



Data-driven approach for analysing the influence of crossing pedestrian on traffic characteristics at urban midblock sections

J. Athira ^{1*}, Sreechitra², G. Anju³, Yogeshwar V. Navandar⁴,
Hareshkumar Dahyabhai Golakiya⁵, Ashish Dhamaniya⁶

¹ Research Scholar, NIT Calicut, Kerala, India, 673601

² PG Student, NIT Calicut, Kerala, India,

³ PG Student, NIT Calicut, Kerala, India,

⁴ Assistant Professor, NIT Calicut, Kerala, India, 673601

⁵ Assistant Professor, Dr. S. and S. S. Ghandhy Government Engineering College, Surat, Gujarat, 395007, India

⁶ Associate Professor, Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat, 395007, India

Abstract

In developing countries like India, pedestrians crossing in the mid blocks is a common occurrence, and it has two major consequences from the traffic point of view. First, When there are pedestrian crossing the street, cars slow down because they are trying to prevent crashes, which lowers the overall capacity. Second, such pedestrian engagement in traffic flow may irritate drivers, resulting in a high number of accidents between pedestrians and vehicles. The current research focused on examining the impact of undesignated midblock pedestrian crossings on vehicular flow. The study attempts to quantify the impact of such pedestrian crossings on traffic parameters, such as vehicle capacity and speed, using data-driven approaches like Support vector regression (SVR), Locally weighted regression (LOESS), and Gaussian processes regression (GPR). Data is modelled using data-driven techniques, which show improved modelling performance. Information was gathered from midblock segments of six-lane urban arterial in many Indian towns, such as Surat, Baroda, Noida, and Udhana. Among which, Baroda Surat and Noida are taken as non-base sections where pedestrian crossflow is present and Udhana location is considered as base section where there is no pedestrian crossing. Videographic survey is used for collecting data, and extracted using Avidemux 2.6 software. With varying degrees of pedestrian crossflow, the speed of each vehicle category and the capacity of the base and non-base sections were estimated and validated against field data. MAPE and RMSE values were used to conduct a comparative study and relative evaluation of data-driven techniques and conventional theoretical pedestrian models. The results show that proposed data-driven approaches outperformed the conventional model for different vehicle categories.

Keywords: Pedestrian crossing, Speed, Data-driven techniques, Capacity.

* Corresponding author: J. Athira (athirajp94@gmail.com)

1. Introduction

Walking is a means of transportation that can be used in conjunction with other modes. For short journeys of 1-2 kilometres, walking is the ideal means of transportation. Between any particular trip origin and destination, walking serves as a feeder mode. Pedestrian facilities like crossing and walking are inadequate in underdeveloped countries like India. This results in the sharing of the roads by pedestrians with driving automobiles. Instead of using designated crossing facilities, pedestrians usually choose to cross the road at points that are advantageous for reaching different kinds of land uses. Crossings like this are forbidden. At at-grade facilities, zebra crossings are present, although moving traffic rarely yields to pedestrians. As a result, people usually choose to cross the road at undesignated sections. Figure 1 shows an undesignated pedestrian crossing. At these crossings, people and traffic flow interact. As a result, cars must change their course or slow down to avoid colliding with pedestrians. This causes the flow velocity to drop, which in turn reduces the capacity of some sections of the path. However, it is important to note that capacity is not as significant as speed in urban road section. This is because a road section might allow for relatively continuous flow and thus have a higher vehicle capacity, the intersections at both ends of the street section determine the maximum rate at which vehicles can actually proceed through the road network. The present study is conducted at urban midblock sections which is free from all mentioned side frictions except the undesignated pedestrian crossings. Undesignated roads are those that do not have grade-separated facilities, pedestrian signals, or zebra crossings. In order to examine how crossing pedestrians affect traffic dynamics in urban midblock sections, the study will make use of data-driven approaches. In the big data era, skilled data processing becomes essential to improving traffic model accuracy and reliability. In a future world where vehicles may link to each other and to infrastructure, all possible travel scenarios and modes must be represented.



Figure 1: Undesignated pedestrian crossings

The study is controlled to 1 base section and 3 non-base sections from 3 cities in India. Here, the non-base section refers to the midblock sections with pedestrian crossing only as a friction factor. The 3 non-base sections considered are Surat, Baroda, and Noida. Initially the study was planned for the prediction of the speed of vehicles and capacity of roadway but later on, sensitivity analysis was also done for the results.

The limitations of traditional models have prompted to look into alternative approaches to model estimation that combine flexible data-driven components. These techniques have been applied to a variety of transportation-related applications. In recent years, many machine learning approaches have been utilised in transportation research.

