	Accuracy	Link Type
(Billings & Yang, 2006)	Arima Model	Travel time, volume	-	Urban
(D'Angelo, Al-Deek, & C., 1999)	Nonlinear time series model	Travel time, volume, speed, occupancy	-	Freeway
(Zhang & Rice, 2003)	Linear Regression	Travel time, volume	Error= 8% at zero lag Error= 13 % at a time lag of 30 min	Freeway
(Liu, Wang, Yang, & Zhang, 2017)	LSTM-DNN (Deep Learning) Linear regression Ridge regression Lasso Regression ARIMA DNN	Travel time, volume	-	Freeway
(Sharma, Singh, & Upteti, 2021)	Random Forest Regressor and Linear Regression	Travel time, Volume	RFR=83 % LR=78%	Freeway
(Fan & Gurmu, 2015)	Kalman Filter Algorithm Artificial Neural Network	Travel time and Volume		Freeway
(Wu, Ho, & Lee, 2004)	Support Vector Regression	Travel time and traffic volume	Error <= 5%	Freeway
(Yi, Jung, & Bae, 2017)	Deep neural network	Travel time and traffic volume	99%	Freeway
(Murni, et al., 2020)	K-Nearest Neighbors Regression method with Tensorflow	Travel time and volume	88.18%	Highway
(Chien & Kuchipudi, 2003)	Kalman filtering algorithm	Travel time and volume		Freeway

2.3 Effect of Large Vehicle on Performance of Traffic Stream

Numerous studies have explored the impact of heavy vehicles on traffic system performance, with a predominant focus on simulation-based modeling and analysis. Afandizadeh et al. (2021) investigated the impact of heavy vehicles (HV) on traffic flow characteristics on an undivided two-lane two-way highway in Mazandaran, Iran. The simulation, conducted using AIMSUM software, indicated that a rise in heavy vehicle rates led to a reduction in the average speed of traffic flow by up to 15% at a

40% HV rate. Additionally, an increase in the number of lane changes was observed, with a rise of about 20% corresponding to a 24% increase in the presence of heavy vehicles. This model showed the effect of the proportion of heavy vehicles on travel time, average speed, delay time and lane changing number (Afandizadeh, Nasiri, & Mirzahossein, 2021). Roh et al. (2017) investigated the influence of heavy vehicles (HVs) on traffic flows through the analysis of real-time AVC data. The study aimed to assess the relationship among average speed, HV ratio, flow rate, and the number of lanes. Results revealed a decrease in average speed as both flow rate and HV ratio increased for six-lane, eight-lane highways, and the four-lane highway. (Roh, Park, & Kim, 2017). Gao et al. (2020) predicted the impact of large-scale vehicles on the average speed of flow, vehicle speeds under different vehicle mixing rates for three groups of v/c ratio (Gao, Xu, Jia, & Dong, 2020). Lu et al. (2016) developed a microscopic simulation model and proposed continuous function for each vehicle type and its parameter using the traffic data generated by simulation model. The study indicated that (i) in addition to traffic volume, traffic composition has significant influence on travel time of vehicle and (ii) the respective estimations for travel time of heterogenous flows could greatly improve their estimation accuracy (Lu, Meng, & Gomes, 2016).

2.4 Travel Time Function

Travel time functions delineate the correlation between the travel time along a road and the traffic volume on that road (Rose, Taylor, & Tisato, 1989). This definition serves as the foundation for numerous research endeavors, where researchers often adapt and customize the function to accommodate diverse situations, aiming to identify the most accurate and effective function.

1. In 1964, the Bureau of Public Road in the United States established an empirical correlation called the BPR (Bureau of Public Road) latency function shown in Eq. 2.1 (Manual, 1964).

$$f(x) = t\left(1 + \alpha\left(\frac{x}{C}\right)^{\beta}\right) \qquad \textbf{2.1}$$

Where, t= indicates the trip time on an unoccupied route,

x= traffic flow in PCU

C= specifies the capacity of a route, and the constants

α, β= route-specific features which may affect the impact of flow to capacity ration on travel time.

The α and β values are evaluated based on statistical data from a route. Generally, β takes a value from 1 to 4.

2. Wu et al. (2006) developed an empirical function of a single link travel time function for all the vehicle type using the BPR- type function shown in Eq. 2.2 (Wu, Florian, & He, 2006).

$$t = t^0 \left(1 + 0.15 \left(\frac{x}{c}\right)^4\right) \qquad \textbf{2.2}$$

Where, t is the average travel time of the study road segment, t^0 denotes free flow travel time, x is the traffic flow in PCU and C is road capacity in PCU.

3. Lu et al. (2016) estimated the component travel time function for heterogeneous traffic flows on freeway and developed a microscopic traffic simulation based four-step model. The study purposed individual function for each vehicle type and its parameters are estimated using traffic data generated by a microscopic traffic simulation model is shown in Eq. (2.3), Eq. (2.4), Eq. (2.5) and Eq. (2.6) (Lu, Meng, & Gomes, 2016).

For cars,
$$t_{a,0} = t^0_{a,0} * \left(1 + 0.29(1 + \rho_1)^{2.62} \times \left(\frac{Q}{y}\right)^{1.97}\right) \qquad \rho_0 > 60\% \qquad \textbf{2.3}$$

$$t^0_{a,0} * \left(1 + 0.62 \left(\frac{Q}{y}\right)^{1.26}\right) \qquad \rho_0 < 60\% \qquad \textbf{2.4}$$

For heavy trucks $$t_{a,1} = t^0_{a,1} * \left(1 + 0.12 \times \left(\frac{Q}{y}\right)^{1.87}\right)$$ ρ0 >60% **2.5**

$$t^0_{a,1} * \left(1 + 0.10 \times \left(\frac{Q}{y}\right)^{1.26}\right)$$ ρ0 <60% **2.6**

Where, $t_{a,0}$= Average link travel time of car

$t_{a,1}$= Average link travel time of heavy trucks

$t^0_{a,0}$= Free flow speed of car

$t^0_{a,1}$= Free flow speed of heavy truck

ρ_0 = Percentage of car traffic

ρ_1 = Percentage of heavy truck

Q and y are the volume and capacity of link in PCU respectively.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Study Area

This study is focused on the Dhankhola-Bhaluwang Road segment, which forms part of the East-West highway (H01) and spans a distance of 12.9 kilometers. The road section is situated in the Dang and Kapilvastu districts of Lumbini Province. The road corridor stretches from the eastern point in Dhankhola (at 82°48'7.36"E, 27°47'3.90"N) to the western end in Bhaluwang (at 82°44'52.76"E, 27°50'19.05"N), running in an east-west direction. The road comprises a two-lane, two-way undivided carriageway, surfaced with blacktop but lacks a shoulder and footpath. Along the corridor, there is visible ribbon development, with small hotels and local market activities spreading throughout. Major settlement areas along the route include Dhankhola, Kalakate, and Bhaluwang. The road experiences mixed traffic conditions, and its Annual Average Daily Traffic (AADT) was recorded as 8076 and 6886 (excluding motorcycles and rickshaws) during the 2021/2022 period (Highway Management Information System (HMIS), n.d.). The road section is situated in rolling terrain, and there are various curves and crossings along the way, as indicated in Figure 3-1.

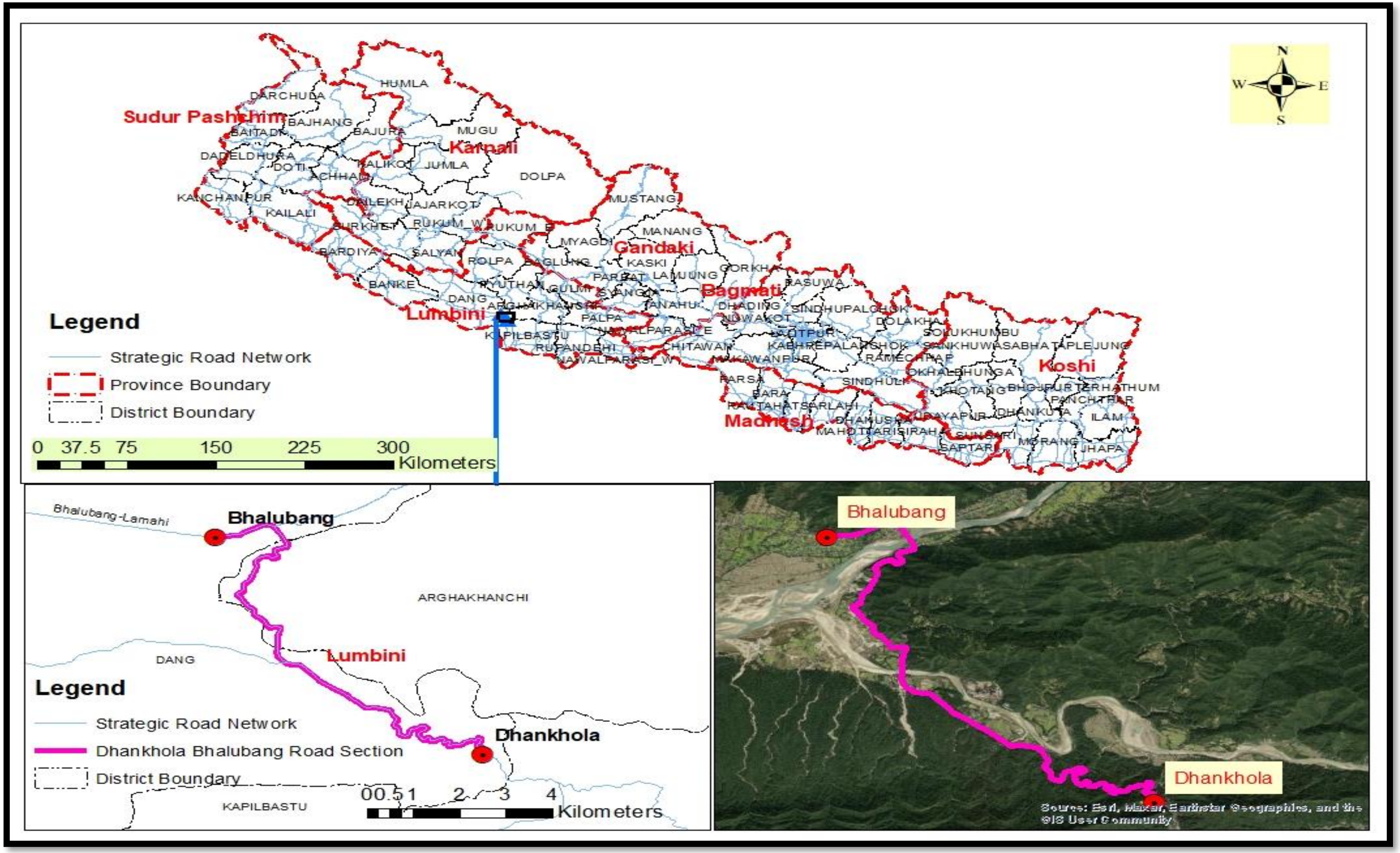


Figure 3-1: Study Area

3.2 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 is shown in Figure 3-2.

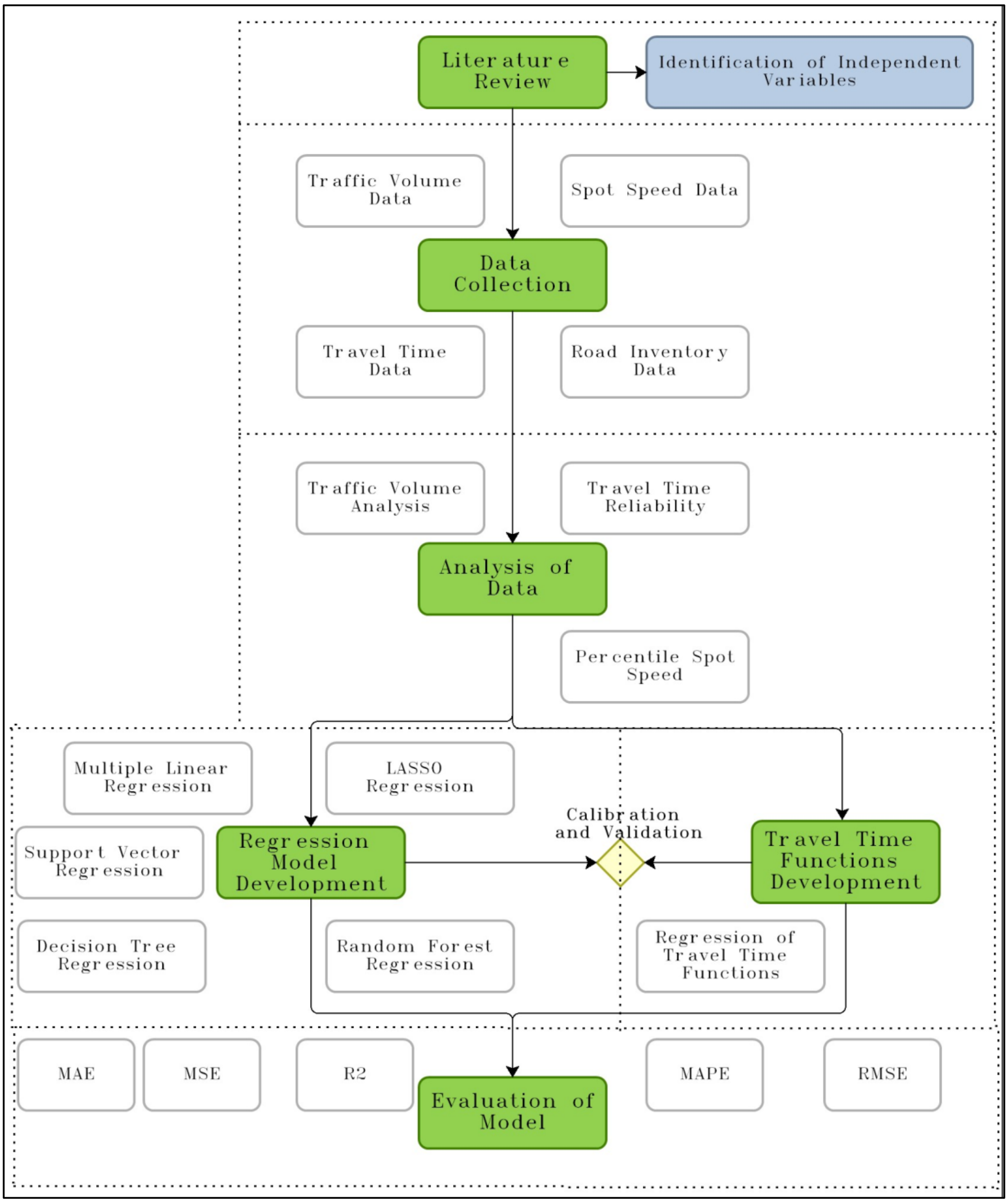


Figure 3-2: Methodological framework of the study

3.2.1 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 objective of this desk study was to identify potential measurable variables that could be associated with the prediction of travel time. For this study, three independent variables were selected to predict travel time.

The first independent variable utilized in the study was through traffic volume, as it was observed that the variability in traffic time is influenced by fluctuations in both demand and capacity (Tu, Van Lint, & Van Zuylen, 2007). This variable plays a crucial role in understanding how changes in traffic volume affect travel time prediction. The second independent variable considered in the research was opposing traffic volume. The act of overtaking on undivided roads, where vehicles utilize the opposing lane to pass slower vehicles, presents a complex and significant maneuver, especially in the presence of oncoming vehicles from the opposite direction (Asaithambi & Shravani, 2017). This variable was deemed essential in comprehending how opposing traffic impacts travel time estimation. The third independent variable focused on the percentages of heavy vehicles in through traffic. These vehicles exhibit distinct lagging-leading behavior in relation to other vehicle types due to the presence of heavy vehicles on the road (Ahmed, Drakopoulos, & Ng, 2013). By considering these three independent variables, the research sought to enhance the accuracy and comprehensiveness of travel time prediction models.

Another objective of the desk study was to ascertain the necessary data for developing and calibrating a simulation model in VISSIM. Vital input data for the development of the simulation model for the road corridor included traffic parameters such as demand (volume), vehicle characteristics, speed distribution dataset, and geometric features of the corridor. These factors were crucial in constructing an accurate and reliable simulation model in VISSIM that could effectively represent the real-world conditions and dynamics of the road corridor (L. & Turochy, 2019).

In addition, standard data sheets were prepared for the collection of traffic volume data, travel time data and speed data.

3.2.2 Data Collection

Various traffic surveys were performed to obtain primary data for the study.

- **Traffic Volume Count**

The videographic recording survey was used to record the video footage of traffic volume to collect the traffic volume data. The movement of vehicles in and out of corridors was recorded for three days from 12th June, Monday to 14th June, Wednesday, 2023 using cameras. Two cameras, one at the entry and the other at the exit, were installed to capture the directional flow of traffic. The typical layout of the position of cameras for recording video footage is shown in Figure 3-3 and the photographs taken during installation of camera is shown in Figure 3-4.

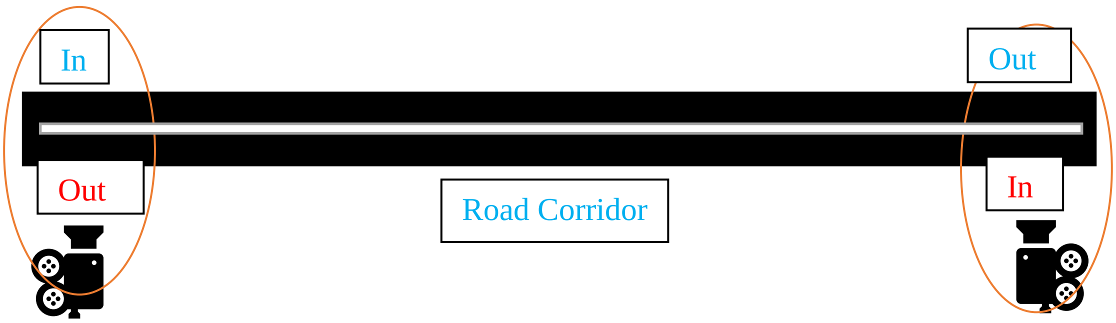


Figure 3-3: Typical layout for camera's position for recording video footages

Figure 3-4: Installation of camera at sites

A manual method was used to count the traffic volume data by viewing the recorded video. Two enumerators had been used in traffic volume count. 72-hour classified and directional traffic volume data were extracted at the interval of 30 minutes. Only motorized traffic was taken into consideration in traffic volume count. Figure 3-5 and

Table 3-1 shows the types and classification used in both counting stations. The classification adopted is consistent with Department of road practice for traffic volume count (Highway Management Information System (HMIS), n.d.).

Table 3-1: Classification of motorized vehicles

Vehicle Type	Vehicle Characteristics
Multi-Axle Truck	Standard / heavy trucks, trailers/articulated. (≥3 axles)
Heavy Truck	Standard / heavy trucks, trailers/articulated. (2 axles)
Mini Truck	Mid-sized trucks with single rear-axle (usually 4-wheeled, <8 tons GVW)
Big Bus	Buses having seating capacity of 35-50 seats
Mini-bus	Medium size buses having seating capacity of 20-35 seats.
Micro-bus	Small buses and vans having seating capacity of 10-15 seats.
Car	Passenger car taxis and vans (≤ 5 seats).
Utility Vehicles	Pickups or 4-wheeled vehicles with single/twin cabin and load compartment (open/hooded), Light freight vehicles
Tractor	Farm tractors
Four Wheel Drive	Vehicles strictly having four-wheel gears (seating approximately 10) such as Mitsubishi Pajero, Prado etc.
Three-Wheeler	Electrical or gasoline/LPG fueled 3-wheeled vehicles (excluding power tillers, farm tractors)
Power Tiller	Motorized four-wheel vehicles used for carrying goods and mainly driven by hands and not steering.
Motorcycle	Motorized two wheelers such as scooters and motorcycles

				
Multi-Axle Truck	Heavy Truck	Mini Truck	Standard Bus	Mini-bus

				
Micro-bus	Car	Utility Vehicle	Tractor	4 Wheel Drive
				
Three-Wheeler	Power Tiller	Motorcycle		

Figure 3-5: Vehicle classification adopted for traffic volume count

- **Speed Survey**

The spot speed for different elements of the road was measured by using a radar gun. The spot speed survey was conducted on a total of 17 locations, which consist of 8 straight segments, four curves, two bridge approaches, and two intersection approaches. To conduct the speed survey, two individuals were assigned: one operated the radar gun, while the other recorded the radar gun readings on the datasheet. In each segment, data was collected for a duration of 1 hour or until 100 vehicles (excluding motorcycles) had been recorded in each direction. To ensure unbiased results, the surveyors took measurements discreetly. After a vehicle passed the surveyor's location, the radar gun was used to capture the speed by targeting the rear end of the vehicle, thereby ensuring that the driver remained unaware of the ongoing survey. Spot speed data were manually recorded on a prearranged Excel sheet, with each section of the road element properly segmented. The unit of measurement used for spot speed was kilometers per hour (kmph). The photographs taken during the fieldwork of speed data collection is shown in Figure 3-6 and Figure 3-7 shows the locations where speed survey was conducted.

