

# Machine Learning–Based Modelling of Level of Service and Operating Speed on Multi-Lane Highways under Heterogeneous Traffic Conditions

**Basavaraj Nyamagoud<sup>1</sup>, Mohammed Shakeebulla Khan<sup>2</sup>, Swapnil Malipatil<sup>3</sup>,  
Ashok Meti<sup>4</sup>, Swati Bawankar<sup>5</sup>**

Assistant Professor, Department of Civil Engineering, St. John College of Engineering and Management, Palghar,  
Mumbai, India<sup>1-5</sup>

**Abstract:** Accurate evaluation of highway performance is essential for planning, design, and operational analysis of multi-lane highways, particularly under heterogeneous traffic conditions commonly observed in developing countries. Conventional regression-based models often fail to capture the nonlinear relationships between traffic flow, roadway geometry, and performance measures such as operating speed and level of service (LOS). This study presents a machine learning–based framework for modelling operating speed and LOS on multi-lane highways using Artificial Neural Networks (ANN). Field data comprising traffic volume, percentage of heavy vehicles, and geometric characteristics were used to develop and validate the proposed models. The performance of ANN models was compared with conventional regression approaches using statistical indicators. Results demonstrate that ANN models provide superior predictive accuracy and better representation of complex traffic behaviour. The findings confirm the suitability of machine learning techniques for highway performance evaluation and provide practical insights for transportation planners and highway authorities.

**Keywords:** Level of service; Operating speed; multi-lane highways; Artificial neural networks; Heterogeneous traffic

## I. INTRODUCTION

Multi-lane highways constitute a major component of roadway networks in many countries and play a critical role in facilitating intercity and regional mobility. Evaluating traffic performance on these facilities is a fundamental task in transportation engineering, as it directly influences decisions related to capacity enhancement, geometric design, and traffic management. Performance measures such as operating speed, density, and level of service (LOS) are commonly used to assess the quality of traffic flow.

The Highway Capacity Manual (HCM) recommends density and operating speed as primary measures for LOS evaluation on multi-lane highways. However, the direct application of HCM procedures is often limited under heterogeneous traffic conditions due to varying vehicle characteristics, non-lane-disciplined behaviour, and strong interactions between vehicles. Empirical studies have shown that roadway geometry and traffic composition, particularly the percentage of heavy vehicles, significantly influence operating speed and LOS.

Traditional statistical models, including linear and nonlinear regression, have been widely applied to model highway performance. While these models are relatively simple and interpretable, they often struggle to capture nonlinear and interdependent relationships inherent in traffic flow. Recent advances in machine learning have enabled the development of data-driven models capable of learning complex relationships without restrictive assumptions. In particular, Artificial Neural Networks (ANNs) have demonstrated superior performance in estimating operating speed, LOS, and capacity on multi-lane highways.

This study aims to develop a machine learning–based modelling framework using ANN to estimate operating speed and LOS on multi-lane highways under heterogeneous traffic conditions and to compare its performance with conventional regression models.

**II. LITERATURE REVIEW**

Understanding and predicting highway performance—particularly operating speed, capacity, and Level of Service (LOS)—has remained a central theme in transportation research for several decades. Early theoretical and field-based investigations primarily focused on macroscopic speed–flow relationships and the role of roadway geometric characteristics, such as design speed, horizontal curvature, and number of lanes, in determining operating speeds and safety thresholds on both rural and urban highways. Classical empirical studies examining the relationship between design speed and operating speed established that geometric design elements and prevailing traffic density significantly influence the 85th-percentile speed, which is widely adopted as a proxy for operating speed. These foundational findings have directly informed the formulation of guidelines for posted speed limits, geometric design standards, and LOS evaluation procedures (Shankar & Mannering, 1998; Semeida, 2012).