2. Previous Approaches

Marisamynatha and Vedagiri (2013) developed a practical delay model by examining pedestrian delay at signalised junction crosswalks. Gao and colleagues (2012) investigated how pedestrians affect junction delay. Latent variable approaches were employed by Cantillo et al. (2015) to simulate pedestrian crossing behaviour. By creating speed models and evaluating capacity, Golakhiya and Dhamania (2019) investigated the impact of pedestrian crossings on traffic variables. According to Golakhiya et al.'s (2019) study, capacity declined as pedestrian crossflow increased when considering the impact of midblock pedestrian crossings on speed and capacity. Models of speed prediction for urban arterials were developed by Dhamania and Chandra (2013) without taking pedestrian impacts into account. Numerous research presented novel approaches. A high-performance pedestrian gap acceptance model was suggested by Lyons et al. (2001). Vehicle-pedestrian interaction at crossroads and midblock sections was investigated by Ujjal et al. (2016). Gaussian processes and support vector machines were introduced by Xie and Huynh (2010) as tools for forecasting urban traffic. Multiple linear regression modelling was used by Kadali and Vedagiri (2013) to simulate pedestrian road crossing behaviour. Antoniou et al. (2013) used emerging sensors to forecast the short-term traffic situation. A data-driven nonparametric regression model for estimating urban signalised traffic flow was suggested by Yoon and Chang (2014). Kadali et al. (2014) used artificial neural networks (ANN) to study pedestrian mid-block road crossing behaviour. Data-driven car-following models were created by Papathanasopoulou and Antoniou (2015). The fields of traffic modelling, controlled lane behaviour, pedestrian delay, and microscopic pedestrian behaviour were all studied by Kouskoulis et al. (2018), Golakhiya and Dhamaniya (2021), Sharifi and Burris (2019), and Chattaraj (2021).

Previous literatures were used to identify the study's gaps. Some of the gaps identified are as follows:

1. On undivided urban midblock sections, the illegal pedestrian crossings effects on traffic dynamics and road capacity can be studied.
2. Pedestrian modelling can be done using data-driven techniques
3. Comparative analysis of theoretical and data-driven techniques can be done
4. Can compare models with additional goodness-of-fit measures
5. Among various algorithms available, some of the most efficient algorithms that can be used for pedestrian modelling are locally weighted regression, gaussian processes regression and support vector regression.

3. Research Objectives

From a comprehensive literature review and the identified research gaps in previous studies, the research objective for the present study has been formulated as follows:

1. To analyze the influence of crossing pedestrian on traffic characteristics at urban midblock sections using Data-driven approach.

4. Research Methodology

A literature review was conducted in order to have a better understanding of the many aspects of pedestrian impact. The key variables that influence the forecast of speed and capacity have been discovered. The gaps were discovered based on the literature review. Figure 2 depicts the methodology framework for this research, which gives a general structure of the research work based on the study plan. Traffic operations will be influenced by a variety of parameters, including traffic density, speed, and composition. Among these, speed has a significant impact on comfort, service quality, and traffic safety. In addition to traffic variables, environmental factors, and road geometry can influence speed at any given location. All of these factors, however, stay unchanged during the observation period. When traffic volume and highway geometry factors remain constant, the relationship between vehicles and their manoeuvrability becomes increasingly important in traffic management. Manoeuvrability is influenced by the size, shape, acceleration-deceleration characteristics, and driving ability of the vehicle. The "free flow speed" is the speed at which each kind of vehicle may travel at its preferred level depending on the characteristics of the driver and the vehicle when traffic volume and density are low in a traffic stream. The relative interaction between each type of vehicle causes its speed to drop as the volume grows. When pedestrian cross-flow affects such portions, the vehicle-pedestrian interaction affects the speed of each kind of vehicle. Different vehicle kinds are predicted by taking into account variables like the number of people crossing a certain road segment, the density, and the free flow speed of various vehicle categories, among other pertinent aspects.