Figure 3-6: Spot speed survey using radar gun

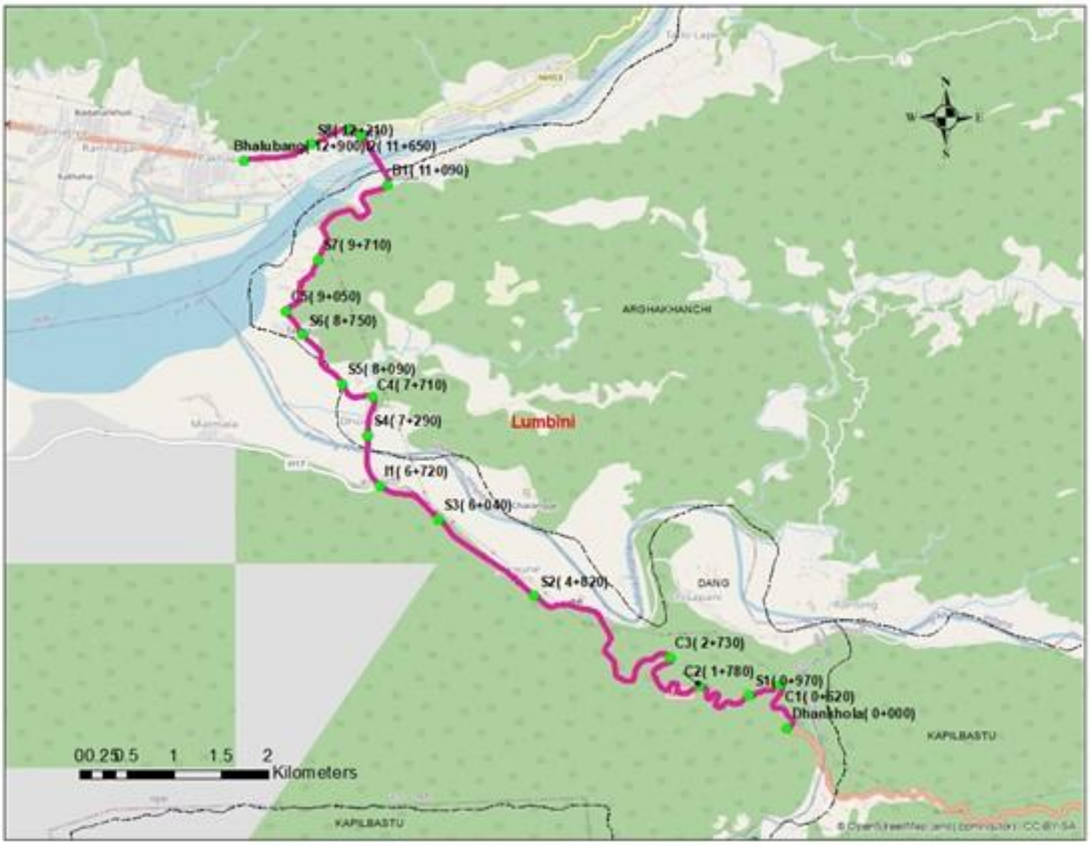


Figure 3-7: Locations for spot speed survey

- **Road Inventory Survey**

Road inventory and features of the roadside environment were collected during the fieldworks. Bridges, curves, settlement were recorded along with geometric parameter of road corridor.

- **Travel Time Determination**

Based on video footage of in and out of the corridor, arrival and departure times of vehicles (excluding motorcycle and non-motorized vehicles) for both directions were determined by viewing and identifying the vehicle. Afterward, travel time was computed by using the arithmetic operation of arrival and departure time.

$$Travel\ Time = Departure\ Time - Arrival\ Time$$

A total of 2,882 samples of travel time for the Dhankhola-Bhaluwang direction and 2,887 samples for the Bhaluwang-Dhankhola direction were collected. These travel time data points were then averaged at 30-minute intervals.

3.2.3 Analysis of Data

Collected data were further processed to develop various prediction models and simulation model.

- **Traffic Volume Analysis**

The traffic volume data is presented and processed for additional analysis using two measures: Passenger Car Unit (PCU) and Average Daily Traffic (ADT). To accurately assess the impact of each classified vehicle in mixed traffic conditions, the counts of various vehicle types are transformed into a standardized PCU unit. This is achieved by multiplying the total number of each vehicle type by its corresponding PCU factor. The PCUs assigned to different types of vehicles, used for subsequent analysis, shown in Table 3-2.

Table 3-2: Equivalency factors in terms of PCU

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

SN	Vehicle Type	PCU Equivalency Factor
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
Source: (Highway Management Information System (HMIS), n.d.)		

Moreover, calculations were performed for both directions to determine the through traffic volume, opposing traffic volume, and proportion of heavy slow-moving vehicles. For hourly traffic volume, 30 min interval traffic volume was multiplied by two. Ultimately, this resulted in 144 datasets (three-day datasets), each containing four columns: travel time (minutes), through traffic volume (PCU/hr.), percentage of heavy vehicles in through traffic, and opposing traffic volume (PCU/hr.).

- **Speed Data Analysis**

To analyze speed data, 98th/ 85th/ 50th and 15th percentile speeds of classified vehicle were calculated using ranking method (Hou, Carlos, & Edara, 2012). Ranking method was used to compute the 98th/ 85th/ 50th and 15th percentile speed. The measured speed data were arranged in descending order and ranked them. The least measured speed was ranked 1st and the highest speed was ranked nth. Then the ranked number for individually measured speed was divided by n value and multiplied by 100 to obtain the percentile. The measured speed which has 85th percent was termed as 85th percentile speed similarly for others.

- **Travel Time Reliability**

Travel time reliability in terms of 90th or 95th percentile travel times was calculated for both directions (Administration, 2010). This result was not valued for further model development, but knowing the reliability helps in identifying categorical variation of travel time in mixed traffic conditions

3.2.4 Regression Model Development

Various methods and analyses were done to develop a travel time prediction model using regression techniques. 48-hours dataset was used to train the model and the remaining 24-hours dataset were used to validate the model. A distinct model was developed for each direction. Additionally, K-fold cross validation and hyper tuning were employed in machine learning regression to fine-tune the model's hyperparameters and assess its performance across various subsets of the data, ensuring robustness and accuracy in predictions.

- **Multiple Linear Regression**

The multi linear regression models were fitted to the data. Using least squared regression, the best-fitted line was determined using Excel software. In multiple linear regression analysis, the goal is to simultaneously address the variation of independent variables in relation to dependent variables. The formulation of the multivariate regression analysis model is represented as shown in Eq 3.1.

$$y = \beta + \sum_{i=1}^{n} \beta i x_i + \epsilon \qquad \textbf{3.1}$$

In this context, 'y' represents the dependent variable, 'x' represents the independent variables, 'β' represents the constant parameter, and ' ϵ ' represents the error.

- **Lasso Regression**

The lasso regression model was fitted to the data using a gradient descent algorithm in Python using the scikit-learn library. LASSO, introduced by Tibshirani in 1996 (Tibshirani, 1996), is a pioneering technique for variable selection in regression. It achieves this by minimizing the residual sum of squares while ensuring that the absolute values of the coefficients are below a certain constant. It is a well-known method for sparse regression, effectively regulating the parameter β under sparse assumptions. Originally designed in the context of least squares, its fundamental framework can be summarized as follows:

Given a sample with N cases, each comprised of p covariates and a single outcome, where y_i represents the response variable and $x_i = (x_{i1}, x_{i2}, ..., x_{ip})^T$ is the covariate vector for the i^{th} case, and $\beta = (\beta_1, \beta_2, ..., \beta_p)^T$, the objective of LASSO is to solve the following optimization problem:

$$arg \min_{\beta_0, \beta \in \Re p} \frac{1}{N} \sum_{i=1}^{N} (y_i - \beta_0 - x^T \beta)^2 \quad \textbf{3.2}$$

s.t. $\sum_{j=1}^{p} |\beta_j| \leq t$

The parameter $t \geq 0$ serves as a predefined tuning parameter that governs the degree of regularization. When t assumes a large value, nearly all coefficients tend towards zero. Conversely, for smaller t values, the LASSO method reduces some estimated coefficients to zero. Let's denote X as the N×p matrix of covariates, where N is the sample size and p are the number of covariates. Additionally, y represents the response vector, serving as the expected output. This can lead to a more concise representation of Eq. 3.2.

$$arg \min_{\beta_0, \beta \in \Re p} \frac{1}{N} ||y - \beta_0 I - X\beta^T||_2^2 \quad \textbf{3.3}$$

s.t. $||\beta||_1 \leq t$

where $||Z||_p = \left(\sum_{i=1}^{N} |Z_i|^p\right)^{1/p}$ is the standard l^p norm. Since $\hat{\beta} = \bar{y} - \bar{x}^T \beta$, so that

$$y_i - \bar{\beta}_0 - x_i^T \beta = y_i - (\bar{y} - \bar{x}^T \beta) - \bar{x}^T \beta \quad \textbf{3.4}$$

$$= (y_i - \bar{y}) - (x_i - \bar{x})^T \beta,$$

We can write Eq. 3.3 as

$$\text{arg} \min_{\beta_0, \beta \in \Re p} \left\{ \frac{1}{N} ||y - X\beta^T||_2^2 \right\} \quad \textbf{3.5}$$

s.t. $||\beta||_1 \leq t$

The LASSO estimator $\hat{\beta}$ can be expressed in the Lagrangian form as Eq. 3.6.

$$\mathrm{L}(\beta,\lambda) = \min_{\beta\in\Re p}\left\{\frac{1}{N}||y - X\beta^T||_2^2 + \lambda||\beta||_1\right\} \quad \textbf{3.6}$$

Here, the tuning parameter λ≥0 plays a crucial role in striking a balance between the empirical error and the sparsity of model parameters. It's worth noting that the precise relationship between λ is contingent on the specific data at hand.

In practice, the combination of k-fold cross-validation with the LARS algorithm, referred to as LassoLarsCV, is commonly employed to determine the optimal regularization parameter (λ) for the Lasso regression model. In this particular investigation, we opt for a 5-fold cross-validation approach to estimate λ or t. The constant parameter λ is determined by minimizing the loss function defined in Eq. 3.6, wherein the regression coefficient vector β is iteratively estimated using the LARS algorithm, progressively reducing the residual error until it reaches a sufficiently small threshold or falls below a predefined constant.

During k-fold cross-validation, the training dataset is divided into approximately equal-sized subsets. Then, k training iterations are performed, with each iteration using a different subset for validation. The average test error across these k iterations serves as the evaluation metric for the regression model trained on $\mathrm{N}-\frac{N}{K}$ samples. Consequently, a range of λ values is obtained, and from this range, the optimal λ value is selected, corresponding to the lowest estimated generalization error. Additionally, hyperparameter tuning algorithm is employed to further optimize the performance of the model.

- **Decision Tree Regression**

The necessary libraries such as numpy, matplotlib and sklearn were imported in python. The regression dataset was loaded and preprocessed and separated it into dependent variable and independent variable. The feature such as StandardScaler from sklearn.preprocessing was used to normalized the data. Lastly, the Decision Tree Regression model was developed from DecisionTreeRegressor, specifying the kernel and hyperparameters. The model was trained firstly and then evaluation metrics were evaluated using test data. The basic framework of decision tree regression can be summarized as below:

Decision tree regression is a variation of the decision tree classifier, tailored for approximating real-valued functions like class proportions. Its construction, much like the classifier, relies on binary recursive partitioning, an iterative process that divides the data into partitions. Initially, all training samples contribute to establishing the tree's structure. The algorithm then dissects the data using every conceivable binary split and opts for the one that divides the data into two segments, minimizing the sum of squared deviations from the mean in each segment. This splitting process is then recursively applied to the new branches. The cycle persists until each node reaches a user-defined minimum node size, denoting the number of training samples at the node, and becomes a terminal node.

Because the tree is built from training samples, it can potentially suffer from overfitting when the full structure is reached. This can harm the regression accuracy of the tree when applied to unseen data, potentially limiting its generalization capability. Consequently, a pruning process is often implemented, utilizing a validation dataset and a user-specified cost complexity factor. Pruning aims to minimize the sum of the output variable variance in the validation data, considering both the number of terminal nodes and the cost complexity factor. This factor represents the complexity cost per node. In the pruning process, nodes are removed starting from the last grown node and proceeding in reverse order. The schematic representation of decision tree regression is shown in Figure 3-8.

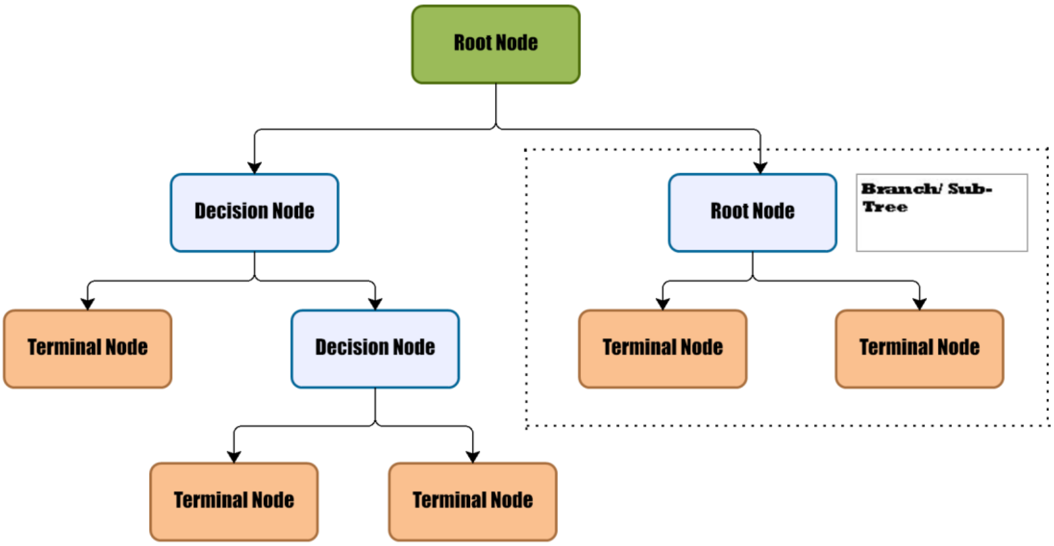


Figure 3-8: Schematical representation of Decision Tree Regression

- **Random Forest Regression**

The study made use of computer vision and deep learning libraries that are based on Python programming language. To implement and test the project, tool like Python and libraries such as Numpy, Matplotlib, Scipy, Pandas, SKlearn, Pickle, and Svm were utilized. The training process for the Deep Q-learning technique will be carried out using NVIDIA GPUs. The basic framework of random forest regression can be summarized as below:

Random Forest (RF) is a regression method that leverages the collective power of multiple Decision Tree (DT) algorithms for classification or regression tasks regarding a variable (Breiman, 2001). When provided with an input vector (x), consisting of the analyzed values of distinct evidential features for a specific training area, RF constructs a set of K regression trees and computes the average outcome. Once these K trees, denoted as $\{T(x)\}_1^K$to K, are developed, the resulting RF regression predictor can be expressed as follows:

$$f_{rf}^{K}(x) = \frac{1}{K}\sum_{K=1}^{K} T(x) \qquad \textbf{3.7}$$

Random Forest (RF) employs a technique called bagging to enhance the diversity of trees and mitigate correlation among them. Bagging involves creating training data by randomly resampling the original dataset with replacement. This means that some data points may be used more than once, while others may not be used at all, resulting in increased stability and robustness against slight variations in input data, ultimately leading to higher prediction accuracy (Breiman, 2001) When growing a tree in RF, the algorithm selects the best feature and split point from a randomly chosen subset of evidential features, reducing the correlation between trees and subsequently lowering the generalization error (Breiman, 2001). Notably, RF trees grow without pruning, making them computationally lightweight.

Additionally, samples that are not selected for the training of a specific tree in the bagging process form an out-of-bag (oob) subset. These oob elements can be used by the tree to evaluate performance, enabling RF to compute an unbiased estimation of generalization error without requiring an external test data subset (Breiman, 2001). As the number of trees increases, the generalization error converges, preventing overfitting. RF also provides insights into the relative importance of different evidential features. This is particularly valuable in high-dimensional data scenarios, allowing for the selection of the most influential features in multi-source studies. To assess feature importance, RF systematically varies one input evidential feature while keeping the others constant, measuring the resulting decrease in accuracy through oob error estimation (Breiman, 2001) The schematic representation of decision tree regression is shown in Figure 3-9.

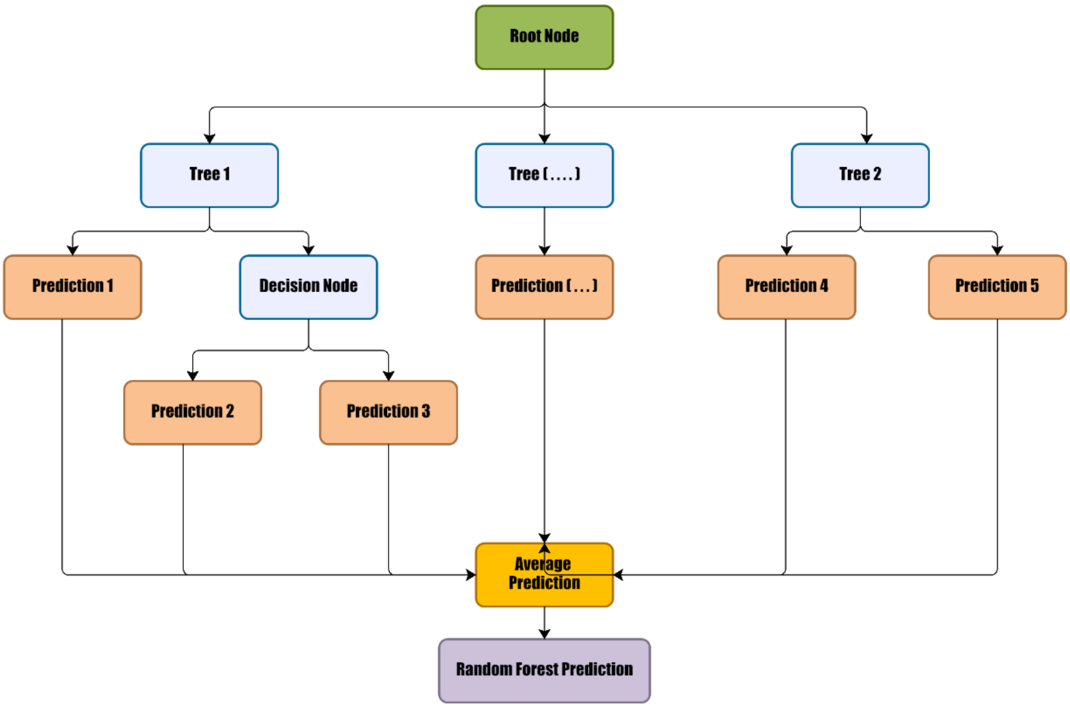


Figure 3-9: Schematical representation of Random Forest Regression

- **Support Vector Regression**

The necessary libraries such as numpy, matplotlib and sklearn were imported in python. The regression dataset was loaded and preprocessed and separated it into dependent variable and independent variable. The feature such as StandardScaler from sklearn.preprocessing was used to normalized the data. Lastly, the SVR model was developed from sklearn.svm, specifying the kernel and hyperparameters. The model was trained firstly and then evaluation metrics were evaluated using test data. The fundamental framework of support vector regression can be summarized as below:

Support Vector Regression (SVR) is a widely utilized method for both linear and non-linear regression tasks, known for its effectiveness in prediction and curve fitting. SVR is rooted in the principles of Support Vector Machine (SVM), where support vectors are data points that lie closer to the generated hyperplane within an n-dimensional feature space, distinctly segregating the data points around it. The SVR model conducts the fitting process illustrated in Figure 3-10. The generalized equation for the hyperplane is given by $y=wX+b$, where w represents the weights and b is the intercept when $X=0$. The margin of tolerance is denoted by epsilon (ϵ).

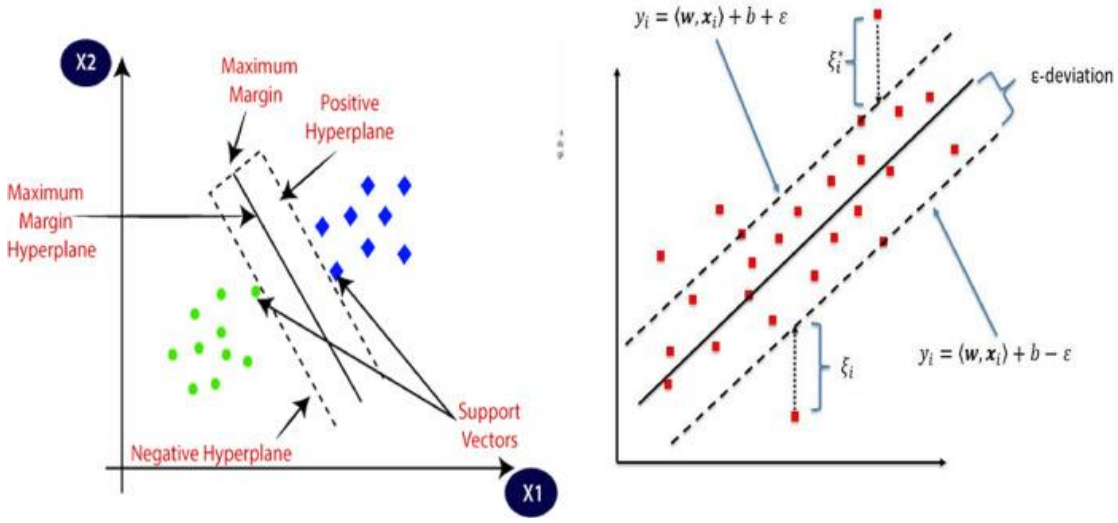


Figure 3-10: Support vector regression model for linear regression fitting where X1= X and X2 = y are the features and label in our case.