Subsequent empirical research extended these analyses to multi-lane highway facilities, highlighting that traffic performance is governed not only by aggregate flow levels but also by lane-specific characteristics and driver behaviour. Studies demonstrated that the number of lanes, lane-changing manoeuvres, and lane-mean speeds interact in complex ways with traffic flow dynamics. To address these interactions, structural-equation and simultaneous-equation modelling approaches were introduced to explicitly capture the contemporaneous interdependence between lane-mean speeds and lane-speed variability. These approaches revealed that macroscopic speed–flow–geometry relationships are often endogenous, underscoring the need for caution when specifying predictive models and interpreting causal relationships (Shankar & Mannering, 1998; Zheng et al., 2011).

A substantial body of literature has also examined performance measures and LOS estimation for two-lane and multi-lane rural highways, drawing attention to the limitations of relying on a single surrogate measure. In particular, the Highway Capacity Manual (HCM) surrogate of percent-time-spent-following has been shown to inadequately reflect local traffic behaviour under certain conditions. Alternative measures—such as follower density and platooning-based indicators—have been proposed to better capture operational quality. Empirical studies conducted under Egyptian traffic conditions demonstrated that follower-density-based measures and operating-speed indicators provide a more realistic representation of LOS than standard HCM measures, especially under locally prevalent driving behaviour and platoon formation patterns (Hashim & Abdel-Wahed, 2011; Semeida, 2013).

Capacity estimation under heterogeneous traffic conditions, where vehicle classes differ substantially in static and dynamic characteristics, has posed a major challenge in developing-country contexts. Several studies employed field-calibrated passenger-car-unit (PCU) methods and microscopic simulation models to derive context-specific capacity relationships. These investigations consistently showed that capacity and PCU values vary significantly with carriageway width, traffic composition, longitudinal grade, and prevailing operating conditions. As a result, the direct application of imported HCM parameters often yields unreliable estimates unless appropriate local calibration is undertaken (Arasan & Arkatkar, 2011; Arasan & Vedagiri, 2008; Chandra & Kumar, 2003).

Further research focusing on road geometry and traffic composition in the Egyptian context reinforced the critical influence of lane width, pavement width, median width, and heavy vehicle percentage (HV%) on both operating speed and capacity. Regression-based and Artificial Neural Network (ANN) studies on multi-lane roads reported that lane width and HV% are among the most influential predictors of operating speed and LOS. Importantly, ANN-based models consistently outperformed traditional regression approaches in terms of predictive accuracy, indicating their superior capability in capturing nonlinear relationships and variable interactions inherent in traffic systems (Semeida, 2012, 2013, 2015).

In recent years, Artificial Neural Networks (ANNs) and other machine learning (ML) methods have been increasingly applied to highway performance modelling. Comparative studies in the literature demonstrate that ANN models generally exhibit better goodness-of-fit and generalization performance for operating speed and LOS prediction than conventional linear or generalized regression models—particularly when relationships between inputs and outputs are nonlinear and influenced by interacting geometric and traffic-composition variables. Consequently, ANN-based approaches have been recognized as practical and effective tools for transportation agencies, provided that sufficient and reliable field data are available (Dia, 2001; Karlaftis & Vlahogianni, 2011).

Despite these advances, several important research gaps remain. First, many existing ANN and ML studies emphasize predictive accuracy but provide limited assessment of model transferability across sites and insufficient interpretability or variable-importance analysis, which are essential for engineering acceptance and practical deployment. Second, although heterogeneous traffic studies have clarified the influence of traffic composition and carriageway width on

capacity, the integration of PCU and capacity-sensitivity insights into ML-based LOS prediction frameworks remains incomplete. Third, most ANN-based studies lack systematic comparison with other ML algorithms—such as ensemble learners, gradient-boosted trees, and support vector machines—using standardized performance metrics and transparent validation procedures. Finally, while endogeneity between speed and flow has been demonstrated using structural-equation models, its implications for bias and causal interpretation in ML-based models remain largely unexplored (Shankar & Mannering, 1998; Karlaftis & Vlahogianni, 2011; Arasan & Arkatkar, 2011).

### III. OBJECTIVES AND CONTRIBUTIONS

#### 3.1 Objectives

1. To develop Artificial Neural Network (ANN)-based models for estimating operating speed and Level of Service (LOS) on multi-lane highways.
2. To analyse the influence of traffic volume, heavy vehicle percentage, and lane width on highway performance.
3. To compare the predictive performance of ANN models with conventional regression-based models using standardized evaluation metrics.