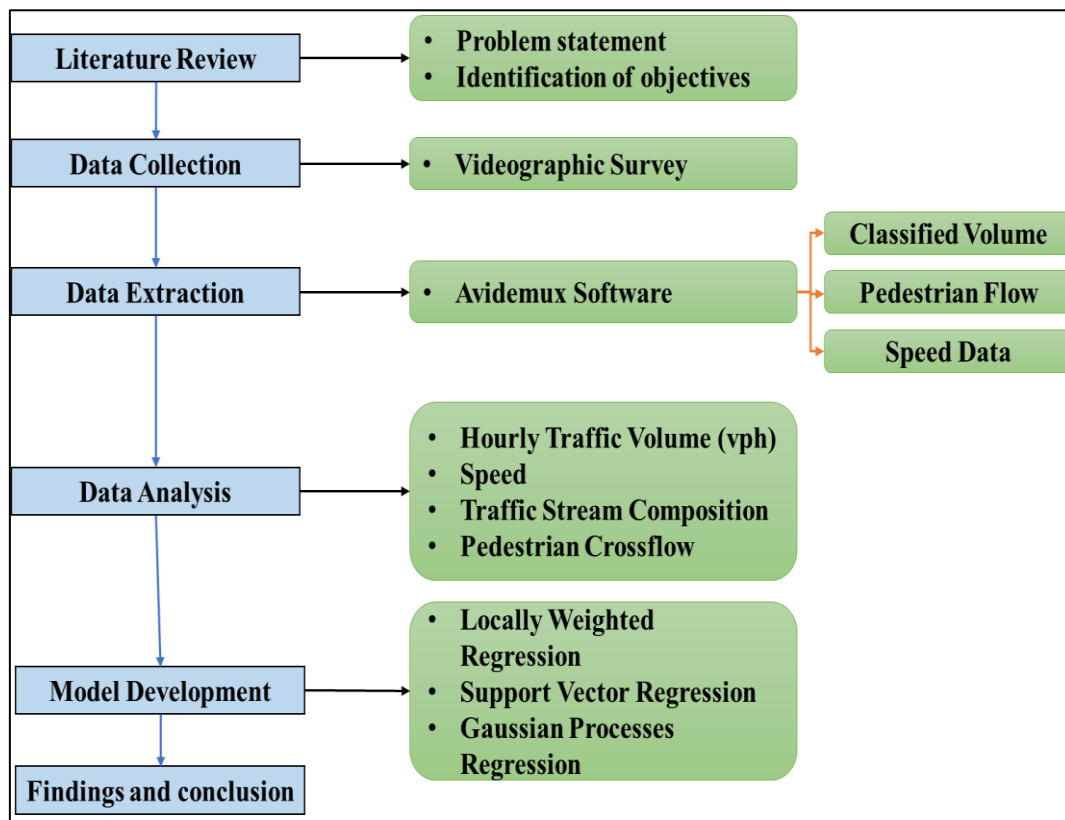


Figure 2: Flowchart for proposed methodology

The capacity of urban midblock sections is assessed with pedestrian crossings in account. Sections with side frictions (pedestrian cross-flow) are examined to see how each vehicle type's speed is affected by pedestrian-vehicle contact. Also, many studies concluded that data driven modelling shows superior performance compared to theoretical modelling. Hence the present study will also use data driven techniques to model the pedestrian and traffic behaviour at undesignated midblock crossings and compare with conventional models as a reference benchmark. There are various data-driven algorithms available for the application to the above said traffic scenario in order to predict the speed and capacity. Locally weighted regression (LOESS), gaussian processes regression (GPR), support vector regression (SVR), and other methods are among the most helpful ones. Thus, by using density and pedestrian cross-flow as independent or explanatory factors and speed as the dependent variable, this study likewise employs these methods. The obtained results are validated against field data using mean absolute percentage error (MAPE) and root mean squared error (RMSE). The results are also compared with the speed predicted by conventional mathematical models using these goodness-of-fit measures. Finally, using the predicted speeds capacity of each section is calculated. The calculated capacity at both non-base sections and base sections are compared to identify the effect of the pedestrian crossing on capacity at non-base sections. Also, the capacity obtained using data-driven techniques is validated for accuracy by comparing it with field data.

5. Data Collection

A videographic survey was used to gather the study's essential data, and Avidemux 2.6 software was used to extract the data. Data on six-lane divided roads with pedestrian crossings and without pedestrian crossings were gathered from important Indian cities including Baroda, Noida, and Surat. Vehicle speeds were measured across 500-meter sections with consistent traffic conditions and no access points. Furthermore, flow and speed measurements were made easier by a 60-meter longitudinal trap located in the middle of each segment (Figure 3). In order to meet the study's goals of examining how pedestrian crossings affect traffic operations and developing data-driven models, this data-gathering approach was crucial.

On normal weekdays from 6:00 a.m. to 6:00 p.m., data is collected using video cameras of high resolution at each of the designated places. On the study site, a 500-meter stretch of road is measured. Throughout the entire length of the trap, cameras were mounted on stands and configured to record views. With 60 km/h speed restrictions for automobiles and 40 km/h for large vehicles like buses, the design vehicle is a car. Videographic surveys are conducted at designated places, and traffic data is recorded. All of the cars in the stream are divided into five categories, as illustrated in Table 1.