Image Source: (Kleynhans, Montanaro, Gerace, & Kanan, 2017)

3.2.5 Travel Time Function Development

The process of development of travel time function is shown in Figure 3-11.

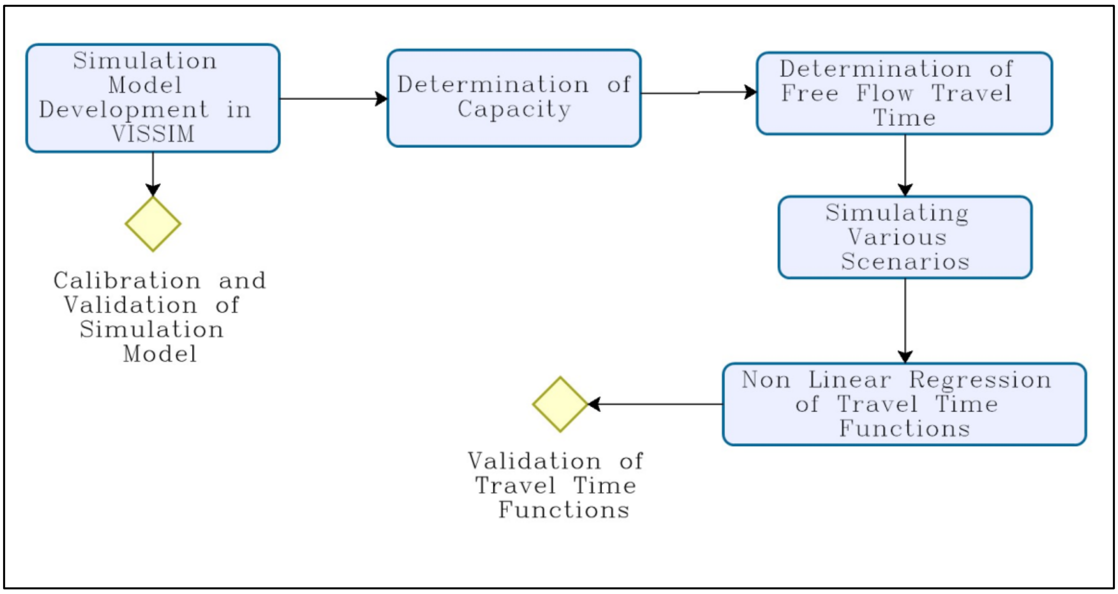


Figure 3-11: Process of determining Travel Time Functions

A simulation model was developed in VISSIM software using geometry data collected from the field and vehicle characteristics data. The classified vehicles are now grouped into two categories: light vehicles and heavy vehicles. Light vehicles consist mainly of a passenger car, four-wheel drive, microbuses, and utility vehicles, whereas a heavy vehicle includes buses and trucks. The model was calibrated by adjusting vehicle performance and driving behavior parameters (S.M.P. & Ramadurai, 2013). The model's validation with real-world scenarios was conducted using the Mean Absolute Percentage Error (MAPE) for travel time. This validation process involved using travel time data spanning a three-day period from 10 PM to 6 AM, resulting in a total of 48 datasets.

After calibration and validation, capacity was determined by simulating light vehicles and recording average speeds with respect to varying traffic volumes during a one-hour simulation period. Then, a graph is plotted based on simulation results of volume and speed, and the point where the volume reaches its maximum with low traffic speed on the graph is noted as capacity. Similarly, free flow travel time for light and heavy vehicles was also determined by simulating each type of vehicle under free flow conditions (low traffic volume) multiple times and averaging the results to determine the free flow travel time for both light and heavy vehicles.

A matrix of simulations was created, encompassing 1442 distinct scenarios. Each scenario was defined by varying percentages of heavy vehicles in through traffic, through traffic volume, and opposing traffic volume. These individual scenarios were then simulated using a VISSIM model that had been previously calibrated and validated. In each simulation, the travel times for both light and heavy vehicles were recorded. As a result, a matrix of dimensions 1442x5 was compiled, with each row representing a specific scenario. The five columns of each row denoted the travel time of light vehicles, travel time of heavy vehicles, the percentage of heavy vehicles in through traffic, through traffic volume, and opposing traffic volume.

The previously developed travel time functions mentioned in Section 2.4 were modified by incorporating opposing traffic volume in such a way that the curve properties of travel time functions would not be changed. Then, nonlinear regression analysis was performed to estimate the parameters of travel time functions with the simulated datasets for both light vehicles and heavy vehicles in Python. For heavy vehicles, two cases were performed—one that does not consider the percentage of heavy vehicles in the travel time function of heavy vehicles, and the other considers this effect. The travel time functions were then validated using simulated data, consisting of a dataset that was not used in the nonlinear regression during the calibration of parameters for the travel time functions.

3.2.6 Evaluation of Model

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. (3.8), Eq. (3.9), Eq. (3.10), Eq. (3.11), Eq. (3.12) and Eq. (3.13) respectively.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|OV_i - PV_i| \quad \textbf{3.8}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(OV_i - PV_i)^2 \quad \textbf{3.9}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{(OV_i - PV_i)^2}{OV_i} \quad \textbf{3.10}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(OV_i - PV_i)^2}{n}} \quad \textbf{3.11}$$

$$R = \frac{\sum_{i=1}^{n}(OV_i - OV_{mean})(PV_i - PV_{mean})}{\sqrt{\sum_{i=1}^{n}(OV_i - OV_{mean})^2 \ \sum_{i=1}^{n}(PV_i - PV_{mean})^2}} \quad \textbf{3.12}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(OV_i - PV_i)^2}{\sum_{i=1}^{n}(OV_i - PV_{mean})^2} \quad \textbf{3.13}$$

Where, OV_i and PV_i represent the observed and predicted values of the variable, respectively. “n” is the total number of data points and OV_{mean} and PV_{mean} refer to the average model output value representing the observed and predicted values, respectively.

CHAPTER 4: TRAFFIC DATA ANALYSIS

4.1 Traffic Volume Analysis

Directional classified traffic volume data in 30 min interval for three days at Dhankhola and Bhaluwang sections is shown in APPENDIX- A

4.1.1 Traffic Volume at Bhaluwang

Average Daily Traffic (ADT) of Bhaluwang section is found to be 5637 vehicle per day. The ADT with excluding motorcycle is 3029 vehicle per day. Figure 4-1 shows the percentage composition of different vehicle types at Bhaluwang section in averaged 24 hour of period. Motorcycles have the highest proportion, accounting for 46.27% of the total vehicles. Other significant categories include heavy trucks (10.75%), four-wheel-drive vehicles (11.96%), and cars (9.53%). Various other vehicle types, such as buses, utility vehicles, and tractors, make up smaller proportions of the total vehicle composition.

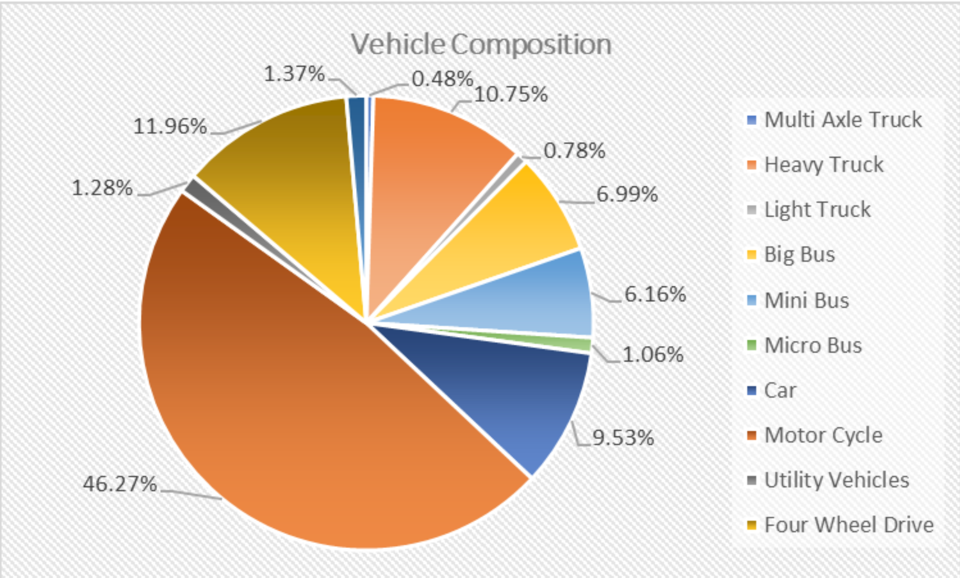


Figure 4-1: Vehicle composition at Bhaluwang

Table 4-1 shows the summary of directional classified 72-hour traffic volume count at Bhaluwang road section. It provides a comprehensive overview of traffic data for two directions: Dhankhola to Lamahi and Lamahi to Dhankhola at Bhaluwang. It shows the three-day grand total, average daily traffic (ADT), and ADT in Passenger Car Units (PCU) for various vehicle types.

Table 4-1: Summary of directional classified 72-hour Traffic Volume Count at Bhaluwang

	Dhankhola to Lamahi			Lamahi to Dhankhola		
Vehicle Type	**3 Day Grand Total**	**Average Daily Traffic**	**ADT in PCU**	**3 Day Grand Total**	**Average Daily Traffic**	**ADT in PCU**
Multi Axle Truck	36	12	48	45	15	60
Heavy Truck	912	304	912	907	302	906
Light Truck	67	22	33	67	22	33
Big Bus	608	202	606	578	192	576
Mini Bus	518	172	430	527	175	438
Micro Bus	87	29	44	94	31	47
Car	834	278	278	779	259	259
Motor Cycle	3913	1304	652	3914	1304	652
Utility Vehicles	106	35	35	111	37	37
Four Wheel Drive	998	332	332	1028	342	342
Tractor	104	34	51	131	43	65
Motorized Three-Wheeler	283	94	71	293	97	73
Total	8466	2818	3491	8474	2819	3487

4.1.2 Traffic Volume at Dhankhola

Average Daily Traffic (ADT) of Bhaluwang section is found to be 3170 vehicle per day. The ADT with excluding motorcycle is 2317 vehicle per day. Figure 4-2 shows the percentage composition of different vehicle types at Bhaluwang section in averaged 24 hour of period. Motorcycles have the highest representation, accounting for 26.91% of the total vehicles. Other significant categories include heavy trucks (22.78%), four-wheel-drive vehicles (15.96%), and cars (9.53%). Various other vehicle types, such as buses, utility vehicles, and tractors, make up smaller proportions of the total vehicle composition.

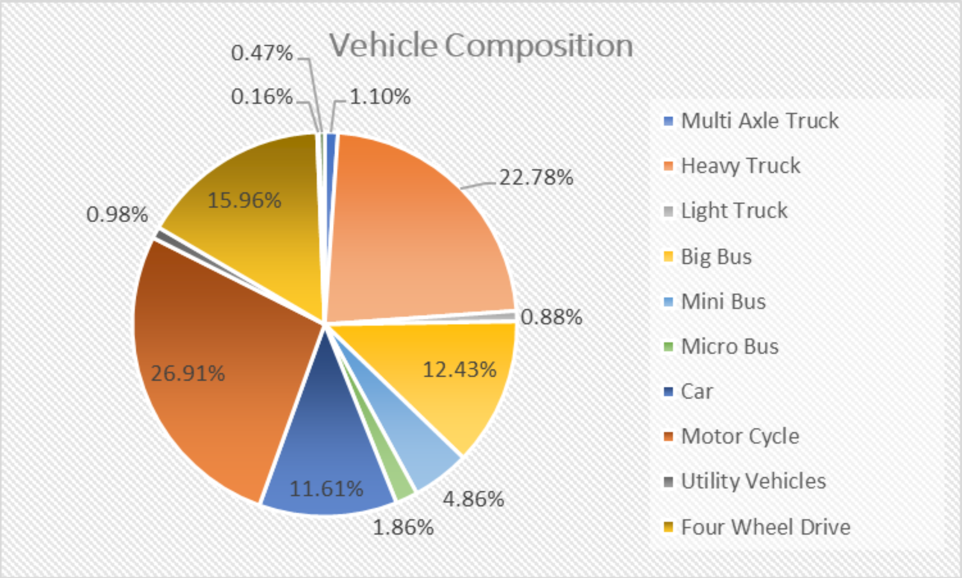


Figure 4-2: Vehicle composition at Dhankhola

Table 4-2 shows the summary of directional classified 72-hour traffic volume count at Bhaluwang road section. It provides a comprehensive overview of traffic data for two directions: Chanauta to Bhaluwang and Bhaluwang to Chanauta. It shows the three-day grand total, average daily traffic (ADT), and ADT in Passenger Car Units (PCU) for various vehicle types.

Table 4-2: Summary of directional classified 72-hour Traffic Volume Count at Dhankhola

	Chanauta to Bhaluwang			Bhaluwang to Chanauta		
Vehicle Type	3 Day Grand Total	Average Daily Traffic	ADT in PCU	3 Day Grand Total	Average Daily Traffic	ADT in PCU
Multi Axle Truck	48	16	64	57	19	76
Heavy Truck	1089	363	1089	1077	359	1077
Light Truck	42	14	21	44	14	21
Big Bus	595	198	594	588	196	588
Mini Bus	233	77	193	232	77	193
Micro Bus	87	29	44	91	30	45
Car	568	189	189	537	179	179
Motor Cycle	1243	414	207	1318	439	220
Utility Vehicles	44	14	14	51	17	17
Four Wheel Drive	754	251	251	767	255	255

	Chanauta to Bhaluwang			Bhaluwang to Chanauta		
Vehicle Type	**3 Day Grand Total**	**Average Daily Traffic**	**ADT in PCU**	**3 Day Grand Total**	**Average Daily Traffic**	**ADT in PCU**
Tractor	7	2	3	10	3	5
Motorized Three-Wheeler	22	7	6	24	8	6
Total	4732	1574	2674	4796	1596	2681

4.2 Speed Analysis

Spot speed survey dataset of 17 location in both directions is shown in APPENDIX-B.

The diverse range of vehicles has been divided into two categories during the speed study, a crucial step in the development of the VISSIM model. In this research, trucks and buses are categorized as heavy vehicles, while cars, four-wheel drives, utility vehicles, and micro-cars are considered light vehicles. Sampling of vehicles of speed survey and analysis were selected in proportional to the observed traffic volume to the best extent possible. Following the survey, data compilation was carried out according to road sections.

4.2.1 Percentile Speed Distribution at Dhankhola to Bhaluwang Section

Table 4-3 illustrates the percentile speed distribution of light and heavy vehicles in various segments of roads along the Dhankhola to Bhaluwang direction. Furthermore, it indicates that light vehicles are consistently faster than heavy vehicles in all sections of the road. A substantial reduction in speed is observed, especially in curve sections compared to others, while high speeds are evident in the straight sections of the road.

Table 4-3: Percentile speed distribution at Dhankhola-Bhaluwang section in km/hr.

Dhankhola to Bhaluwang	Percentile Speed	Light Vehicles	Heavy Vehicles
Straight Section	98^{th} percentile	66.22	38.86
	85^{th} percentile	45.65	34

Dhankhola to Bhaluwang	Percentile Speed	Light Vehicles	Heavy Vehicles
	50th percentile	37	32
	15th percentile	31	26
Curve Section	98th percentile	40	32
	85th percentile	38	31
	50th percentile	32	27
	15th percentile	25	23.1
Bridge Approach	98th percentile	43	34
	85th percentile	35	33
	50th percentile	30	26
	15th percentile	26	21.7
Junction	98th percentile	52.72	40
	85th percentile	42	37.5
	50th percentile	37	32
	15th percentile	31.6	24

4.2.2 Percentile Speed Distribution at Bhaluwang to Dhankhola Section

Table 4-4 illustrates the percentile speed distribution of light and heavy vehicles in various segments of roads along the Bhaluwang to Dhankhola direction. Furthermore, it indicates that light vehicles are consistently faster than heavy vehicles in all sections of the road. A substantial reduction in speed is observed, especially in curve sections compared to others, while high speeds are evident in the straight sections of the road.

Table 4-4: Percentile speed distribution at Bhaluwang-Dhankhola section in km/hr.

Bhaluwang to Dhankhola	Percentile Speed	Light Vehicles	Heavy Vehicles
Straight Section	98th percentile	62	35
	85th percentile	43	31
	50th percentile	35	29
	15th percentile	29	23
Curve Section	98th percentile	36	30

Bhaluwang to Dhankhola	Percentile Speed	Light Vehicles	Heavy Vehicles
	85th percentile	34	29
	50th percentile	28.2	25
	15th percentile	22	21
	98th percentile	38	32
	85th percentile	31	31
	50th percentile	27	24
Bridge Approach	15th percentile	23	20
	98th percentile	47	38
	85th percentile	37	35
	50th percentile	33	30.3
Junction	15th percentile	28	22

4.3 Travel Time Reliability

Travel time reliability refers to the consistency and predictability of travel times on given transportation route or corridor. Most approaches compare high-delay days to those with an average delay. The most effective methods of assessing travel time reliability are 90th or 95th percentile travel times (Administration, 2010). The mean, 90th and 95th percentile travel times of categorized vehicle under mixed traffic condition for both directions is presented based on 72-hour travel time dataset shown in APPENDIX-C.

4.3.1 Travel Time Reliability at Dhankhola to Bhaluwang Section

Table 4-5 represents travel time in terms of mean, 90th percentile reliability and 95th percentile reliability of Dhankhola to Bhaluwang direction. A total of 2,882 vehicles, for which travel time was collected, were recorded. The 90th percentile reliability, representing the travel time that 90% of the vehicles are expected to be below, varies between 31.36 minute for big bus and 52.03 minute for multi-axle truck. Mean travel times range from 24.83 minute to 35.68 minute. The results indicate that the multi-axle truck has the highest travel time, while the big bus has the lowest, followed by the four-wheel drive, car and light truck, and car.

Table 4-5: Percentile travel time reliability at Dhankhola-Bhaluwang section

Vehicle Type	Sample size	Mean Travel Time (minute)	90th Percentile Reliability (minute)	95th Percentile Reliability (minute)
Big Bus	513	24.83	31.36	34.39
Car	534	25.87	33.93	36.52
Four Wheel Drive	681	25.60	32.87	35.02
Heavy Truck	797	30.96	37.74	39.57
Light Truck	23	26.65	36.56	39.78
Micro Bus	66	26.76	31.84	36.37
Mini Bus	211	26.44	33.73	35.28
Multi Axle Truck	26	35.68	52.03	54.52
Utility Vehicles	31	26.74	34.77	37.29
Total	2882	27.04	35.38	37.56

4.3.2 Travel Time Reliability at Bhaluwang to Dhankhola Section

Table 4-6 represents travel time in terms of mean, 90th percentile reliability and 95th percentile reliability of Bhaluwang to Dhankhola direction A total of 2,887 vehicles, for which travel time was collected, were recorded. The 90th percentile reliability, representing the travel time that 90% of the vehicles are expected to be below, varies between 29.6 minute for big bus and 50.58 minute for multi-axle truck. Mean travel times range from 23.49 minute to 35.65 minute. The results indicate that the multi-axle truck has the highest travel time, while the big bus has the lowest, followed by the utility vehicles, light truck and microbus.

Table 4-6: Percentile travel time reliability at Bhaluwang to Dhankhola section

Vehicle Type	**Sample size**	**Mean Travel Time (minute)**	**90^{th} Percentile Reliability (minute)**	**95^{th} Percentile Reliability (minute)**
Big Bus	479	23.49	29.60	31.58
Car	517	28.05	35.92	37.29
Four Wheel Drive	712	30.38	38.50	39.70
Heavy Truck	805	31.37	40.05	43.75
Light Truck	36	26.76	35.60	39.21
Micro Bus	64	27.23	34.42	36.13
Mini Bus	207	30.96	37.64	39.42
Multi Axle Truck	31	35.65	50.58	50.72
Utility Vehicles	36	26.38	34.93	38.22
Total	2887	29.03	37.61	39.74

CHAPTER 5: TRAVEL TIME PREDICTION MODEL

In this chapter, the travel time models are developed illustrating the relationship among travel time, through traffic volume, opposing traffic volume, and the proportion of heavy vehicles. Traffic volume is expressed as PCU/hr., and the proportion of heavy vehicles is expressed as a decimal, while travel time is measured in minutes. The separate models are developed for distinct direction of travel. The data for development of model is arranged for each 30-minute interval since the average travel time is close to 30 minutes, and it is expected that the traffic volume within these intervals has a significant impact on the travel time of the respective vehicles that arrived during that time period. A 48-hour dataset is used to construct the model, whereas a 24-hour dataset is used to validate it. The data used for development of model is shown in APPENDIX-D.