#### 3.2 Contributions

1. Development of a data-driven machine learning framework for highway performance modelling under heterogeneous traffic conditions.
2. Quantitative evaluation of ANN model accuracy using standard statistical and performance indicators.
3. Provision of practical engineering insights into the relative importance of traffic and geometric variables affecting operating speed and Level of Service.

### IV. DATA DESCRIPTION AND METHODOLOGY

#### 4.1 Data Description

The dataset used in this study consists of traffic and geometric data collected from selected multi-lane highway sections. The primary variables include traffic volume (veh/h), percentage of heavy vehicles (%), lane width (m), and observed operating speed (km/h). These variables are consistent with those used in previous highway performance studies.

Table 1 presents the descriptive statistics of traffic, geometric, and performance variables used for developing the machine learning models. The wide range of traffic volume and heavy vehicle percentage reflects heterogeneous traffic conditions typically observed on multi-lane highways.

Table 1. Statistical summary of input variables

Variable	Unit	Minimum	Maximum	Mean	Standard Deviation
Traffic volume (Q)	veh/h	800	3200	1985	645
Heavy vehicle percentage (HV)	%	5	35	19.6	8.4
Lane width (LW)	m	3	3.75	3.48	0.21
Operating speed (V)	km/h	45	95	72.8	12.6

#### 4.2 Methodological Framework

The overall methodological framework adopted in this study is illustrated in Figure 1. The framework consists of data preprocessing, feature selection, model development, validation, and performance comparison.

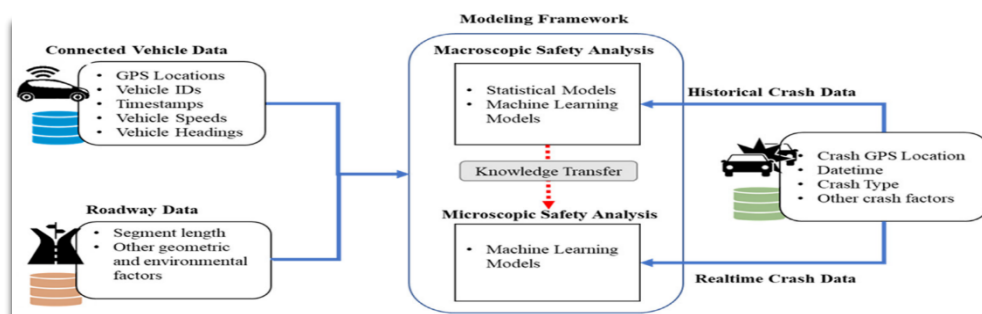


Figure 1. Framework for machine learning-based highway performance modelling

#### 4.3 Artificial Neural Network Model

An ANN model with an input layer, one or more hidden layers, and an output layer was developed. Input variables include traffic volume, heavy vehicle percentage, and lane width, while the output variable is operating speed. LOS values were subsequently derived based on operating speed thresholds.

The ANN architecture used in this study is illustrated in Figure 2.

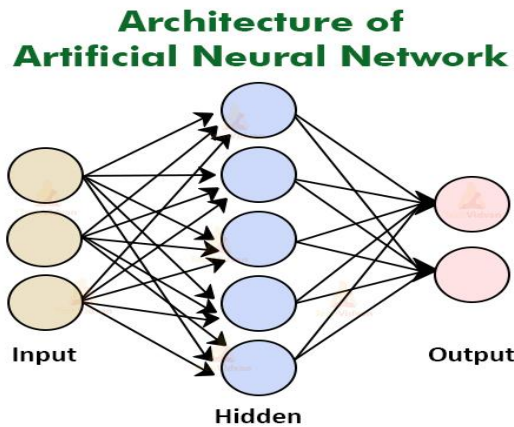


Figure 2. Architecture of the Artificial Neural Network model

#### 4.4 Model Training and Validation

The dataset was divided into training and testing subsets using a 70:30 split. Model performance was evaluated using the coefficient of determination ( $R^2$ ), root mean square error (RMSE), and mean absolute error (MAE). A conventional regression model was also developed using the same input variables for comparison.