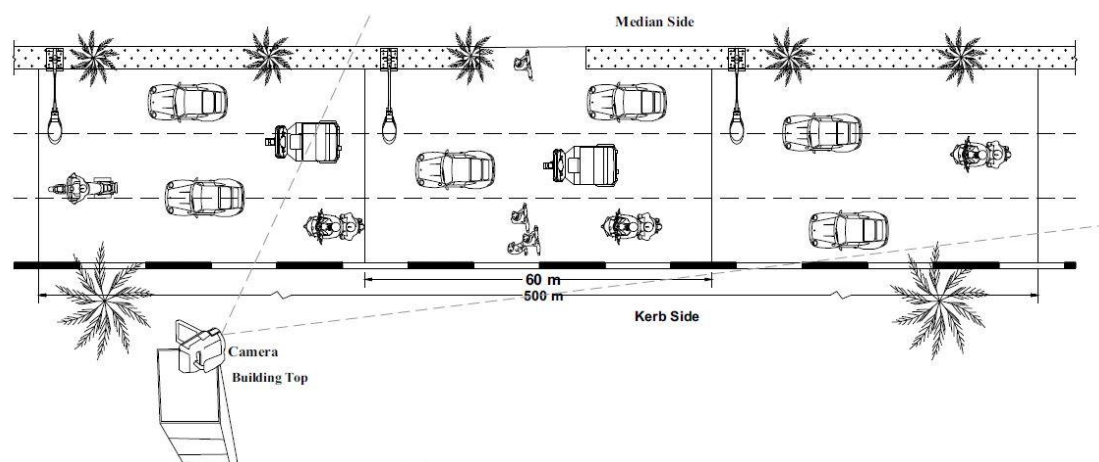


Figure 3: Method of data collection

Table 1: Sizes of vehicle categories

Type of vehicle	Length (m)	Width (m)	Rectangular plan area (m ²)
Two -Wheeler (2w)	1.87	0.64	1.20
Three-Wheeler(3w)	3.20	1.4	4.48
Small car (SC)	3.72	1.44	5.36
Big car (BC)	4.58	1.77	8.11
Heavy vehicle (HV)	10.10	2.43	24.54

Because of the wide range of mechanical properties and physical dimensions of automobiles in India, cars are divided primarily into two categories: small cars (SC) and big cars (BC). SC refers to cars with engines up to 1400cc with typical length of 3.72m and width of 1.44m. BC refers to cars having a 2500cc engine with average lengths of 4.58m and widths of 1.77m. The maximum length and width of the vehicle are used to determine its size. If a vehicle category has more than one type of vehicle, the average measurements are used (Indo-HCM, 2017). For validation purposes, traffic data was collected from a location in Surat with similar geometric features.

6. Data Extraction and Analysis

To get the relevant information, the recorded video is viewed on a computer. Avidemux 2.6 is a program that records the moment a vehicle enters and exits the trap. It does this by converting 1 second of video into 25 frames. At 5-minute intervals, this software retrieves classified volume count data from the recorded video. The video is played frame by frame in order to maintain the necessary accuracy. Every vehicle category is computed in a single round to achieve the best precision possible in volume counting, and then the subsequent vehicle category is repeated. For the purpose of measuring speed, the entry and departure times of the 60-meter trap are precisely recorded every minute with an accuracy of 40 milliseconds (ms) for every kind of vehicle. The 1-minute traffic volume data was then converted into 5-minute categorized volume counts. In the non-base sections, pedestrian crossflow was assessed in addition to volume and speed counts.

Data Analysis is the process of analysing, cleansing, manipulating, and modelling data in order to identify usable information, make conclusions, and support decision-making. Outlier detection is an important part of data analysis. There are two simple methods for outlier detection which are, by scatter plot method and box plots. These two methods are

used in this study for the removal of outliers. Speed, Traffic composition, and hourly traffic volume are extracted in this study. The traffic compositions at selected sites are shown in Table 2. The average composition is also included for reference, however, the 5-minute flow rate varies considerably depending on the composition of the vehicle. For all the locations considered in this study, the composition of two-wheelers compared to other categories of vehicles is high followed by the percentage of three-wheelers as highest. While comparing the total vehicle composition at the selected study locations, the heavy vehicle proportions are relatively less. Similarly, the variation of pedestrian crossflow and the flow rate are also noted in Table 2. Udhana section does not have pedestrian crossflow since it is a base section with no pedestrian crossing. Pedestrian crossflow ranged from 84 ped/h to 1,404 ped/h, as shown in Table 2.