5.1 Statistical Characteristics of Variables

Statistical characteristics such as mean, standard deviation, minimum and maximum value of each variable in both directions are expressed in Table 5-1. The data shows travel time of Bhaluwang to Dhankhola has more variation than Dhankhola to Bhaluwang direction.

Table 5-1: Statistical characteristics of variables

Variable	**Unit**	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
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

Variable	Unit	Mean	Standard Deviation	Minimum	Maximum
Opposing Volume	PCU/Hr.	105.30	35.80	26.00	206.00
Percentages of Heavy Vehicles in Through Traffic	Decimal Proportion	0.35	0.15	0.17	0.88

5.2 Correlation of Variables

Correlation analysis is a method used to examine the relationships between different variables. In this research, the Pearson correlation coefficient was employed to understand the associations among the variables. The Pearson correlation coefficient (ρxy) is calculated by dividing the covariance of two variables (COV (X, Y)) by the product of their individual standard deviations (σx, σy) (Taylor, 1990).

Figure 5-1 shows the relationships between travel time and independent variables in two directions of the road: "Bhaluwang to Dhankhola" and "Dhankhola to Bhaluwang." The values in the matrix range from -1 to 1, with 1 indicating a perfect positive correlation, -1 representing a perfect negative correlation, and 0 denoting no correlation. In both directions, travel time exhibits a moderate strong positive correlation with through traffic volume, opposing traffic volume and percentages of heavy vehicles in through traffic, implying that as these variables increase, travel time tends to travel time increases as well.

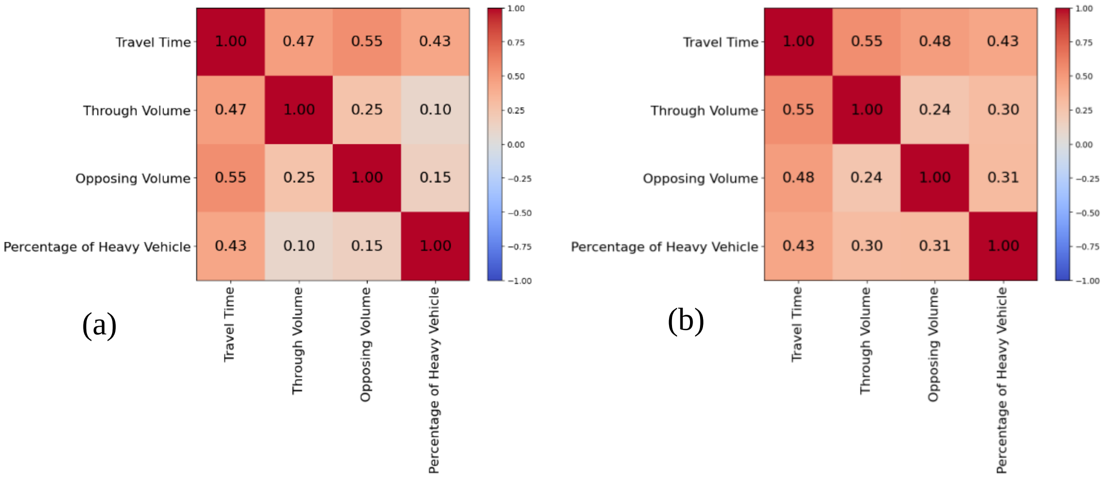


Figure 5-1: Correlation Matrix: (a) Dhankhola to Bhaluwang direction (b) Bhaluwang to Dhankhola direction

5.3 Travel Time Prediction Model

Several techniques have been used to develop a travel time prediction model. This involves the application of diverse supervised machine learning regression techniques to establish a model capable of predicting travel time by considering elements like through traffic volume, opposing traffic volume and the percentage of heavy vehicles in the through traffic. The dataset covers a span of three days and encompasses a total of 144 data points, which are then divided into training and test sets to facilitate various regression analyses. Specifically, the training set incorporates data from two days, comprising 96 observations, while the test set consists of data from one day, totaling 48 observations. This analysis is independently carried out for both travel directions.

5.3.1 Multiple Linear Regression Model

The Multiple Linear Regression tool in Excel was employed to develop a travel time prediction model. By fitting a linear equation to the data, the tool enables the estimation of how changes in the predictor variables are associated with changes in the target variable. The prediction model developed from the training dataset is used to predict travel time in the test dataset, followed by the evaluation of performance metrics using actual travel times in the test dataset.

Table 5-2 shows the outcomes of the analysis of variance (ANOVA), indicating a noteworthy relationship between the variables. The critical value from the F-table at a significance threshold of 0.05 stands at 2.7, which is surpassed by the computed F-values of 24.33 for the Dhankhola to Bhaluwang direction and 27.40 for the Bhaluwang to Dhankhola direction. This signifies the rejection of the null hypothesis in favor of the alternative hypothesis, underscoring a substantial relationship among the groups employed in the multiple linear regression model.

Table 5-2: Statistical ANOVA measure of Multiple Linear Regression Model

	DF	*SS*	*MS*	*F*	*Significance F*
Dhankhola to Bhaluwang Direction					
Regression	3	279.57	93.19	24.33	0.00
Residual	92	352.44	3.83		
Total	95	632.01			
Bhaluwang to Dhankhola Direction					
Regression	3	524.00	174.67	27.40	0.00
Residual	92	586.48	6.37		
Total	95	1110.48			

Table 5-3 depicts the regression result of a multiple linear regression analysis. The independent variables, namely through volume, opposing volume and percentage of heavy vehicle in through traffic, exhibit statistically significant relationships with travel time, as indicated by their p-values which is lower than significance level of 0.05 for both directions.

Table 5-3: Regression results from Multiple Linear Regression Model

	Coefficients	Standard Error	P-value
Dhankhola to Bhaluwang Direction			
Intercept	19.243	1.067	0.000
Through Volume	0.026	0.006	0.000
Opposing Volume	0.019	0.003	0.000
Percentages of Heavy Vehicles in Through Traffic	2.803	1.468	0.047
Bhaluwang to Dhankhola Direction			
Intercept	16.848	1.755	0.000
Through Volume	0.035	0.006	0.000

	Coefficients	Standard Error	P-value
Opposing Volume	0.039	0.008	0.000
Percentages of Heavy Vehicles in Through Traffic	5.970	2.212	0.008

The model in the form of equations from the multiple linear regression model for predicting travel time is shown in Eq. 5.1 and Eq. 5.2 for the Dhankhola to Bhaluwang direction and Bhaluwang to Dhankhola direction, respectively.

Travel Time (min)= *19.243 +0.026* Through Volume (PCU/hr.) +0.019* Opposing Volume (PCU/hr.) +2.803* Percentages of Heavy Vehicles in Through Traffic (Decimal)* **5.1**

Travel Time (min)= *16.848 +0.035* Through Volume (PCU/hr.) +0.039* Opposing Volume (PCU/hr.) +5.970* Percentages of Heavy Vehicles in Through Traffic (Decimal)* **5.2**

Table 5-4 presents performance metrics for travel time prediction models in two directions. In the Dhankhola to Bhaluwang direction, the model demonstrates good accuracy with a Mean Absolute Error (MAE) of 0.95, Mean Squared Error (MSE) of 1.32, Root Mean Squared Error (RMSE) of 1.15, Mean Absolute Percentage Error (MAPE) of 3.68%, and an R-squared (R^2) value of 0.44. Similarly, for the Bhaluwang to Dhankhola direction, the model performs well with a MAE of 1.52, MSE of 3.67, RMSE of 1.92, MAPE of 5.58%, and an R^2 value of 0.45. Multiple linear regression model demonstrates significant predictive capability in both directions.

Table 5-4: Regression metrics from Multiple Linear Regression Model

Direction	MAE	MSE	RMSE	MAPE	Adj.R^2
Dhankhola to Bhaluwang	0.95	1.32	1.15	3.68%	0.44
Bhaluwang to Dhankhola	1.52	3.67	1.92	5.58%	0.45

The comparison between actual and predicted values of travel time prediction model, based on multiple linear regression model, for both directions from Dhankhola to Bhaluwang direction and Bhaluwang to Dhankhola direction, is shown in Figure 5-2(a) and Figure 5-3(a) respectively. Additionally, the distribution of errors between the predicted and actual values is shown in Figure 5-2(b) and Figure 5-3(b) respectively.

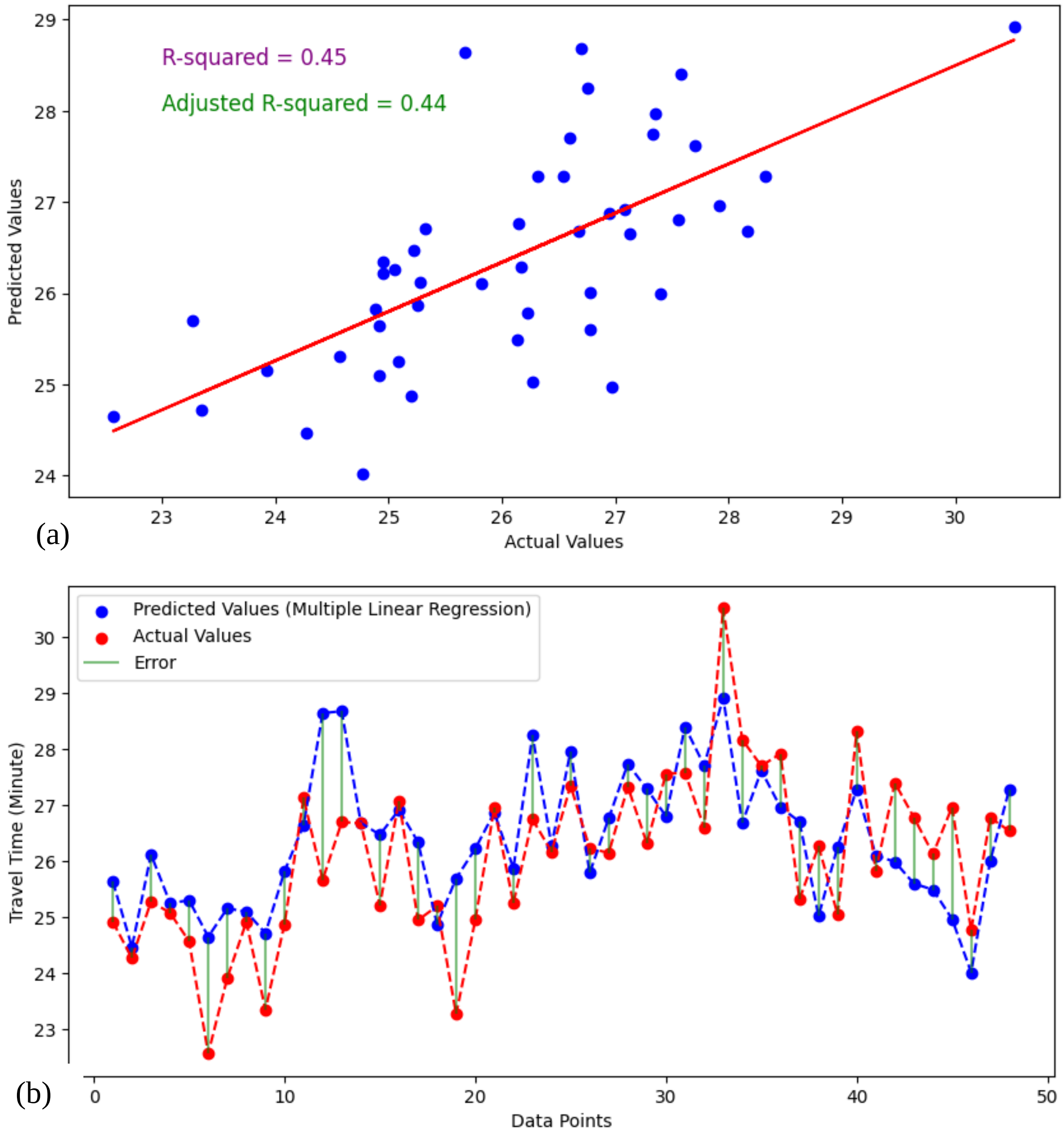


Figure 5-2: Travel Time Prediction based on Multiple Linear Regression model for Dhankhola to Bhaluwang direction (a) Prediction (b) Distribution of tested and predicted error

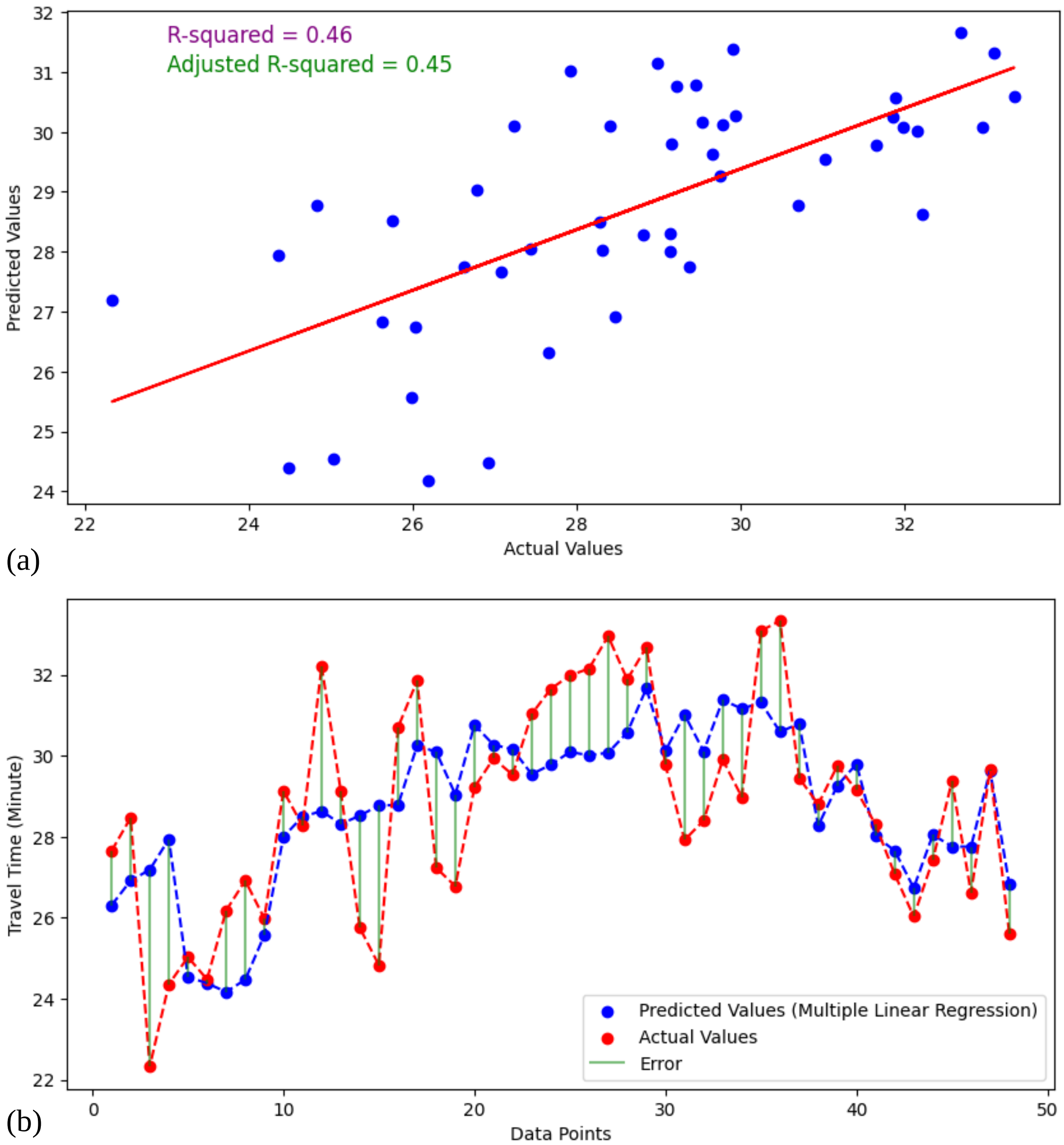


Figure 5-3: Travel Time Prediction based on Multiple Linear Regression model for Bhaluwang to Dhankhola direction (a) Prediction (b) Distribution of tested and predicted error

5.3.2 LASSO Regression

The methodology employed in this section leverages LASSO Regression, a powerful technique in statistical modeling known for its ability to perform variable selection and regularization. The Python code implementation used for this purpose is extensively documented in APPENDIX-E.

Table 5-5 presents performance metrics for travel time prediction models in two directions. In the Dhankhola to Bhaluwang direction, the model demonstrates good accuracy with a Mean Absolute Error (MAE) of 0.87, Mean Squared Error (MSE) of

1.14, Root Mean Squared Error (RMSE) of 1.07, Mean Absolute Percentage Error (MAPE) of 3.37%, and an R-squared (R^2) value of 0.48. Similarly, for the Bhaluwang to Dhankhola direction, the model performs well with a MAE of 1.37, MSE of 3.75, RMSE of 1.93, MAPE of 5.57%, and an R^2 value of 0.49. LASSO regression model demonstrates significant predictive capability in both directions.

Table 5-5: Regression metrics from LASSO Regression Model

Direction	MAE	MSE	RMSE	MAPE	Adj.R^2
Dhankhola to Bhaluwang	0.87	1.14	1.07	3.37%	0.48
Bhaluwang to Dhankhola	1.37	3.75	1.93	5.57%	0.49

The comparison between actual and predicted values of travel time prediction model, based on LASSO regression model, for both directions from Dhankhola to Bhaluwang direction and Bhaluwang to Dhankhola direction, is shown in Figure 5-4(a) and Figure 5-5(a) respectively. Additionally, the distribution of errors between the predicted and actual values is shown in Figure 5-4(b) and Figure 5-5(b) respectively.