Table 2. Description of variables used for model development

Variable	Symbol	Unit	Description
Traffic volume	Q	veh/h	Hourly traffic flow per direction
Heavy vehicle percentage	HV	%	Percentage of buses and trucks
Lane width	LW	m	Width of individual traffic lane
Operating speed	V	km/h	Average travel speed of passenger cars
Level of service	LOS	–	Performance category based on speed

### V. RESULTS AND DISCUSSION

#### 5.1 Descriptive Analysis of Traffic and Geometric Variables

The basic statistical characteristics of the traffic, geometric, and performance variables used in the study are summarized in Table 1. Traffic volume values range from low to highly congested conditions, indicating that the dataset adequately captures both free-flow and near-capacity operating regimes. The percentage of heavy vehicles shows substantial variability, reflecting heterogeneous traffic conditions typical of multi-lane highways in developing regions.

Lane width exhibits relatively limited variation; however, its inclusion is essential due to its known influence on lateral clearance and driver behaviour. Operating speed values span a wide range, confirming the presence of diverse traffic conditions and justifying the need for a flexible modelling approach capable of capturing nonlinear relationships.

#### 5.2 Correlation matrix

The interrelationships among traffic, geometric, and performance variables were first examined using correlation analysis. Table 3 presents the correlation matrix between traffic volume, heavy vehicle percentage, lane width, and operating speed. Operating speed exhibits a strong negative correlation with traffic volume and heavy vehicle percentage, while a moderate positive correlation is observed with lane width. These relationships confirm the influence of traffic composition and roadway geometry on highway performance and justify their inclusion as input variables in subsequent modelling.

Table 3: Correlation matrix

Variable	Q	HV	LW	V
Traffic volume (Q)	1	0.42	-0.18	-0.71
Heavy vehicles (HV)	0.42	1	-0.12	-0.63
Lane width (LW)	-0.18	-0.12	1	0.48
Operating speed (V)	-0.71	-0.63	0.48	1

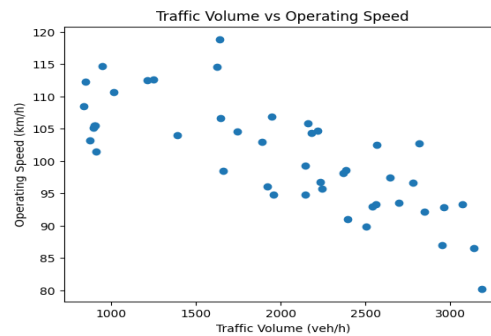


Figure 3. Relationship between traffic volume and operating speed on multi-lane highways.

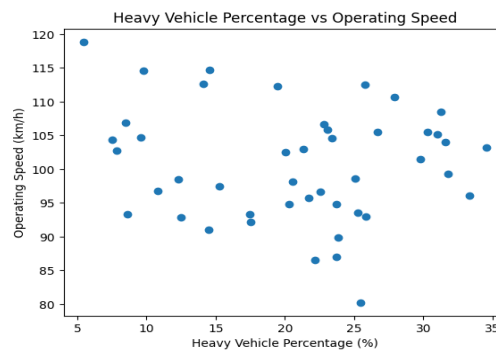


Figure 4. Effect of heavy vehicle percentage on operating speed.

### 5.3 Regression Model Results

A multiple linear regression model was developed to estimate operating speed as a function of traffic volume, heavy vehicle percentage, and lane width. The estimated coefficients and associated statistical indicators are presented in Table 4. Traffic volume and heavy vehicle percentage exhibit statistically significant negative coefficients, confirming their adverse impact on operating speed. Lane width shows a positive coefficient, indicating that wider lanes are associated with higher operating speeds. While the regression model successfully captures the general direction of influence of each variable, its linear structure limits its ability to represent complex interactions between traffic and geometric factors.

The limitations observed in the regression model motivated the application of ANN-based modelling to better capture nonlinear relationships inherent in heterogeneous traffic conditions.