Table 2: Composition of different vehicle categories at study locations

Section	Base or Non-Base	2w (%)	3w (%)	SC (%)	BC (%)	HV (%)	Observed range	
							Pedestrian crossflow (ped/hr)	Traffic volume (veh/hr)
1- Udhana	Base	32	28	26	8	6	-	372-4716
2- Surat	Non-Base	53	35	9	1	2	108 -708	3000-5364
3- Baroda	Non-Base	46	11	36	6	1	84 - 516	624 - 3432
4- Noida	Non-Base	34	21	33	8	4	276 - 1404	3372-4836

The values of the pedestrian crossflow range from very low to very high. It is therefore difficult to assess the capacity loss caused by the reported pedestrian traffic. To mitigate this issue, a range of speed estimates are employed, which assess the effect of pedestrian crossflow on the velocities of various vehicle classes at certain traffic volumes and proportions in the traffic composition. These forecasts are based on data-driven techniques. Furthermore, the stream speed is determined by these speeds, and capacity decrease is computed for each volume of pedestrian cross-flow using speed-volume curves.

7. Model Development

The limitations of traditional models have prompted to investigate alternate ways for model estimate that combine flexible data-driven components. In recent years, many machine learning approaches have been utilised in transportation research. Figure 4 depicts the complete data-driven model creation process. Training and application are the two stages of the strategy. Initially, the model's relevant explanatory variables are identified, and the necessary surveillance data is collected. In the training stage, traffic models are estimated using the surveillance data that is currently available, and these models are then employed in the application step. Finding underlying patterns in the available data that match similar traffic situations is the first step in the training process.

Flexible regression approaches are used in data analysis to build representative models that are stored in a knowledge base. In the application step, the model is pulled from the database and used to estimate the response variable whenever fresh data becomes available. After evaluating the expected values, the model is improved in the next iterations. This chapter provides a summary of computational models that outperform traditional models constrained by theory-based functional constraints in data fitting, all

without the need for explicit functional forms. Alternative data-driven algorithms for estimating speeds in modelling techniques such as Locally Weighted Regression (LOESS), Gaussian Processes Regression (GPR), and Support Vector Regression (SVR) are discussed in this work. The purpose of this work is twofold: to describe and compare alternatives to conventional speed-density models, especially those that are relevant to pedestrian modelling.

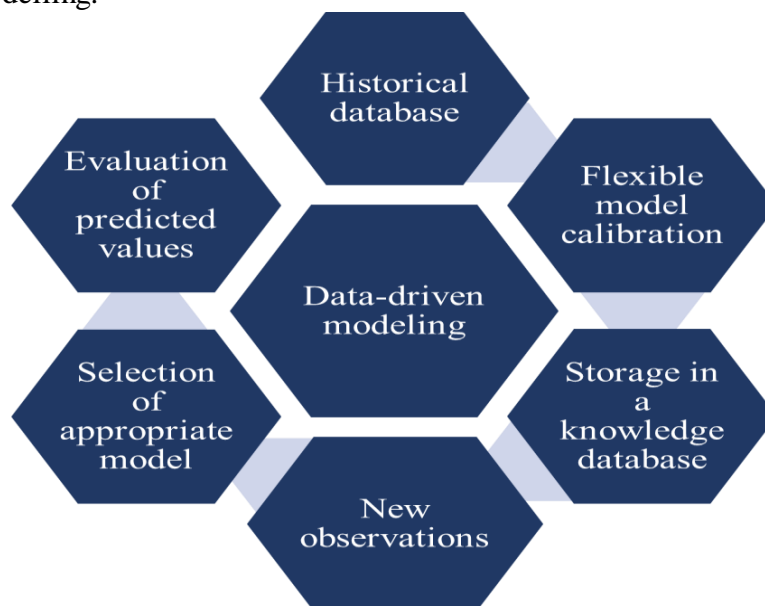


Figure 4: Data- driven model development

7.1. Locally Weighted Regression (LOESS)

There are some important flexible regression algorithms available for speed estimates. According to a critical literature evaluation, LOESS performs better in this situation. As a result, LOESS was used to estimate speeds, with number of vehicles and pedestrian crossflow as explanatory factors. Typically, locally weighted regression uses Equation 1.