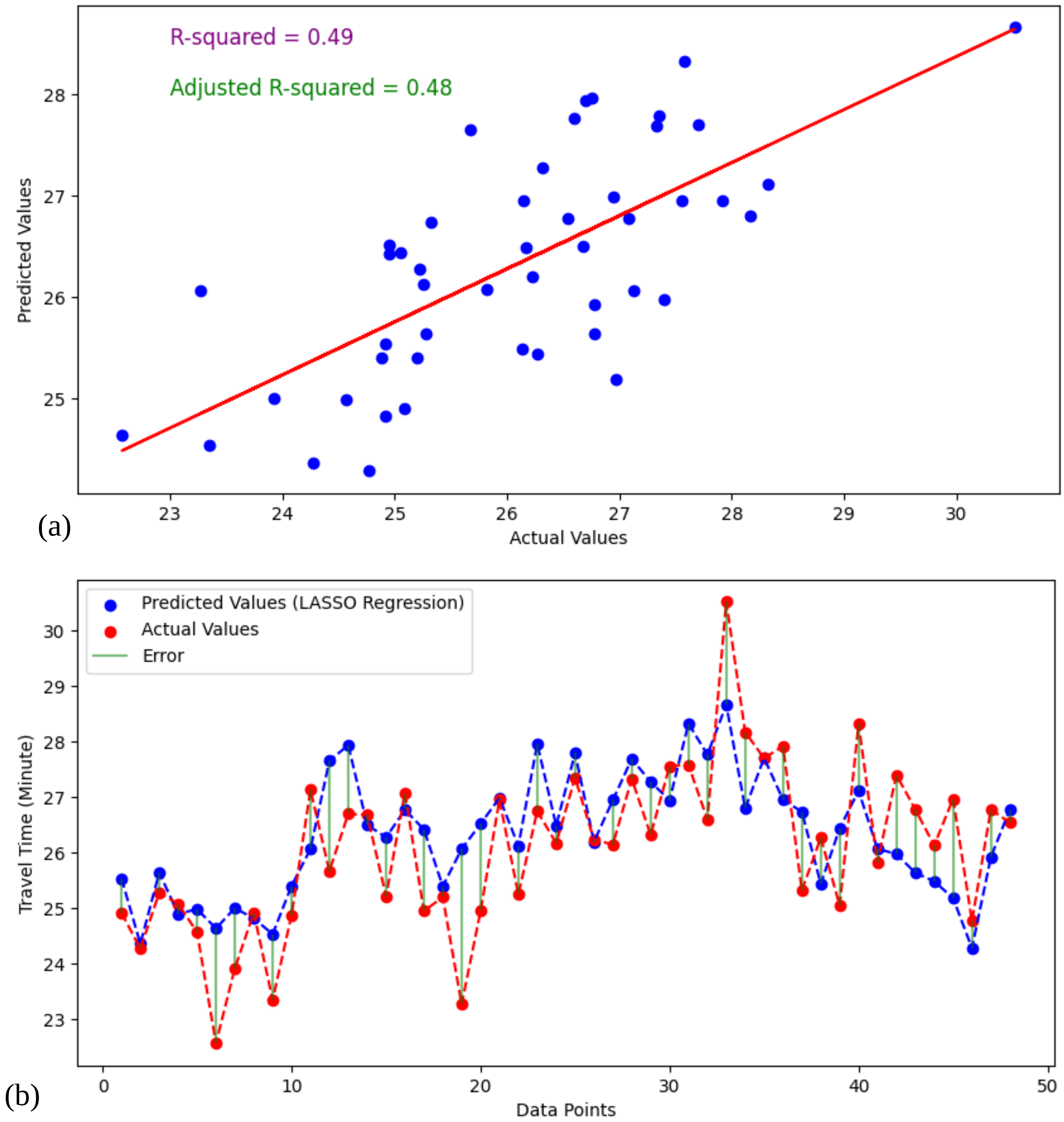
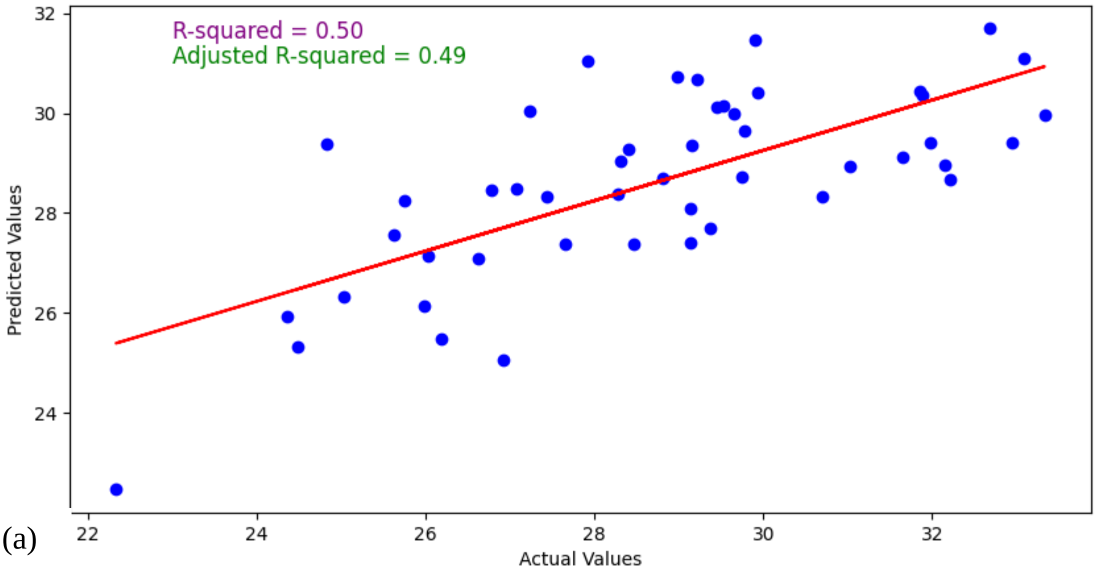
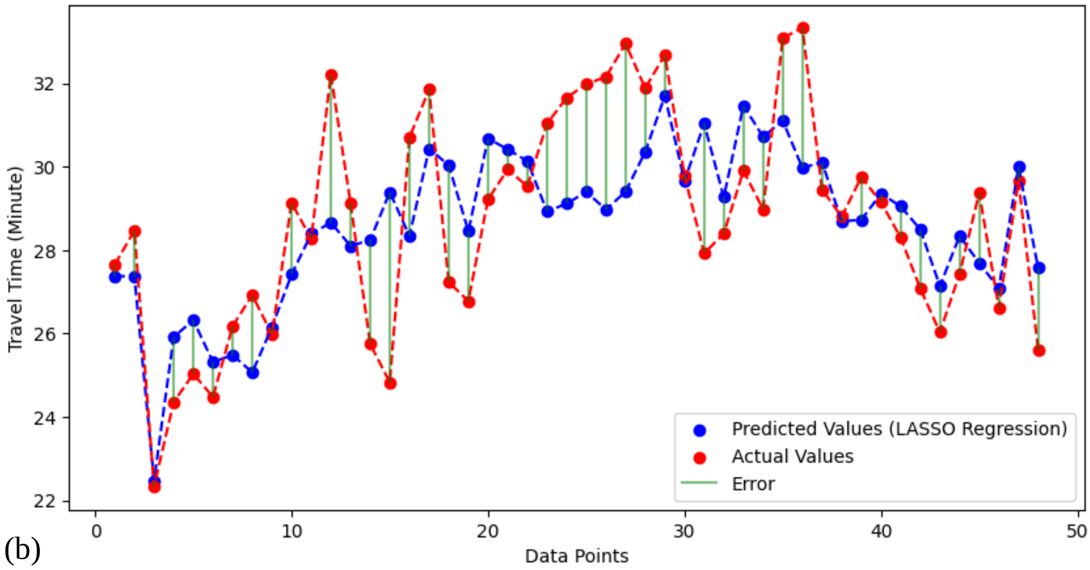


Figure 5-4: Travel Time Prediction based on LASSO Regression model for Dhankhola to Bhaluwang direction (a) Prediction (b) Distribution of tested and predicted error



(a)



(b)

Figure 5-5: Travel Time Prediction based on LASSO Regression model for Bhaluwang to Dhankhola direction (a) Prediction (b) Distribution of tested and predicted error

5.3.3 Decision Tree Regression

The methodology employed in this section incorporates Decision Tree Regression, a robust technique widely utilized in predictive modeling due to its ability to capture complex relationships within datasets. The implementation of Decision Tree Regression in Python is extensively detailed in APPENDIX-E.

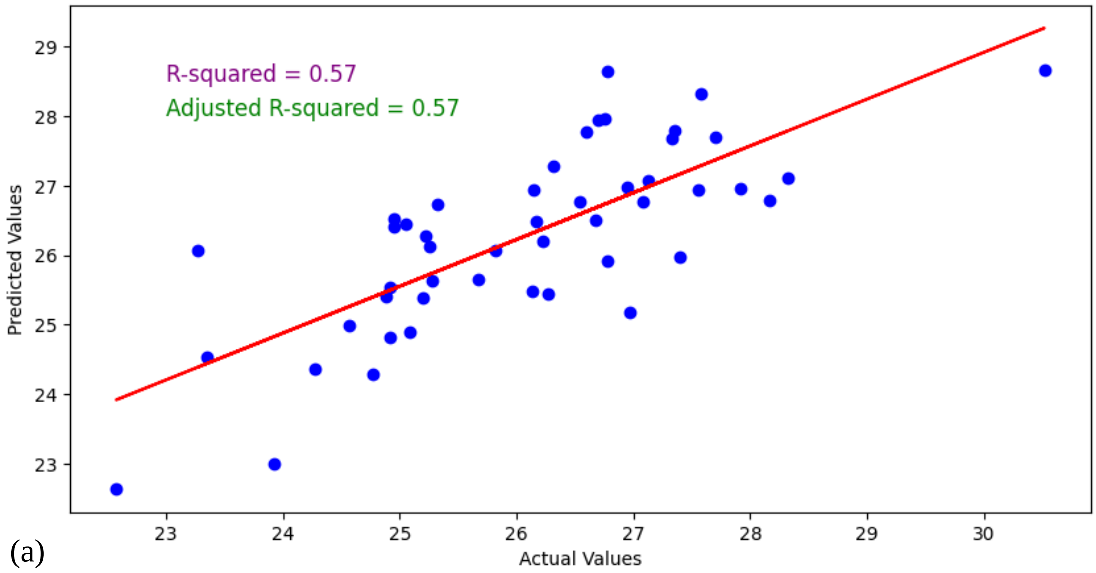
Table 5-6 presents performance metrics for travel time prediction models in two directions. In the Dhankhola to Bhaluwang direction, the model demonstrates good accuracy with a Mean Absolute Error (MAE) of 0.81, Mean Squared Error (MSE) of 0.9, Root Mean Squared Error (RMSE) of 0.95, Mean Absolute Percentage Error

(MAPE) of 2.58%, and an R-squared (R^2) value of 0.57. Similarly, for the Bhaluwang to Dhankhola direction, the model performs well with a MAE of 1.31, MSE of 2.53, RMSE of 1.59, MAPE of 4.59%, and an R^2 value of 0.61. Decision Tree Regression model demonstrates significant predictive capability in both directions.

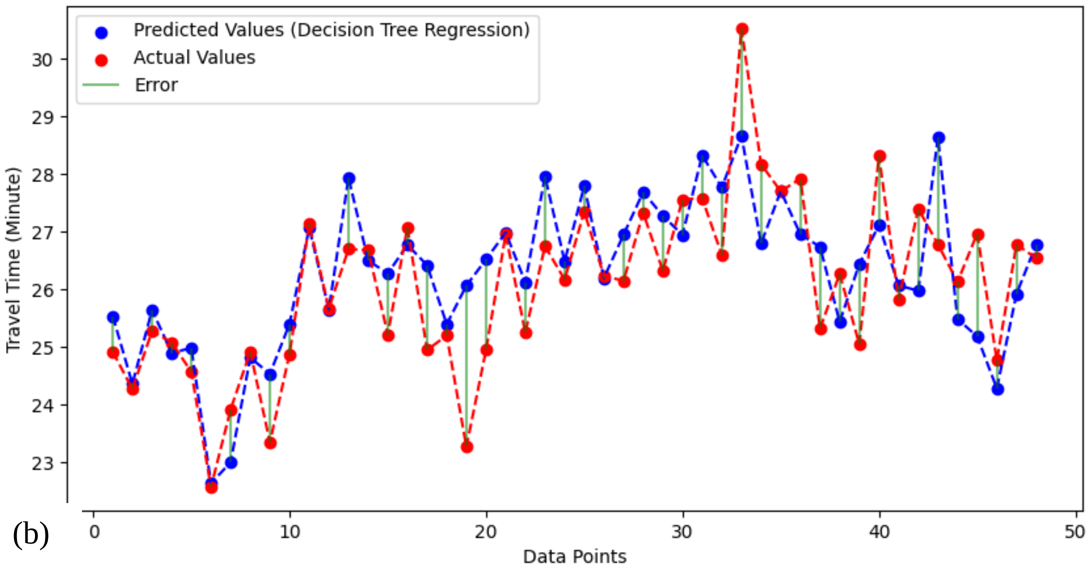
Table 5-6: Regression metrics from Decision Tree Regression Model

Direction	MAE	MSE	RMSE	MAPE	Adj.R^2
Dhankhola to Bhaluwang	0.81	0.9	0.95	2.58%	0.57
Bhaluwang to Dhankhola	1.31	2.53	1.59	4.59%	0.61

The comparison between actual and predicted values of travel time prediction model, based on decision tree regression model, for both directions from Dhankhola to Bhaluwang direction and Bhaluwang to Dhankhola direction, is shown in Figure 5-6(a) and Figure 5-7(a) respectively. Additionally, the distribution of errors between the predicted and actual values is shown in Figure 5-6(b) and Figure 5-7(b) respectively.



(a)



(b)

Figure 5-6: Travel Time Prediction based on Decision Tree Regression model for Dhankhola to Bhaluwang direction (a) Prediction (b) Distribution of tested and predicted error

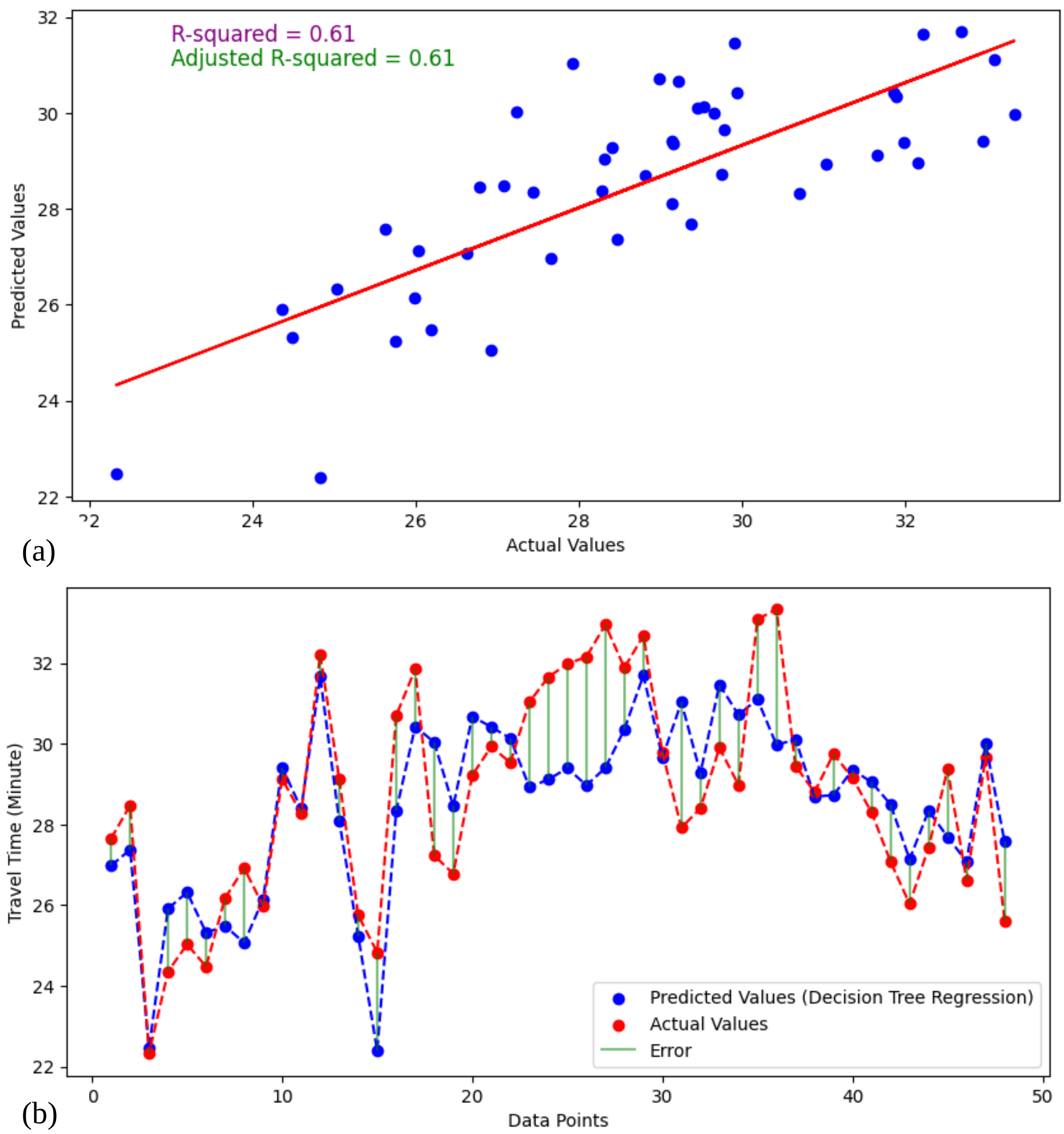


Figure 5-7: Travel Time Prediction based on Decision Tree Regression model for Bhaluwang to Dhankhola direction (a) Prediction (b) Distribution of tested and predicted error

5.3.4 Support Vector Regression

The methodology adopted in this study incorporates Support Vector Regression (SVR), a powerful technique known for its effectiveness in modeling complex relationships and handling non-linear data. The Python code implementation for SVR is meticulously outlined in APPENDIX-E.

Table 5-7 presents performance metrics for travel time prediction models in two directions. In the Dhankhola to Bhaluwang direction, the model demonstrates good accuracy with a Mean Absolute Error (MAE) of 0.80, Mean Squared Error (MSE) of 0.88, Root Mean Squared Error (RMSE) of 0.94, Mean Absolute Percentage Error

(MAPE) of 2.53%, and an R-squared (R^2) value of 0.58. Similarly, for the Bhaluwang to Dhankhola direction, the model performs well with a MAE of 1.30, MSE of 2.56, RMSE of 1.60, MAPE of 3.98%, and an R^2 value of 0.61. Support Vector Regression model demonstrate significant predictive capability in both directions.

Table 5-7: Regression metrics from Support Vector Regression Model

Direction	MAE	MSE	RMSE	MAPE	Adj.R^2
Dhankhola to Bhaluwang	0.80	0.88	0.94	2.53%	0.58
Bhaluwang to Dhankhola	1.30	2.56	1.60	3.98%	0.61

The comparison between actual and predicted values of travel time prediction model, based on support vector regression model, for both directions from Dhankhola to Bhaluwang direction and Bhaluwang to Dhankhola direction, is shown in Figure 5-8(a) and Figure 5-9(a) respectively. Additionally, the distribution of errors between the predicted and actual values is shown in Figure 5-8(b) and Figure 5-9(b) respectively.

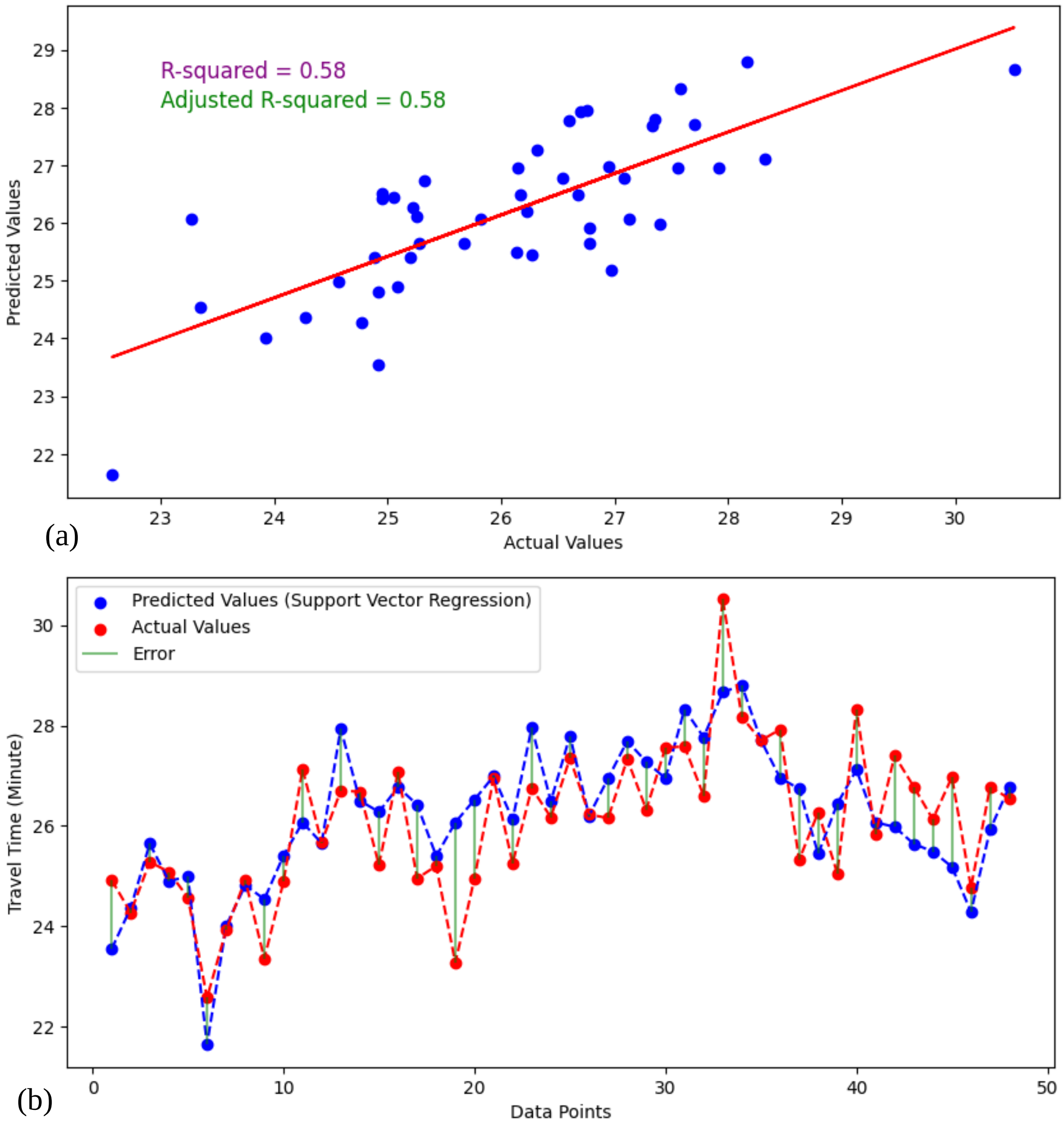


Figure 5-8: Travel Time Prediction based on Support Vector Regression model for Dhankhola to Bhaluwang direction (a) Prediction (b) Distribution of tested and predicted error