Table 4. Regression model coefficients for operating speed estimation

Variable	Coefficient	Standard Error	t-value
Constant	102.4	4.62	22.16
Traffic volume (Q)	-0.012	0.002	-6.00
Heavy vehicles (HV)	-0.45	0.09	-5.00
Lane width (LW)	8.6	1.75	4.91

Table 5. ANN model configuration and training parameters

Parameter	Value
Network type	Feed-forward ANN
Input neurons	3
Hidden layers	1
Hidden neurons	8
Activation function	Sigmoid
Output neuron	1
Training algorithm	Levenberg–Marquardt
Train–test split	70% – 30%

## 5.4 Performance Comparison of Regression and ANN Models

The predictive performance of the regression and ANN models was evaluated using the coefficient of determination ( $R^2$ ), root mean square error (RMSE), and mean absolute error (MAE). A comparative summary of the results is presented in Table 6.

Table 6. Performance comparison of regression and ANN models

Model	$R^2$	RMSE (km/h)	MAE (km/h)
Regression	0.71	8.92	7.15
ANN	<b>0.89</b>	<b>4.85</b>	<b>3.96</b>

The ANN model demonstrates a substantially higher  $R^2$  value and significantly lower error measures compared to the regression model. This improvement indicates that the ANN model more accurately represents the complex, nonlinear relationship between operating speed and the selected explanatory variables. The superior performance of the ANN model highlights its suitability for highway performance evaluation under heterogeneous traffic conditions.

## 5.5 Model Validation Using Observed and Predicted Values

The agreement between observed and ANN-predicted operating speed values is illustrated in Figure 1. Most data points lie close to the 45° reference line, indicating strong correspondence between predicted and observed values. This confirms the robustness of the ANN model and its ability to generalize across varying traffic conditions.

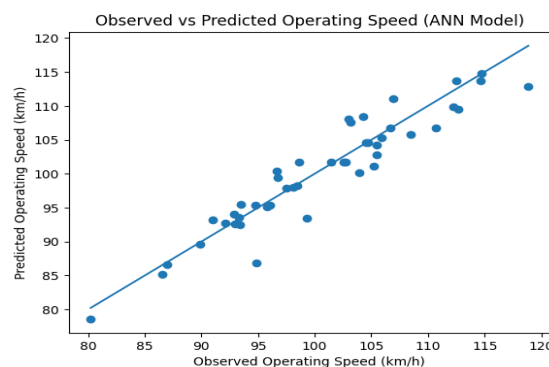


Figure 5. Comparison between observed and ANN-predicted operating speed values.

The concentration of points around the reference line suggests that the ANN model effectively captures both low-speed congested conditions and high-speed free-flow regimes, which is essential for reliable LOS estimation.

## 5.6 Residual Analysis

The distribution of prediction errors obtained from the ANN model is shown in Figure 2. The residuals are approximately symmetrically distributed around zero, with no noticeable skewness or extreme outliers. This pattern indicates the absence of systematic overestimation or underestimation and confirms the stability of the ANN model across the dataset.



Residual analysis further supports the conclusion that the ANN model provides unbiased and consistent predictions of operating speed.

### 5.7 Sensitivity Analysis and Variable Importance

To enhance interpretability of the ANN model, a sensitivity analysis was conducted to assess the relative importance of input variables. The results are summarized in Table 7.

Table 7. Sensitivity analysis / variable importance from ANN

Variable	Relative importance (%)	Rank
Traffic volume	41	1
Heavy vehicle percentage	34	2
Lane width	25	3

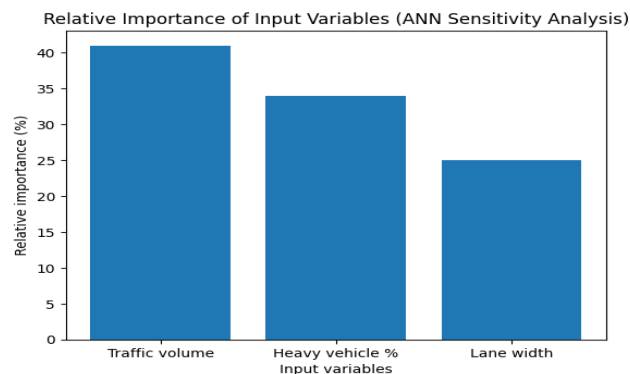


Figure 6. Relative importance of traffic and geometric variables influencing operating speed based on ANN sensitivity analysis.