$$y_i = u(x_i) + \epsilon_i, \text{ where } i = 1, \dots, n \quad (1)$$

where u is the regression function, n is the index of observations and ϵ_i are residual errors, provide an estimate $u(x)$ of each regression surface at any value x in the d -dimensional space of the independent variables. Let,

q be an integer, and $1 \leq q \leq n$

$u(x)$ at x is estimated using the q observations with the closest x_i values. Each of these points is given a weight based on its distance from x , with points closer to x receiving greater weights and those farther away receiving lower weights. The dependent variable is fitted with a function of the independent variables using least squares with these weights. The value of this fitted function at x is then will be like Equation 2:

$$y = u(x) \quad (2)$$

To derive the proper weights for locally weighted regression, a distance function p defined in the space of the independent variables is necessary. For locally weighted regression, a weight function and a neighbourhood size specification are also necessary (where neighbourhood is defined by the q observations whose value is closest to x , as defined by the appropriate distance function). Equation 3 is an example of a weight function (tri-cubic weight function) that is often used:

$$w_i(x) = W \left[\frac{p(x, x_i)}{d(x)} \right] = \left(1 - \left(\frac{x_i - x}{d(x)} \right)^3 \right)^3 \quad (3)$$

where $d(x)$ is the distance between the most distant predictor value and the influence area (x_i and x). In order to fit linear or quadratic independent variables to the centre of neighbourhoods, weighted least squares is used (Cleveland, W. S., 1988).

7.2. Support Vector Regression (SVR)

A relatively new machine-learning technique is Support vector machines (SVM). They are appealing from a theoretical and practical standpoint, and were designed to address categorization problems. After then, SVM technology was used to deal with issues like density estimation and regression (function approximation). The basic principle behind SVM is to construct a "tube" around the fitted regression curve with a predetermined radius, and assume that the error is zero if all of the data fall inside of it. The degree of inaccuracy applied to other observations might be directly correlated with their distance from the tube. The border curves are the decision boundaries. With the most points, the centre curve—also referred to as the hyperplane in higher dimensions—has the most. Consequently, the decision boundaries are extracted at a distance of ε from the hyperplane. Consider Equation 4 for the hyperplane:

$$Y = PX + b \quad (4)$$

The decision-boundary equation thus given by Equation 5 & 6:

$$PX + b = + \varepsilon \quad (5)$$

$$PX + b = - \varepsilon \quad (6)$$

Therefore, the SVR needs to fulfil the conditions given by Equation 7 should be satisfied by every hyperplane:

$$- \varepsilon < Y - (PX + b) < + \varepsilon \quad (7)$$

The major goal is to choose a decision boundary that comprises data points or support vectors that are closest to the original hyperplane and is at a distance of ' ε ' from it. Consequently, only points falling within the margin of tolerance alone, or inside the decision boundary and having the lowest error rate, are considered. The decision boundaries in SVR need to be properly determined or specified. A few kernel functions are available to help with this. There are many other kinds of kernels available for use in SVR, including radial, polynomial, and linear ones. However, because the RBF kernel is versatile and may be applied in circumstances when prior knowledge of the data is lacking, it was chosen for this investigation. In Equation 8, the RBF kernel format is displayed (Awad, M., 2015).

$$K(x_i, x_j) = \exp(-\gamma ||x_i - x_j||^2) \quad (8)$$

The parameter γ specifies how much influence a single training example has. The greater it is, the closer other examples must be to be affected.

7.3. Gaussian Processes Regression (GPR)

A popular machine learning technique for regression analysis, the Gaussian Process (GP) is renowned for its automated evaluation of prediction uncertainty. In general probability, the prediction $f(x)$ at a given point x is a normally distributed random variable

with the formula $f(x) \sim N(\mu(x), \sigma^2(x))$. The prediction mean ($\mu(x)$) and prediction variance ($\sigma^2(x)$) serve as indicators of prediction uncertainty. By directing the wise distribution of useful training data, estimation of prediction uncertainty supports active learning and is essential for decision dependability. When GP is used, iterations are made to predict function output y^* from input x^* using training data $D = (x_i, y_i)$, where y_i is the real-valued label. GP forecasts complex input-output patterns and computes prediction uncertainty by using Gaussian random process realisation to represent the underlying actual function $y(x)$. In terms of random processes, GP specifies a function $f(\cdot)$ with the following characteristics: at any location x , $f(x)$ has a Gaussian distribution; at different locations x_i and x_j , $f(x_i)$ and $f(x_j)$ are correlated random variables, and the degree of the correlation depends on the distance between x_i and x_j . The word "Gaussian" indicates that GP characterises the random process using a Gaussian distribution, which calls for a formal mean function ($\mu(x)$) and covariance function ($\sigma^2 K(x, x^*)$). The total process variance is shown by the correlation function, also known as the kernel function, $K(x, x^*)$, and σ^2 . A single location x has a Gaussian distribution $f(x)$, and $f = [f(x_1), f(x_2), \dots, f(x_n)]$ represents the multivariate Gaussian distribution for any collection of locations $[x_1, x_2, \dots, x_n]$.