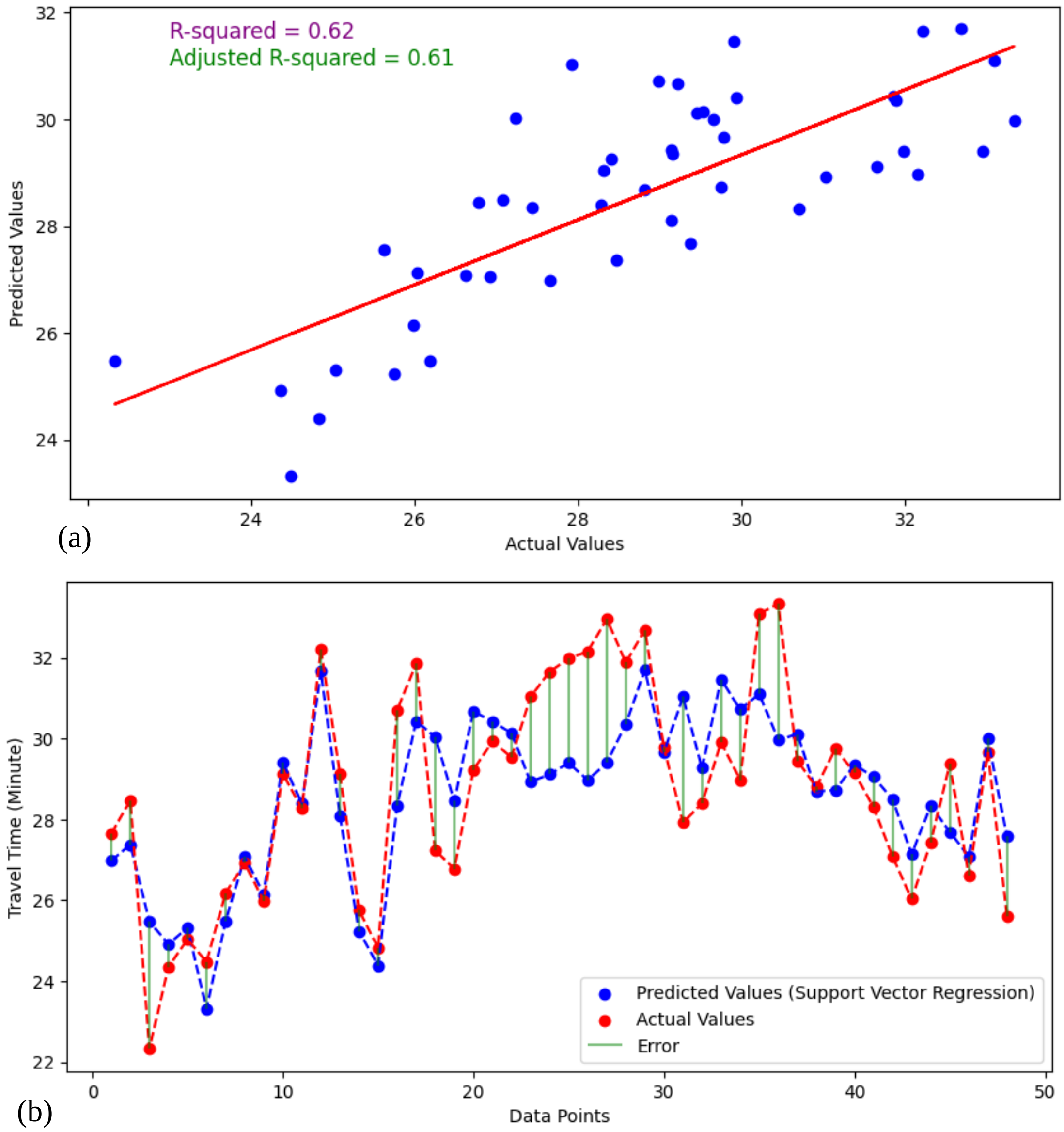


Figure 5-9: Travel Time Prediction based on Support Vector Regression model for Bhaluwang to Dhankhola direction (a) Prediction (b) Distribution of tested and predicted error

5.3.5 Random Forest Regression

The methodology employed in this study incorporates Random Forest Regression, a powerful ensemble learning technique renowned for its ability to capture intricate relationships within datasets. The Python code implementation for Random Forest Regression is meticulously detailed in APPENDIX-E.

Table 5-8 presents performance metrics for travel time prediction models in two directions. In the Dhankhola to Bhaluwang direction, the model demonstrates good accuracy with a Mean Absolute Error (MAE) of 0.74, Mean Squared Error (MSE) of 0.81, Root Mean Squared Error (RMSE) of 0.9, Mean Absolute Percentage Error

(MAPE) of 2.48%, and an R-squared (R^2) value of 0.62. Similarly, for the Bhaluwang to Dhankhola direction, the model performs well with a MAE of 1.16, MSE of 2.14, RMSE of 1.46, MAPE of 3.92%, and an R^2 value of 0.68. Random Forest Regression model demonstrate significant predictive capability in both directions.

Table 5-8: Regression metrics from Random Forest Regression Model

Direction	MAE	MSE	RMSE	MAPE	Adj.R^2
Dhankhola to Bhaluwang	0.74	0.81	0.9	2.48%	0.62
Bhaluwang to Dhankhola	1.16	2.14	1.46	3.92%	0.68

The comparison between actual and predicted values of travel time prediction model, based on random forest regression model, for both directions from Dhankhola to Bhaluwang direction and Bhaluwang to Dhankhola direction, is shown in Figure 5-10(a) and Figure 5-11(a) respectively. Additionally, the distribution of errors between the predicted and actual values is shown in Figure 5-10(b) and Figure 5-11(b) respectively.

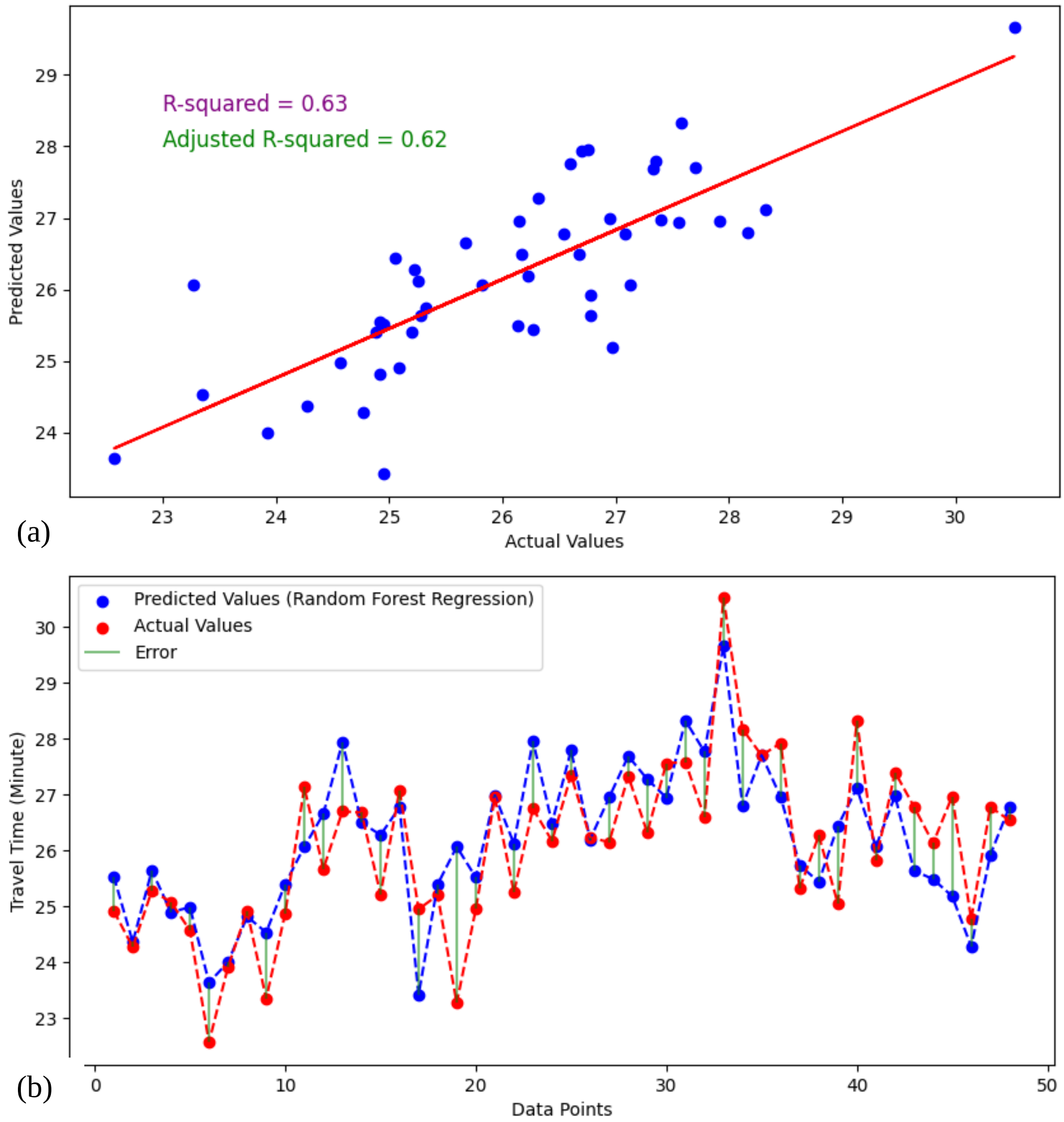
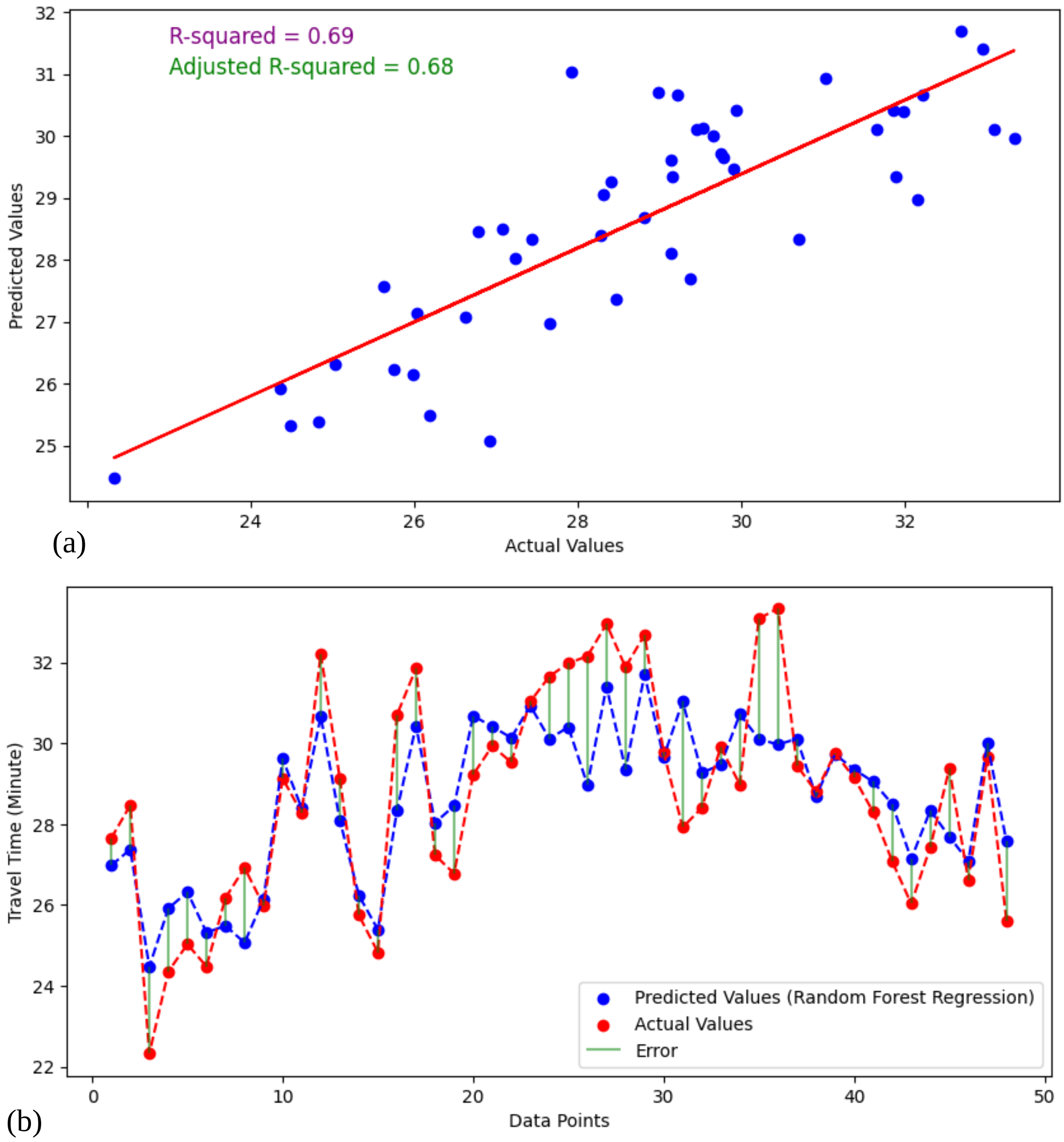


Figure 5-10: Travel Time Prediction based on Random Forest Regression model for Dhankhola to Bhaluwang direction (a) Prediction (b) Distribution of tested and predicted error



(a)
(b)

Figure 5-11: Travel Time Prediction based on Random Forest Regression model for Bhaluwang to Dhankhola direction (a) Prediction (b) Distribution of tested and predicted error

5.4 Discussion

Through the analysis of results of the experimentation, it was possible to draw some interesting findings.

The selected independent variable for predicting travel time is the volume of opposing traffic. In the context of a two-lane, two-way undivided highway, the success of overtaking maneuvers is notably dependent on the level of traffic from the opposite direction, especially in situations with mixed traffic flow (Asaithambi & Shravani, 2017). When the volume of opposing traffic is low, faster vehicles can adeptly overtake slower ones in the oncoming lane. Conversely, when the traffic volume in the oncoming lane is substantial, it hinders the feasibility of overtaking maneuvers. This phenomenon has a direct impact on travel time; greater opposing traffic volume correlates with increased travel time, whereas lower opposing traffic volume is associated with a reduction, as demonstrated in Section 5.2 for both directions.

In Section 2.2, researchers predominantly use through traffic volume to predict travel time on freeways and multilane highways. This choice is supported by a consistent finding in past literature and is further confirmed by the travel time function (Manual, 1964). This reaffirms the significance of traffic volume as a key factor in accurately estimating travel time in these contexts. Based on the findings from Section 5.2, travel time is directly influenced by the through traffic volume. An increase in through traffic volume causes an increase in travel time, and vice versa.

Heavy vehicles, including trucks with multiple axles and large buses, are categorized as slower-moving vehicles. Conversely, four-wheel drives and passenger cars are considered light vehicles and tend to move at higher speeds compared to heavy vehicles. This observation is substantiated by travel time data gathered from field. Given the speed differential, faster-moving vehicles often need to perform overtaking maneuvers, a scenario particularly common on two-lane, two-way undivided highways where all vehicles share a single lane. A higher percentage of heavy vehicles within the through traffic necessitates a greater number of overtaking maneuvers. In order to execute these maneuvers, fast-moving vehicles must temporarily match the speed of the slower-moving vehicle before overtaking. Consequently, an increased proportion of heavy vehicles leads to longer travel times, while a decreased proportion results in shorter travel times. This pattern is supported by the findings presented in Section 5.2, aligning with previous simulation-based research in multilane highway discussed in Section 2.3.

The most important point of this study was may be that existence of a significant nonlinear relationship between travel time and the independent variables in predicting travel time on a two-lane, undivided highway, which is akin to what has been observed in models for multilane highways, where traffic volume is the sole independent variable (Liu, Wang, Yang, & Zhang, 2017).

Finally, when examining the performance of regression models, it is evident from the analysis in Section 5.3 and the subsequent statistical assessment illustrated in Figure 5-12 and Figure 5-13, that Random Forest Regression yielded the most favorable outcomes across various metrics. This aligns with previous research by (Sharma, Singh, & Upteti, 2021) on travel time prediction models. In the Dhankhola to Bhaluwang direction, it achieved scores of 0.74 for MAE, 0.9 for RMSE, 2.48% for MAPE, and 0.62 for R^2. Similarly, in the Bhaluwang to Dhankhola direction, the results were notably consistent, yielding values of 1.10 for MAE, 1.46 for RMSE, 3.92% for MAPE, and 0.68 for R^2.

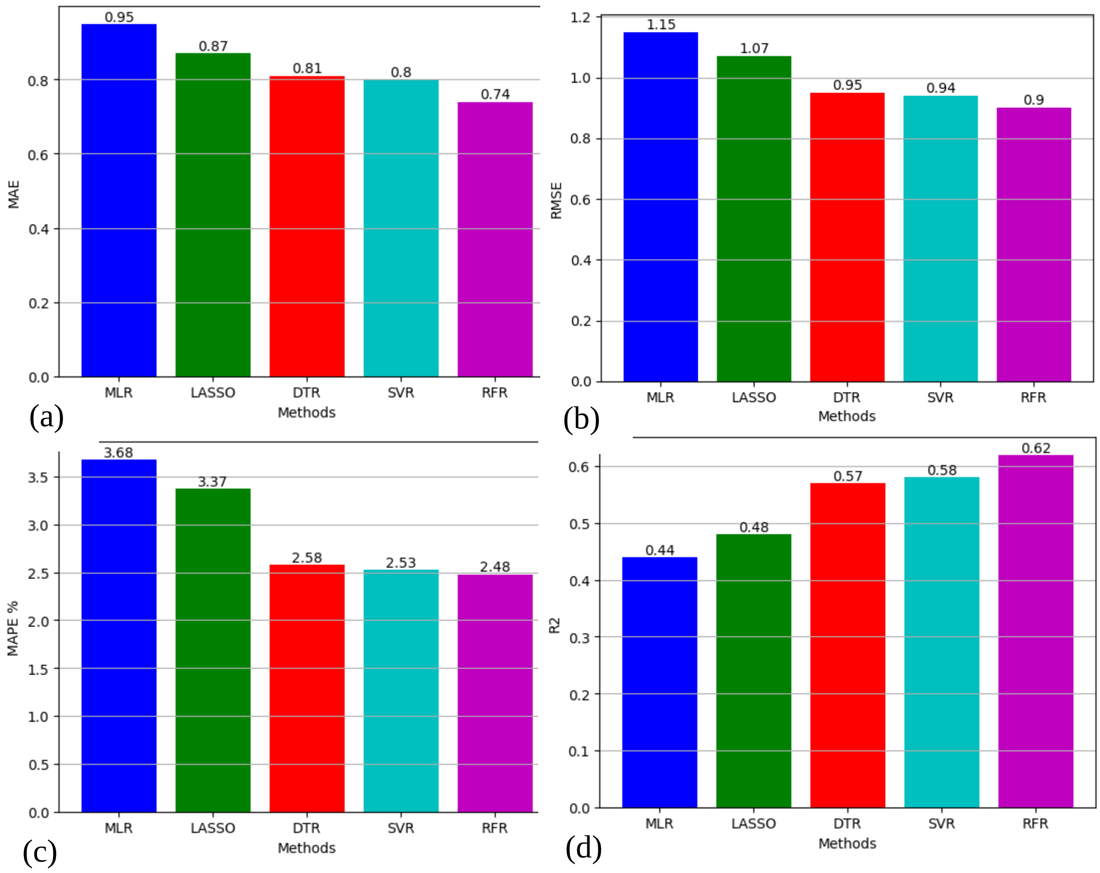


Figure 5-12: Comparison of regression results and statistical errors of prediction models at Dhankhola to Bhaluwang direction: (a) MAE (b) RMSE (c) MAPE (d) R^2

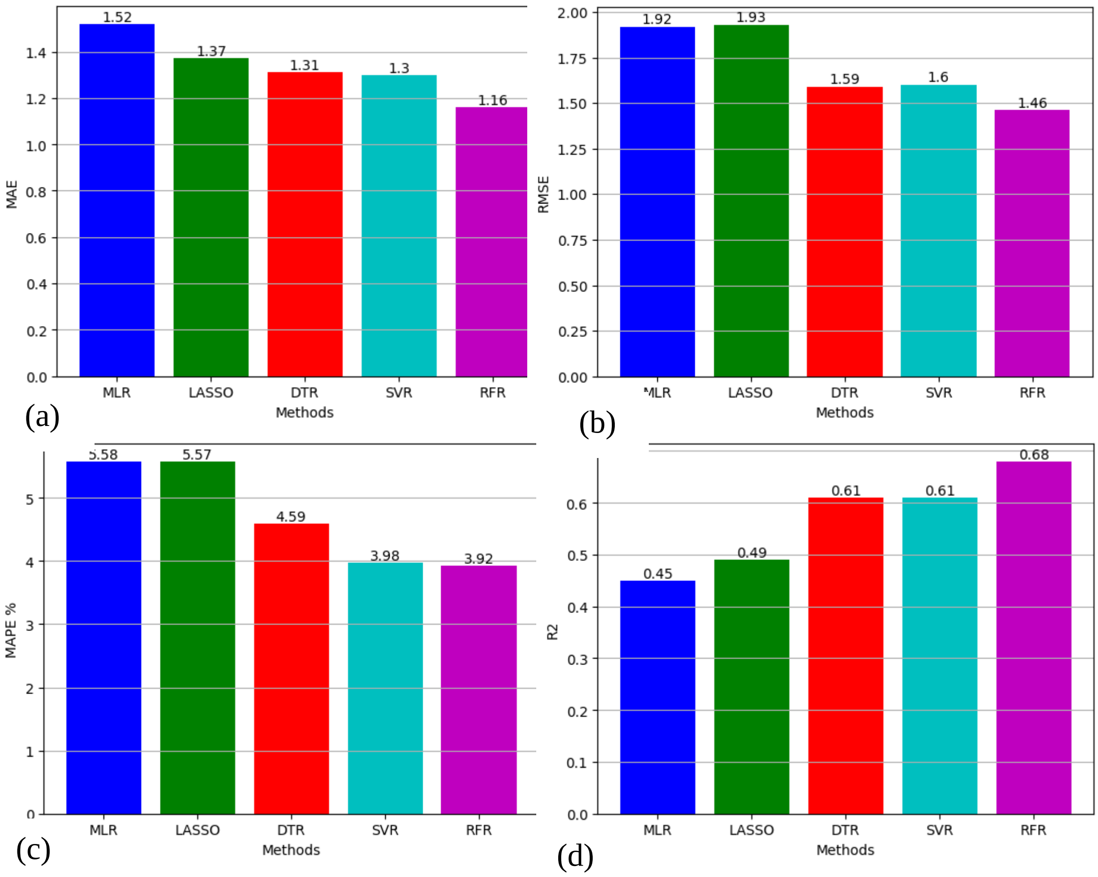


Figure 5-13: Comparison of regression results and statistical errors of prediction models at Bhaluwang to Dhankhola direction: (a) MAE (b) RMSE (c) MAPE (d) R^2

CHAPTER 6: TRAVEL TIME FUNCTION

6.1 Development of Simulation Model

A simulation model was developed for the Dhankhola-Bhaluwang road section, consisting of a two-lane, two-way undivided carriageway. The road alignment was established using a Google base map in VISSIM software. The total length of the network is 12.9 km in each direction. The geometrical layout for the simulation model, including lane width, was measured in a field survey. Overtaking zones were provided along the oncoming lane in both directions of travel. Traffic data used in the development of the model was collected from field observations. The speed distribution of light and heavy vehicles was measured in the field and subsequently incorporated into the model, taking into account vehicle category and road elements such as bridges, curves, straight segments, and intersections. Figure 6-1 shows the sample of simulation model developed in VISSIM software.