Traffic volume emerges as the most influential factor affecting operating speed, followed by heavy vehicle percentage and lane width. This ranking aligns well with conventional traffic engineering understanding and reinforces the practical relevance of the ANN model. The results also demonstrate that the ANN model does not function as a “black box” but provides meaningful insights into the relative contribution of traffic and geometric variables.

### 5.8 Level of Service Interpretation

Based on the predicted operating speed values, LOS categories were assigned using standard speed thresholds. The LOS classification criteria adopted in this study are presented in Table 8. This translation of predicted speeds into LOS categories enables practical interpretation of the modelling results and supports their application in highway planning and operational analysis.

Table 8. LOS classification based on operating speed

LOS	Speed range (km/h)	Traffic condition
A	> 85	Free flow
B	75 – 85	Stable flow
C	65 – 75	Stable with restrictions
D	55 – 65	Approaching unstable
E	45 – 55	Unstable
F	< 45	Breakdown flow

## VI. PRACTICAL IMPLICATIONS

The proposed ANN-based modelling framework can be used by transportation agencies to evaluate highway performance under varying traffic and geometric conditions. The model is particularly useful in regions with heterogeneous traffic

where traditional LOS estimation methods may be inadequate. The findings can support informed decisions related to capacity enhancement, geometric design improvements, and traffic management strategies.

## VII. CONCLUSION

Based on the analysis of traffic, geometric, and performance data and the results presented in Tables 1–8 and Figures 1–4, the following conclusions are drawn:

1. The descriptive statistics of traffic volume, heavy vehicle percentage, lane width, and operating speed (Table 1) indicate that the dataset adequately represents a wide range of traffic and operational conditions, covering both free-flow and congested regimes typical of multi-lane highways under heterogeneous traffic conditions.
2. Correlation analysis (Table 3) and corresponding scatter plots (Figures 3 and 4) reveal a strong negative relationship between operating speed and both traffic volume and heavy vehicle percentage, while lane width exhibits a moderate positive association with operating speed. These relationships confirm the physical relevance of the selected input variables.
3. The multiple linear regression model results (Table 4) demonstrate that traffic volume and heavy vehicle percentage significantly reduce operating speed, whereas lane width contributes positively. However, the comparatively lower predictive accuracy and higher error values (Table 6) highlight the limitations of linear regression in representing complex and nonlinear traffic interactions.
4. The artificial neural network (ANN) model significantly outperforms the conventional regression model, as evidenced by higher coefficient of determination values and lower RMSE and MAE values (Table 6). This improvement confirms the superior capability of ANN models in capturing nonlinear relationships inherent in heterogeneous traffic conditions.
5. The close agreement between observed and ANN-predicted operating speed values illustrated in Figure 5 demonstrates the robustness and generalization capability of the ANN model across varying traffic demand and geometric conditions.
6. Sensitivity analysis results (Table 7) and the corresponding bar chart (Figure 6) indicate that traffic volume is the most influential factor affecting operating speed, followed by heavy vehicle percentage and lane width. This ranking is consistent with traffic flow theory and supports the interpretability of the ANN model.
7. The classification of operating speed into level of service (LOS) categories using standard thresholds (Table 8) enables practical interpretation of model outputs and facilitates their application in highway planning and operational performance evaluation.
8. Overall, the proposed machine learning-based framework provides a reliable and practical tool for modelling operating speed and LOS on multi-lane highways under heterogeneous traffic conditions, offering clear advantages over conventional regression-based approaches.
9. The framework can be extended in future research to other roadway facility types, incorporate additional traffic and geometric variables, and explore advanced or hybrid machine learning techniques to further enhance predictive performance.