By considering observed labels of training data $[y_1, y_2, \dots, y_n]$ as a draw from a multivariate Gaussian distribution, the Gaussian Process (GP) model describes the actual function $y(x)$ as a Gaussian random process. In training, likelihood estimation is used to determine the probable values of $\mu(x)$, σ^2 , and $K(x, x^*)$. Using a similarity-quantifying kernel function—often referred to as a series of 1-dimensional kernel functions for multi-dimensional input—the trained GP assesses prediction uncertainty at unknown locations. Furthermore, It is frequently assumed that the multi-dimensional kernel function $K(x_i, x_j)$ is a series of multiplications of one-dimensional kernel functions for each of the input features when the input includes two or more features (Equation 9).

$$K(x_i, x_j) = \prod_{k=1}^m K(x_i^k, x_j^k) \quad (9)$$

where m stands for the total number of features, often known as the input dimension. This approach is advantageous since it allows for customised correlation structures for particular input attributes. Gaussian kernel will be used in this study. The formula for a one-dimensional Gaussian kernel is shown in Equation 10:

$$K(x_i, x_j) = e^{-\theta(x_i - x_j)^2} \quad (10)$$

The Gaussian kernel in m dimensions is written as Equation 11:

$$K(x_i, x_j) = \exp \left[- \sum_{k=1}^m \theta_k (x_i^k - x_j^k)^2 \right] \quad (11)$$

where θ is a kernel parameter that affects the correlation strength (J. Wang, 2023).

8. Experimental Results

Using collected traffic and pedestrian data, Python was used to develop models for forecasting vehicle speeds in several categories (2w, 3w, SC, BC, and HV). Using a variety of goodness-of-fit metrics, the study confirms the data-driven techniques (LOESS, SVR, and GPR). The effect of vehicle volume and pedestrian cross-flow on vehicle speeds is evaluated at four sites, including base and non-base portions, using locally weighted regression. Vehicle speeds were calculated using the time it took a car to pass a 60-meter trap; for non-base parts, the number of persons crossing the road was

additionally measured for one minute; and traffic data was averaged from one minute to five minutes. Because estimating density in the field is difficult, it was determined by dividing the number of vehicles of a certain class counted in each 5-minute interval by their average speed. The following are the explanatory variables used to predict speed:

1. $\left(\frac{n_{2w}}{V_{2w}}\right)$: n_{2w} = number of two-wheelers passing the section per second
2. $\left(\frac{n_{3w}}{V_{3w}}\right)$: n_{3w} = number of three-wheelers passing the section per second
3. $\left(\frac{n_{SC}}{V_{SC}}\right)$: n_{SC} = number of small cars passing the section per second
4. $\left(\frac{n_{BC}}{V_{BC}}\right)$: n_{BC} = number of big cars passing the section per second
5. $\left(\frac{n_{HV}}{V_{HV}}\right)$: n_{HV} = count of heavy vehicles passing the section per second
6. n_{ped} = number of pedestrians crossing the section per second

$V_{2w}, V_{3w}, V_{SC}, V_{BC}, V_{HV}$ in the above terms represent the average speed of respective category of vehicles. Pedestrian-vehicle interaction was used to quantify how different vehicle types' speeds were affected by crossing pedestrians. The Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) values derived from field data applied to mathematical models created by Golakiya and Dhamaniya, 2019 are used to compare the anticipated outcomes with the observations collected from the field data. The mathematical models developed by them for friction sections (non-base sections here) are displayed as Equation 12, and the parameter values are given in Table 3.

$$V_{mode} = C - a1 \left(\frac{n_{2w}}{V_{2w}}\right) - a2 \left(\frac{n_{3w}}{V_{3w}}\right) - a3 \left(\frac{n_{SC}}{V_{SC}}\right) - a4 \left(\frac{n_{BC}}{V_{BC}}\right) - a5 \left(\frac{n_{HV}}{V_{HV}}\right) - a6 \left(\frac{n_{ped}}{V_{mode}}\right) \quad (12)$$

Table 3: Parameters for regression model (Equation 12)

Mode	C	a1	a2	a3	a4	a5	a6
2w	16.35	68.54	28.36	24.53	49.42	38.56	28.24
3w	13.92	45.61	59.11	28.48	16.92	41.39	28.72
SC	16.96	60.73	33.30	40.43	61.96	70.52	30.87
BC	17.36	72.16	40.40	33.71	79.43	85.39	31.17
HV	13.50	50.91	47.70	23.80	53.40	141.52	4.55

Table 4: RMSE and MAPE value comparison between models

Type of Vehicle	LOESS		Mathematical model		% Improvement (RMSE)	% Improvement (MAPE)
	RMSE	MAPE	RMSE	MAPE		
2w	0.35	2.90	0.62	5.35	44	46
3w	0.33	3.22	1.63	9.97	80	68
SC	0.57	4.87	0.92	7.81	38	38
BC	0.85	7.16	1.11	9.25	23	23
HV	0.82	7.57	2.40	9.96	66	24

Table 4 presents a comparison of speed predictions using LOESS and shows that it performs better than standard models with less error values. On average, there is a 57%

improvement in both RMSE and MAPE. In a similar way, LOESS excels above the other two data-driven strategies (SVR and GPR) with an average percentage improvement in error reduction of 57%.