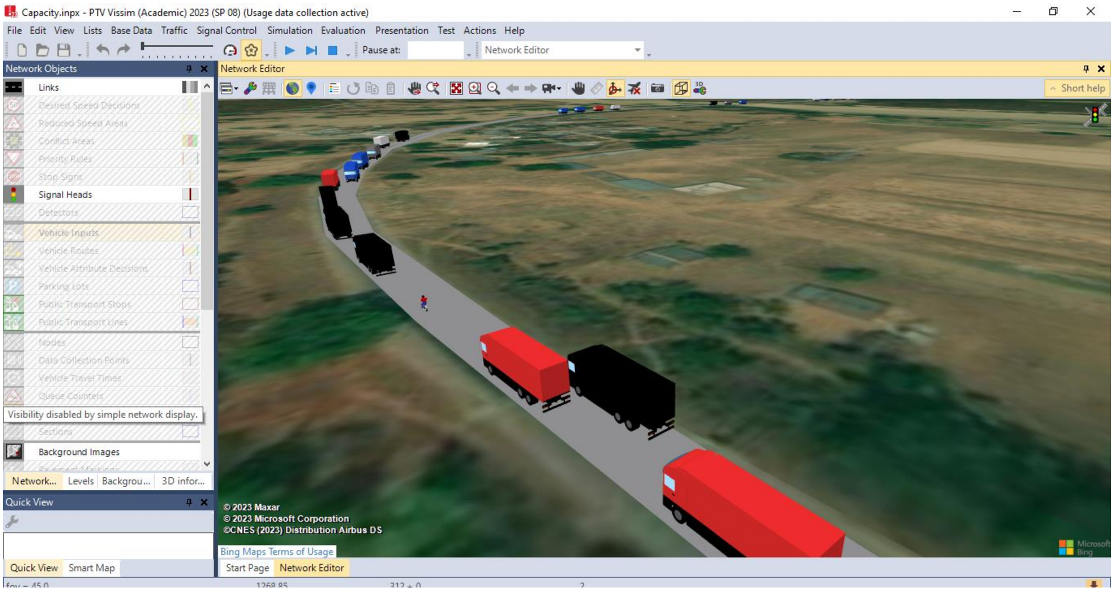


Figure 6-1: Developed VISSIM simulation model (sample)

6.2 Calibration of Simulation Model

The proposed model is then calibrated with the collected data macroscopically and microscopically. Microscopic calibration is conducted for light and heavy vehicles parameters. Vehicular and road geometry characteristics where calibrated using field measurement of properties of individual vehicles characteristics and road layout geometry. The initial driving behavior factor for first trial being used based on the study carried out back in 2013 in Indian heterogeneous traffic condition (Siddharth & Ramadurai, 2013). The initial driving behavior parameters CC_0 and CC_1 (Standstill distance and gap time distribution) is used the study conducted back in 2014 in Indian mixed traffic condition (Mehar, Chandra, & Velmurugan, 2014). The calibrated driving behavior parameter of the simulation model is shown in Table 6-1.

Table 6-1: Calibrated parameters used in simulation model

Gap Time Distribution (CC1)	1.18
Average Standstill Distance (CC2)	1
Additive Part of Safety Distance	0.2
Multiplicative Part of Safety Distance	0.78
Minimum Lateral Distance- bike at 0 Kmph	0.62
Look Ahead Distance-minimum	27.91
Look Back Distance-minimum	14.31
Look ahead distance for overtaking in the opposing lane-minimum	375
Overtaking speed factor in the opposing lane	1.25

6.3 Validation of Simulation Model

The proposed model's validation was conducted using data from the 10 PM to 6 AM timeframe over a period of three days (48 datasets). During validation, the average travel time data were compared to the actual field measurements. Initially, the relationship between travel time and traffic volume, as derived from simulation results, was compared to the field data. Other variables like opposing traffic volume and the proportion of heavy vehicles were kept consistent with the field data in the simulation. Figure 6-2 illustrates the comparison between the travel time-volume relationships derived from both the field data and the simulation. It is evident from the figure that the simulated travel time-volume relationship generally aligns well with the observed field data. Any significant disparities observed under moderate traffic conditions are likely attributable to the inherent stochasticity and instability of traffic flow.

Subsequently, we assessed the model's performance in the time domain by comparing the simulated time series of travel time data with the field data. Figure 6-3 depicts this comparison, revealing a satisfactory match between the simulated travel time and the field data, albeit with notable discrepancies during periods of traffic instability.

The Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) were employed to gauge the overall error in aggregated speed between the simulation results and the field data, yielding values of 2.88 minutes and 9.55% respectively. These error metrics fall within the acceptable range specified by Meng and Weng for CA-based models concerning travel speed (Meng & Weng, 2011) These findings affirm that the proposed model adeptly reproduces disrupted traffic flow in a realistic manner.

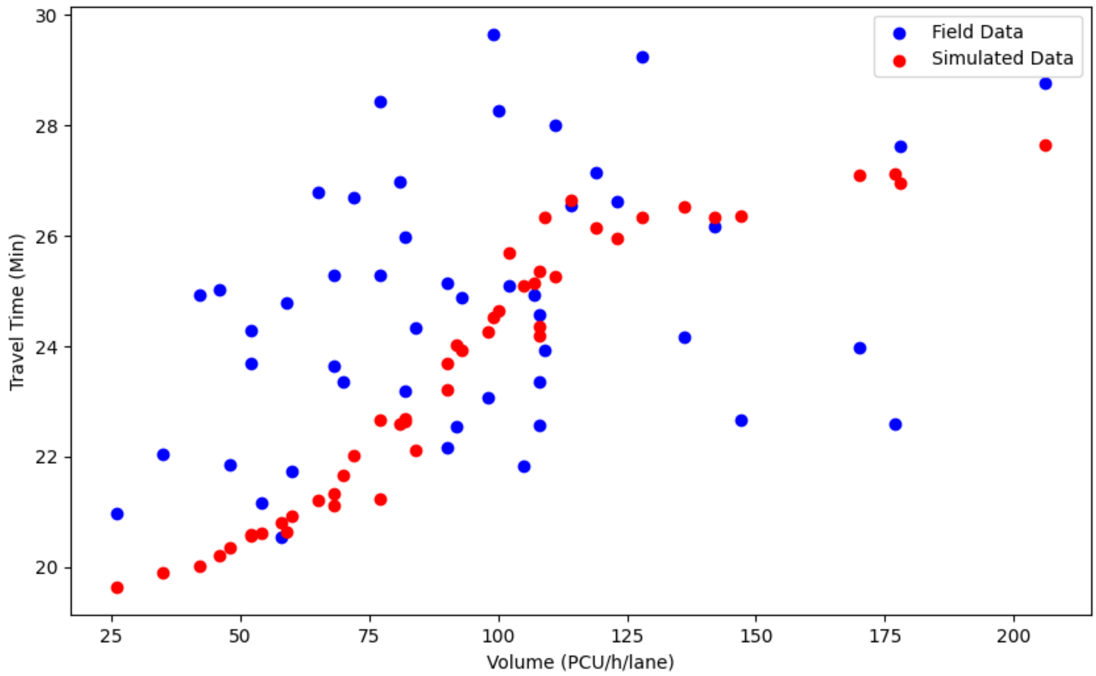


Figure 6-2: Comparison of Travel Time-Volume Relationship from field dCoata and simulation result

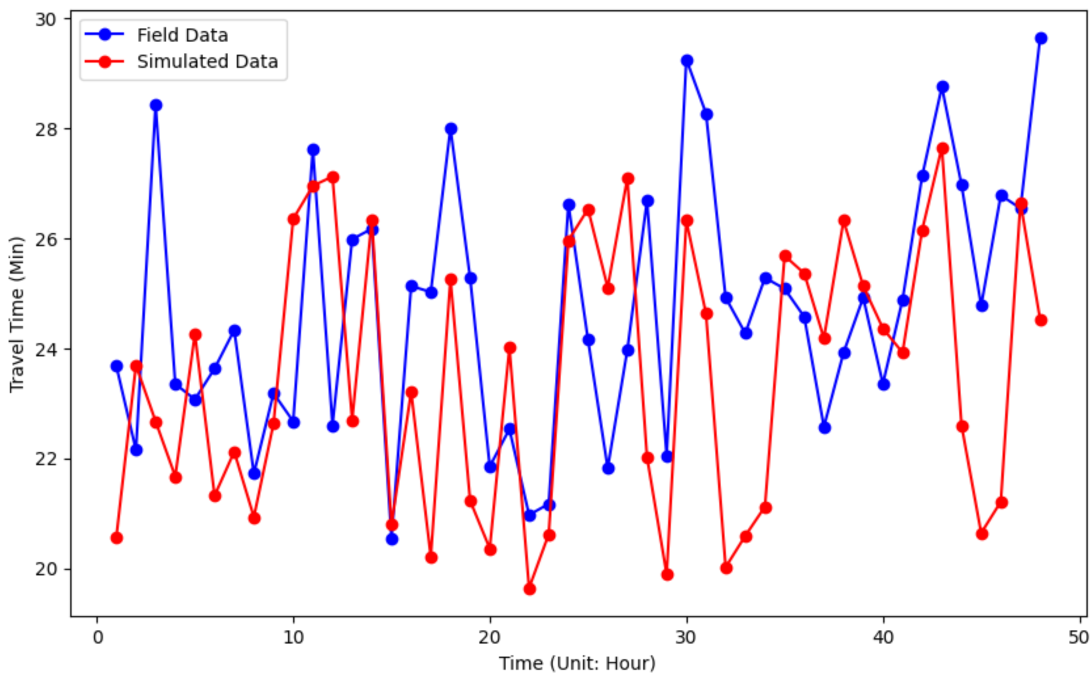


Figure 6-3: Comparison of time series of travel time from field data and simulation result

6.4 Determination of Capacity and Free Flow Time

The calibrated and validated model is used to ascertain both the road capacity and the free flow time of vehicles. To conduct further analysis, only the Dhankhola to Bhaluwang direction is considered as through lane, while the Bhaluwang to Dhankhola direction considered as opposing lane. The vehicles are categorized into two groups: light vehicles and heavy vehicles. This categorization is done because the study seeks to establish a relationship between the impact of heavy vehicles on the travel time of light vehicles. Light vehicles comprise passenger cars, utility vehicles, microbuses, and four-wheel-drive vehicles. Heavy vehicles encompass multi axle trucks, heavy trucks, light trucks, big buses, and minibuses.

Figure 6-4 illustrates the speed-flow correlation of passenger cars on the Dhankhola to Bhaluwang road segment. The graph depicts that the capacity for the Dhankhola to Bhaluwang direction amounts to 1489 PCU/hr./lane. As the road is a two-lane two-way thoroughfare, it can be inferred that the road's overall capacity reaches 2978 PCU/hr.

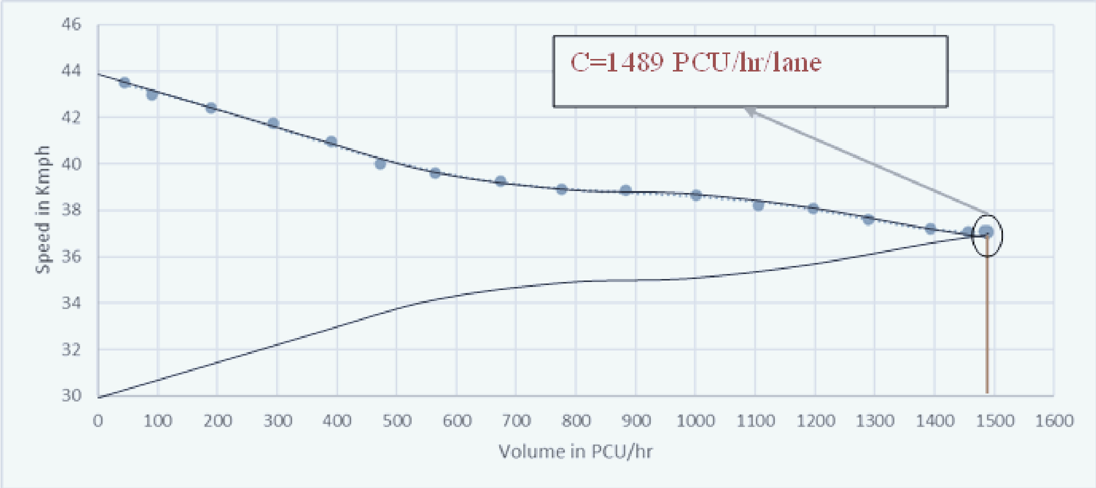


Figure 6-4: Relationship between speed and flow

Figure 6-5 provides free flow travel times in minutes for different vehicle types. Light vehicles have a travel time of 17.787 minutes and heavy vehicles require 20.701 minutes for their travel from Dhankhola to Bhaluwang.

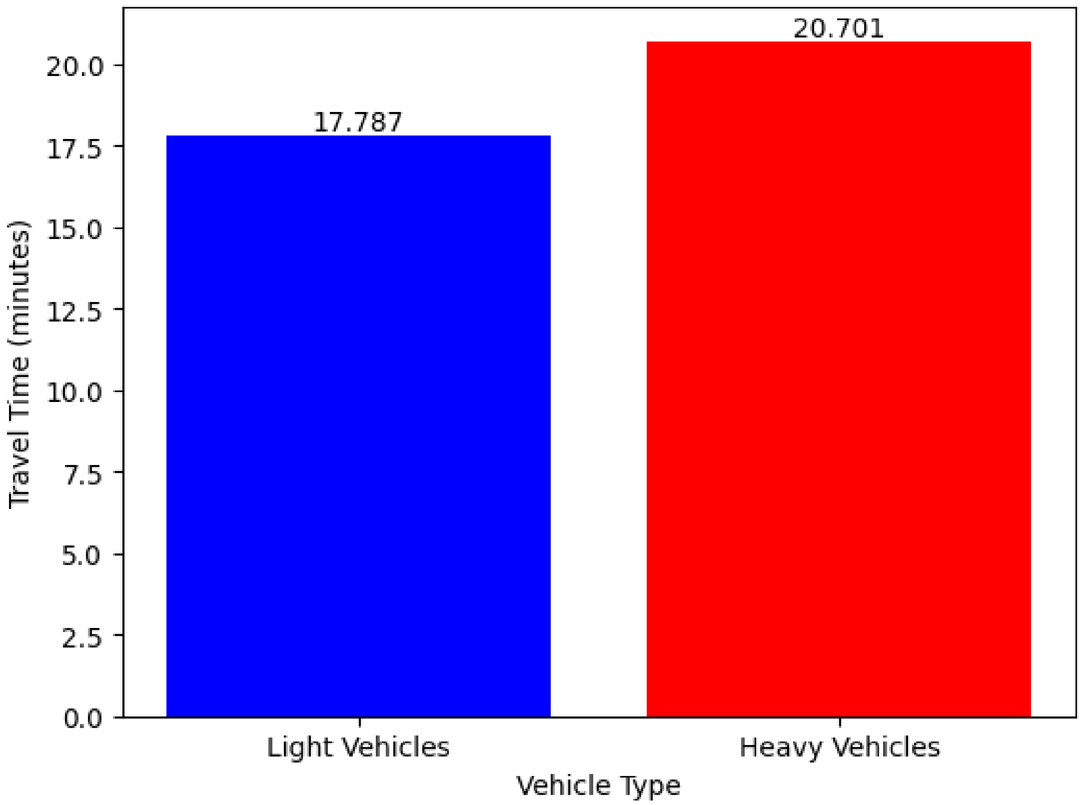


Figure 6-5: Free flow travel time

6.5 Regression Analysis of Travel Time Functions

A simulation matrix comprising 1442 unique scenarios was developed, each characterized by varying percentages of heavy vehicles in through traffic, through traffic volume, and opposing traffic volume. These scenarios were individually simulated using a previously calibrated and validated VISSIM model. For each simulation, the travel times of both light and heavy vehicles were recorded. Consequently, we compiled a 1442x5 matrix containing 1442 rows, with each row consisting of five columns representing travel time of light vehicles, travel time of heavy vehicles, the percentage of heavy vehicles in through traffic, through traffic volume, and opposing traffic volume for each vehicle type. The data obtained from simulation results is presented in APPENDIX-F.

Furthermore, Eq. 2.3 and Eq. 2.5, which is adapted BPR function designed for mixed traffic conditions in freeway as proposed by Lu, Meng and Gomes in 2016 (Lu, Meng, & Gomes, 2016) has been further refined for undivided two lane two way highway to account for the interaction with opposing traffic volume.

For travel function for heavy vehicle, two scenarios have been examined. In Case I, the analysis does not consider the percentages of heavy vehicles in through traffic which is predominately used for travel time function for heavy vehicle in multilane highway or freeway in past studies. However, in Case II, the analysis takes into consideration the relationship with percentages of heavy vehicles in through traffic, which could impact the travel time of heavy vehicles on a two-lane, two-way highway where overtaking maneuvers occur in opposing lanes. Regression techniques were applied to calibrate the parameters shown in Eq. 6.1, Eq. 6.2 and Eq. 6.3 in Python using Scikit-Learn and Statsmodels libraries.

For light vehicles,

$$t_l = t_l^0 * \left(1 + a * (1+\rho_1)^b * [\left(\frac{Q_T}{y}\right)^c + \left(\frac{Q_0}{y}\right)^d]\right) \qquad \textbf{6.1}$$

For heavy vehicles, (Case I)

$$t_h = t_h^0 * \left(1 + a * [\left(\frac{Q_T}{y}\right)^c + \left(\frac{Q_0}{y}\right)^d]\right) \qquad \textbf{6.2}$$

For heavy vehicles, (Case II)

$$t_h = t_h^0 * \left(1 + a * (1+\rho_1)^b * [\left(\frac{Q_T}{y}\right)^c + \left(\frac{Q_0}{y}\right)^d]\right) \qquad \textbf{6.3}$$

Where, $t_{l,0}$= Average link travel time of light vehicles

$t_{h,0}$= Average link travel time of heavy vehicles

t_l^0= Free flow speed of heavy vehicles

t_h^0= Free flow speed of heavy vehicles

ρ_1 = Percentage of heavy vehicles in through traffic

Q_T, Q_O and y are the through traffic volume, opposing traffic volume and capacity of one direction link in PCU/hr. respectively.

a, b, c and d are parameters used to calibrate the function.

The results of the regression, which calibrate the parameter values, have been shown in Table 6-2 for the above equations.

Table 6-2: Calibrated constant parameters in regression analysis

Level	Parameters	Light Vehicles	Heavy Vehicles	
			Case I	Case II
Optimum	a	0.258	0.237	0.172
	b	0.674	-	0.837
	c	1.151	0.846	0.853
	d	0.380	0.072	0.072
Lower 95% Confidence Level	a	0.252	0.234	0.171
	b	0.629	-	0.820
	c	1.067	0.773	0.827
	d	0.353	0.054	0.065
Upper 95% Confidence Level	a	0.264	0.241	0.174
	b	0.719	-	0.854
	c	1.236	0.054	0.880
	d	0.406	0.090	0.078

The travel time functions for light vehicles and heavy vehicles can be expressed as in Eq. 6.4, Eq. 6.5 and Eq. 6.6.