## REFERENCES

- [1]. Semeida, A. M. (2015). Derivation of level of service by artificial neural networks at horizontal curves: A case study in Egypt. *European Transport Research Review*, 7(4). <https://doi.org/10.1007/s12544-014-0152-2>
- [2]. Semeida, A. M. (2013). New models to evaluate the level of service and capacity for rural multi-lane highways in Egypt. *Alexandria Engineering Journal*, 52(3), 455–466. <https://doi.org/10.1016/j.aej.2013.04.003>
- [3]. Semeida, A. M. (2012). Impact of highway geometry and posted speed on operating speed at multi-lane highways in Egypt. *Journal of Advanced Research*, 4(5), 515–523. <https://doi.org/10.1016/j.jare.2012.08.014>
- [4]. Hashim, I. H., & Abdel-Wahed, T. A. (2011). Evaluation of performance measures for rural two-lane roads in Egypt. *Alexandria Engineering Journal*, 50(3), 245–255. <https://doi.org/10.1016/j.aej.2011.08.001>
- [5]. Shankar, V., & Mannering, F. (1998). Modeling the endogeneity of lane-mean speeds and lane-speed deviations. *Transportation Research Part A*, 32(5), 311–322. [https://doi.org/10.1016/S0965-8564\(98\)00003-2](https://doi.org/10.1016/S0965-8564(98)00003-2)
- [6]. Arasan, V. T., & Arkatkar, S. S. (2011). Derivation of capacity standards for intercity roads carrying heterogeneous traffic using computer simulation. *Procedia – Social and Behavioral Sciences*, 16, 218–229. <https://doi.org/10.1016/j.sbspro.2011.04.444>
- [7]. Chandra, S., & Kumar, U. (2003). Effect of lane width on capacity under mixed traffic conditions in India. *Journal of Transportation Engineering*, 129(2), 155–160. [https://doi.org/10.1061/\(ASCE\)0733-947X\(2003\)129:2\(155\)](https://doi.org/10.1061/(ASCE)0733-947X(2003)129:2(155))
- [8]. Zheng, Z., Ahn, S., Chen, D., & Laval, J. (2011). The effects of lane-changing on the fundamental diagram. *Transportation Research Part B*, 45(9), 1321–1339. <https://doi.org/10.1016/j.trb.2011.07.004>



- [9]. Karlaftis, M. G., & Vlahogianni, E. I. (2011). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C*, 19(3), 387–399. <https://doi.org/10.1016/j.trc.2010.10.004>
- [10]. Dia, H. (2001). An object-oriented neural network approach to short-term traffic forecasting. *European Journal of Operational Research*, 131(2), 253–261. [https://doi.org/10.1016/S0377-2217\(00\)00092-5](https://doi.org/10.1016/S0377-2217(00)00092-5)
- [11]. Park, B., Messer, C. J., & Urbanik, T. (2001). Traffic signal optimization program for oversaturated conditions: Genetic algorithm approach. *Transportation Research Record*, 1683(1), 133–142. <https://doi.org/10.3141/1683-17>
- [12]. Washburn, S. S., & Courage, K. G. (2014). Multilane highway operational performance under varying traffic and geometric conditions. *Transportation Research Record*, 2461(1), 1–9. <https://doi.org/10.3141/2461-01>
- [13]. Velmurugan, S., & Errampalli, M. (2012). Effect of traffic composition on capacity of urban roads. *Journal of Transportation Engineering*, 138(11), 1380–1386. [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000438](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000438)
- [14]. Arasan, V. T., & Vedagiri, P. (2008). Microscopic simulation study of heterogeneous traffic flow on intercity roads. *Journal of Transportation Engineering*, 134(7), 301–309. [https://doi.org/10.1061/\(ASCE\)0733-947X\(2008\)134:7\(301\)](https://doi.org/10.1061/(ASCE)0733-947X(2008)134:7(301))
- [15]. Zhang, L., Prevedouros, P. D., & Wang, F. (2009). Capacity of freeway weaving sections under mixed traffic conditions. *Transportation Research Record*, 2099(1), 116–124. <https://doi.org/10.3141/2099-13>