For light vehicles,

$$t_l = t_l^0 * \left(1 + 0.258 * (1 + \rho_1)^{0.674} * [\left(\frac{Q_T}{y}\right)^{1.151} + \left(\frac{Q_0}{y}\right)^{0.380}]\right) \quad \textbf{6.4}$$

For heavy vehicles, (Case I)

$$t_h = t_h^0 * \left(1 + 0.237 * [\left(\frac{Q_T}{y}\right)^{0.846} + \left(\frac{Q_0}{y}\right)^{0.072}]\right) \quad \textbf{6.5}$$

For heavy vehicles, (Case II)

$$t_h = t_h^0 * \left(1 + 0.172 * (1 + \rho_1)^{0.837} * [\left(\frac{Q_T}{y}\right)^{0.853} + \left(\frac{Q_0}{y}\right)^{0.072}]\right) \quad \textbf{6.6}$$

To measure how well the travel time functions model fit the VISSIM simulation data, a number of goodness-of fit statistics and evaluation metrics are evaluated and shown in Table 6-3. Firstly, a high value of coefficient of determination R^2 for light vehicles in Eq. 6.4, (R^2=0.84), illustrated the high predictive power of functions. The R^2 for heavy vehicle by Eq. 6.6 is greater than by Eq. 6.5 indicates the extent to which travel times of heavy vehicles are influenced by the percentage of heavy vehicles in through traffic. Secondly, the low MAPE values reveals that the models used to estimate the travel time of various vehicles have a high level of accuracy, and the high values of F statistics also indicate the function are statistically significant.

Table 6-3: Goodness-of-fit statistics and evaluation metrics of travel time functions

Vehicle Category	F-Statistics	R^2	MAPE	RMSE	MSE	MAE
Light Vehicles Eq. 6.4	121.30	0.84	3.45%	1.091969	1.192397	0.794401
Heavy Vehicles (Case I) Eq. 6.5	1842.00	0.58	3.74%	1.151102	1.325037	1.043213
Heavy Vehicles (Case II) Eq. 6.6	680.60	0.94	1.21%	0.419425	0.175917	0.339819

6.6 Comparison with Standard BPR Functions

To highlight the significance of proposed modified travel time function for the selected road type, further calibration was performed on the standard BPR function shown in Eq. 2.1 proposed in (Manual, 1964) with using simulated data. Eq. 6.7 and Eq. 6.8 are the calibrated standard BPR functions.

For light vehicles,

$$t_l = t_l^0 * \left(1 + 0.592 \left(\frac{Q_T}{C}\right)^{0.532}\right) \quad \textbf{6.7}$$

For heavy vehicles

$$t_l = t_h^0 * \left(1 + 0.444 \left(\frac{Q_T}{C}\right)^{0.281}\right) \quad \textbf{6.8}$$

Table 6-4 shows the goodness-of-fit statistics and evaluation metrics of BPR functions. Upon comparing with Table 6-3, it can be concluded that the proposed modified travel time functions outperform the standard BPR functions. For light vehicles, R2 increases from 0.72 in the standard travel time function to 0.84 in the modified travel time function. Similarly, for heavy vehicles, the value increases from 0.56 to 0.94. The values of MAPE, RMSE, MSE, and MAE are also reduced in the modified travel time functions compared to the standard BPR functions, indicating the significance of the modified travel time function for the selected road type.

Table 6-4: Goodness-of-fit statistics and evaluation metrics of standard BPR function

Vehicle Category	F-Statistics	R^2	MAPE	RMSE	MSE	MAE
Light Vehicles Eq. 6.7	98.00	0.72	4.8%	1.4422	2.0801	1.17015
Heavy Vehicles Eq. 6.8	1802.00	0.56	3.73%	1.182225	1.397657	1.04165

6.7 Validation of Travel Time Functions

Further validation of the calibrated travel time functions was performed by predicting travel times for diverse vehicles and comparing them with new simulation data.

Consequently, various traffic compositions (ρ_1 = 0.25), spanning a range of through volumes (from Q_T = 150 PCU/hr. to 1050 PCU/hr. in increments of 100 PCU/hr.), as well as opposing volumes (Q_O = 350 PCU/hr.), were employed. These volumes had not been used during the development of the travel time functions. Figure 6-6 shows the performance of travel time function for light vehicles. The mean absolute percentage error is 4.15%, indicating a high model accuracy. In the case of heavy vehicles, Case II (MAPE = 0.95%) outperforms Case I (MAPE = 3.63%) shown in Figure 6-7 and Figure 6-8 respectively, indicating that the percentage of heavy vehicles in the through traffic significantly influences the travel time of heavy vehicles on a two-lane, two-way road. The simulation data for the chosen scenario aligns within the 95% prediction intervals of the regression travel time functions for all cases, as illustrated in Figure 6-6, Figure 6-7 and Figure 6-8. These results indicate that the capability of travel time functions in predicting travel time is within the acceptance prediction interval.

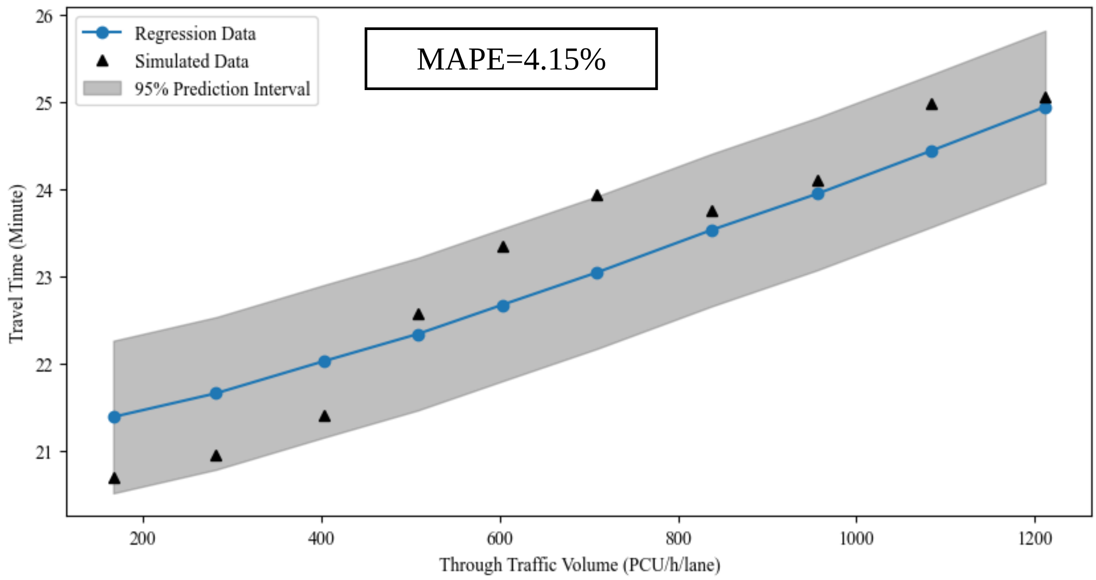


Figure 6-6: Validation of proposed travel time function of light vehicles

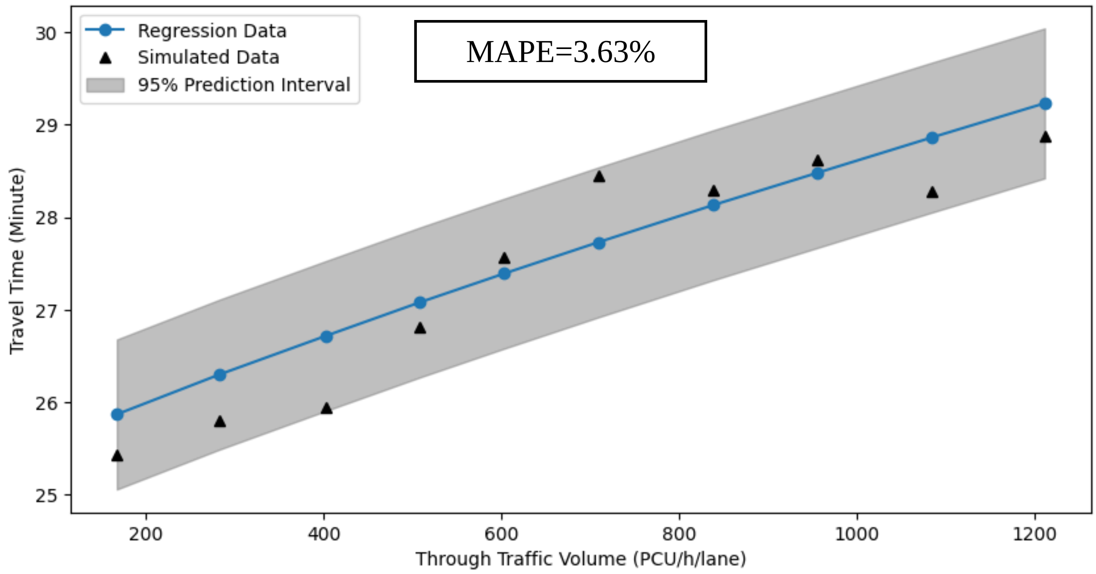


Figure 6-7: Validation of proposed travel time function of heavy vehicles- case I

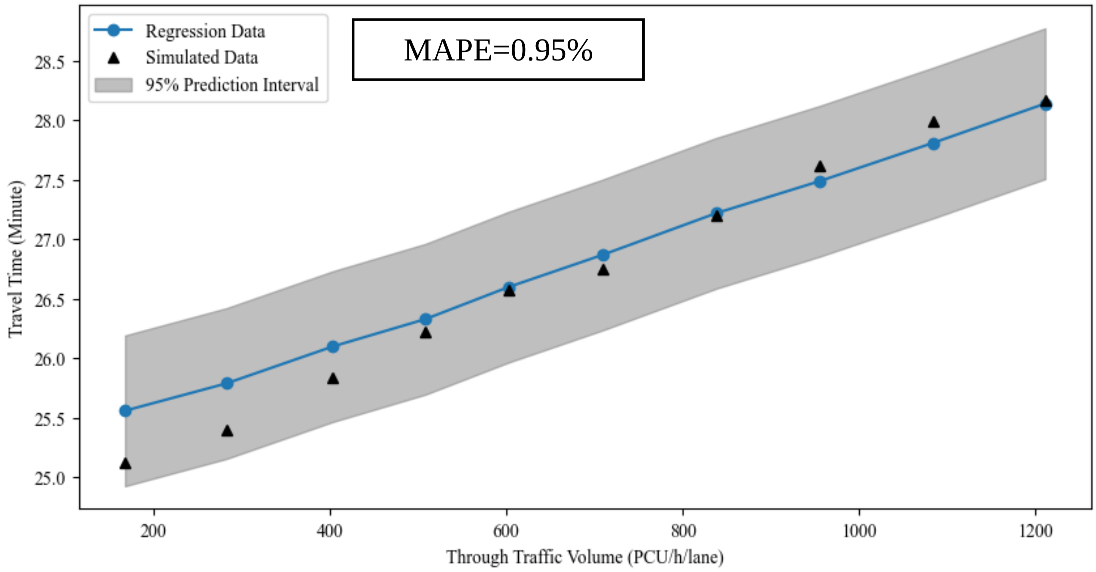


Figure 6-8: Validation of proposed travel time function of heavy vehicles- case II

CHAPTER 7: CONCLUSION AND RECOMMENDATION

7.1 Conclusion

In this research, a comprehensive analysis of 72- hour data sets on the travel time of vehicles were conducted using the data gathered from traffic volume count and speed survey. The primary objective of the study was to compare the performance of travel time prediction models in predicting travel time of Dhankhola-Bhaluwang road section, an undivided highway, relating with through traffic, opposing traffic and proportional of heavy vehicle in through traffic for both directions. Furthermore, this study has developed a microscopic traffic simulation to estimate travel time functions of heterogenous traffic flows on a two-lane two-way highway. The travel time functions were developed to formulate travel time for light vehicles and heavy vehicles using simulation model. The key findings emerged from this study is;

- There was a significant correlation between travel time and the independent variables, which included through traffic volume, opposing traffic volume, and the proportion of heavy vehicles in through traffic.
- The multiple regression model demonstrated a moderate relationship in the variance of travel time and the independent variables, with an R^2 of 0.44 and 0.45 for Dhankhola to Bhaluwang and Bhaluwang to Dhankhola direction respectively. Similarly, the MAPE was found to be 3.68% for the Dhankhola to Bhaluwang direction and 5.58% for the Bhaluwang to Dhankhola direction.
- The evaluation metrics obtained from Random Forest Regression outperform those of other regression models for both directions, followed by Support Vector Regression, Decision Tree Regression, LASSO Regression, and Multiple Linear Regression.
- In Dhankhola to Bhaluwang direction, Random Forest Regression results highest R^2 value (0.62) and lowest MAPE (2.48%) than other models.
- In Bhaluwang to Dhankhola direction, Random Forest Regression results highest R^2 value (0.68) and lowest MAPE (3.92%) than other models.
- The developed modified travel time functions yield better results than the standard BPR functions.

- The obtained travel time function for light vehicles yields an R^2 of 0.84 and a MAPE of 3.45%. Similarly, for the travel time functions of heavy vehicles, in Case I, the R^2 is 0.58 with a corresponding MAPE of 3.74%, and in Case II, the R^2 is 0.94 with a MAPE of 1.21%.
- The travel time function of a heavy vehicle in case II shows better statistical results compared to case I. Therefore, it can be concluded that the proportion of heavy vehicles significantly impacts their own travel time.

7.2 Recommendation

This study demonstrates the potential of various prediction models and simulation based-methodology for predicting travel time in two-lane two-way undivided carriageway road. However, it is important to note that there are several limitations in this study that could affect the predictions, and these limitations are considered as recommendations for future research.

- Conducting individual analyses for morning, afternoon, evening, and night data could lead to improved insights, as visibility may have varying impacts on a two-lane, two-way undivided road during different times of the day.
- 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 in the case of unrestricted two-lane, two-way undivided roads.
- The entire road is considered as a single segment for analysis. However, dividing it into multiple segments and studying them individually may lead to better accuracy and provide insights into the properties associated with each segment.
- Given the high proportion of two-wheeler traffic, including two-wheeler data in the simulation model might further improve the accuracy of the travel time functions.
- The proposed simulation-based methodology for modifying the travel time function can be applied to other roads of a similar type to validate the universality of the proposed travel time functions for the selected road type.

REFERENCES

Administration, F. H. (2010). *Travel Time Reliability.* U.S. Department of Transportation.

Afandizadeh, S., Nasiri, N., & Mirzahossein, H. (2021). 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.

Ahmed, U., Drakopoulos, A., & Ng, M. (2013). Impact of heavy vehicles on freeway operating characteristics under congested conditions. *Transportation Research Record: Journal of the Transportation Research Board, 2396*(1), 28-37. doi:10.3141/2396-04

Asaithambi, G., & Shravani, G. (2017). 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. doi:10.1016/j.jtte.2017.05.004

Billings, D., & Yang, J.-S. (2006). Application of the Arima models to urban roadway travel time prediction - A case study. *2006 IEEE International Conference on Systems, Man and Cybernetics*, 2529-2534.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32. doi:10.1023/a:1010933404324

Chien, S. I.-J., & Kuchipudi, C. M. (2003). Dynamic travel time prediction with real-time and historic data. *Journal of Transportation Engineering, 129*(6), 608-616.

D’Angelo, M. P., Al-Deek, H. M., & C., M. W. (1999). Travel-Time Prediction for Freeway Corridors. *Transportation Research Record, 1676*(1), 184-191.

Fan, W., & Gurmu, Z. (2015). Dynamic travel time prediction models for buses using only GPS Data. *International Journal of Transportation Science and Technology*, 353-366.

Gao, C., Xu, J., Jia, X., & Dong, Y. (2020). Influence of large vehicles on the speed of expressway traffic flow. *Advances in Civil Engineering*, 1-9. doi:10.1155/2020/2454106

Gao, Z., & Zhang, N. (2017). A new relationship between link flow and link travel time. *Proceedings of the 2017 5th International Conference on Frontiers of Manufacturing Science and Measuring Technology (FMSMT 2017)*, 628-634.

Highway Management Information System (HMIS). (n.d.). Retrieved from http://ssrn.aviyaan.com/traffic_controller

Hou, Y., Carlos, C., & Edara, P. (2012). Statistical test for 85th and 15th percentile speeds with asymptotic distribution of sample quantiles. *Transportation Research Record: Journal of the Transportation Research Board, 2279*(1), 47-53. doi:10.3141/2279-06

Kleynhans, T., Montanaro, M., Gerace, A., & Kanan, C. (2017). Predicting top-of-atmosphere thermal radiance using MERRA-2 atmospheric data with Deep Learning. *9*(11), 1133. doi:10.3390/rs9111133

L., N. J., & Turochy, R. E. (2019). Calibration of vissim models for Rural Freeway Lane closures: Novel approach to the modification of key parameters. *Transportation*

Research Record: Journal of the Transportation Research Board, 2673(5), 574-583. doi:10.1177/0361198119842824

Liu, Y., Wang, Y., Yang, X., & Zhang, L. (2017). Short-term travel time prediction by deep learning: A comparison of different LSTM-DNN models. *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 1-8.

Lu, Z., Meng, Q., & Gomes, G. (2016). Estimating link travel time functions for heterogeneous traffic flows on freeways. *Journal of Advanced Transportation, 50*(8), 1683-1698.

Manual, T. (1964). Bureau of Public Roads, US Dept. *Commerce, Urban Planning Division, Washington, DC, USA*.

Mehar, A., Chandra, S., & Velmurugan, S. (2014). Highway capacity through Vissim calibrated for mixed traffic conditions. *KSCE Journal of Civil Engineering, 18*(2), 639-645. doi:10.1007/s12205-014-0440-3

Meng, Q., & Weng, J. (2011). An improved cellular automata model for heterogeneous work zone traffic. *Transportation Research Part C: Emerging Technologies, 19*(6), 1263-1275. doi:10.1016/j.trc.2011.02.011

Murni, Kosasih, R., Fahrurozi, A., Handhika, T., Sari, I., & Lestari, D. P. (2020). Travel time estimation for destination in Bali using KNN-regression method with tensorflow. *IOP Conference Series: Materials Science and Engineering, 854*(1), 012061.

Oh, S., Byon, Y.-J., Jang, K., & Yeo, H. (2015). Short-term travel-time prediction on highway: A review of the data-driven approach. *Transport Reviews, 35*(1), 4-32.

Roh, C.-G., Park, B.-J., & Kim, J. (2017). 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.

Rose, G., Taylor, M. A., & Tisato, P. (1989). Estimating travel time functions for urban roads: Options and issues. *Transportation Planning and Technology, 14*(1), 63-82. doi:10.1080/03081068908717414

S.M.P., S., & Ramadurai, G. (2013). Calibration of vissim for Indian heterogeneous traffic conditions. *Procedia - Social and Behavioral Sciences, 104*, 380-389. doi:10.1016/j.sbspro.2013.11.131

Sharma, P., Singh, P., & Upteti, P. (2021). A Knowledge Based Travel Time Prediction using Regression Technique. *International Journal of Innovative Science and Research Technology, 6*(7), 724-727.

Siddharth, S., & Ramadurai, G. (2013). Calibration of vissim for Indian heterogeneous traffic conditions. *Procedia - Social and Behavioral Sciences, 104*, 380-389. doi:10.1016/j.sbspro.2013.11.131

Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic Medical Sonography, 6*(1), 35-39. doi:10.1177/875647939000600106

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1). doi:10.1111/j.2517-6161.1996.tb02080.x

Tu, H., Van Lint, J. W., & Van Zuylen, H. J. (2007). Impact of traffic flow on travel time variability of freeway corridors. *Transportation Research Record: Journal of the Transportation Research Board, 1993*(1), 59-66. doi:10.3141/1993-09

Wu, C.-H., Ho, J.-M., & Lee, D. (2004). Travel-time prediction with support vector regression. *IEEE Transactions on Intelligent Transportation Systems, 5*(4), 276-281.

Wu, J. H., Florian, M., & He, S. (2006). An algorithm for multi-class network equilibrium problem in PCE of trucks: Application to the SCAG Travel Demand Model. *Transportmetrica, 2*(1), 1-9.

Xu, M., Watanachaturaporn, P., Varshney, P., & Arora, M. (2005). Decision tree regression for soft classification of Remote Sensing Data. *Remote Sensing of Environment, 97*(3), 322-336. doi:10.1016/j.rse.2005.05.008

Yang, S., & Qian, S. (2019). 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.

Yi, H., Jung, H., & Bae, S. (2017). Deep neural networks for traffic flow prediction. *2017 IEEE International Conference on Big Data and Smart Computing*, 328-331.

Zhang, X., & Rice, J. A. (2003). Short-term travel time prediction. *Transportation Research Part C: Emerging Technologies, 11*(3-4), 187-210.

APPENDIX- A

TRAFFIC VOLUME COUNT

APPENDIX-B

SPOT SPEED DATA

APPENDIX-C

TRAVEL TIME DATA

APPENDIX-D

REGRESSION DATA

APPENDIX-E

REGRESSION CODES

APPENDIX-F

SIMULATION RESULTS DATA