Table 5: Speed Comparison (non-base and base sections)

Speed characteristic	Modelling technique	2W		3W		SC		BC		HV	
		BS	NBS	BS	NBS	BS	NBS	BS	NBS	BS	NBS
Mean speed (kmph)	LOESS	56	33	53	29	50	33	49	32	48	32
	SVR	56	35	53	31	51	34	50	35	48	32
	GPR	57	35	53	31	51	35	49	35	50	32

The mean speeds at base and non-base portions with and without pedestrian crossings are contrasted in Table 5. Due to interactions between pedestrians and vehicles, non-base parts have slower speeds. Sensitivity study looked at changes in speed when there was no change in the composition of the vehicle and the flow of pedestrians, and vice versa. Figure 4 (a) and (b) depicts the results of the sensitivity analysis. The graph clearly shows that Except for heavier vehicles, which gradually increase in speed, vehicle speeds drop as the number of pedestrian crossings rises. For big cars, the first decline is less sharp, suggesting a gentler hit with fewer people. Similarly, increasing traffic congestion causes all vehicle classifications to experience notable speed decreases.

The greatest traffic flow achievable on a given roadway using all available lanes is known as road capacity, and it is commonly stated in vehicles per hour or vehicles per day. Several things influence this, including traffic circumstances, road geometry characteristics, environmental considerations, pedestrian influences, and so on. When calculating a section's traffic-carrying capacity, three essential traffic parameters— speed, flow, and density—are critical.

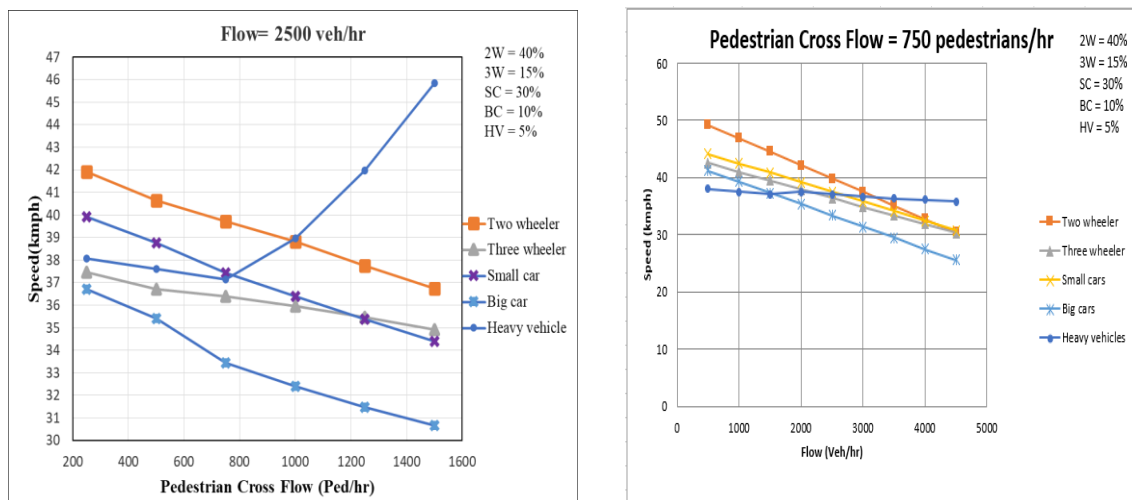


Figure 4: (a) Variation of speed v/s Pedestrian crossflow, (b) Speed v/s traffic flow variation

The emphasis for capacity evaluation is always on flow and speed due to the difficulty of measuring density in the field. Plotting speed-flow curves employed stream speed since numerous vehicle groupings had speeds that varied significantly from one another. Weighted average 1- to 5-minute intervals were used to convert each kind of vehicle's

speed, as calculated by the speed model, to mean stream speed at certain flow and pedestrian crossflow levels. Density was calculated for every 5 minutes of data using stream flow and speed. The capacity values obtained from above analysis for different study locations are shown in Figure 5.

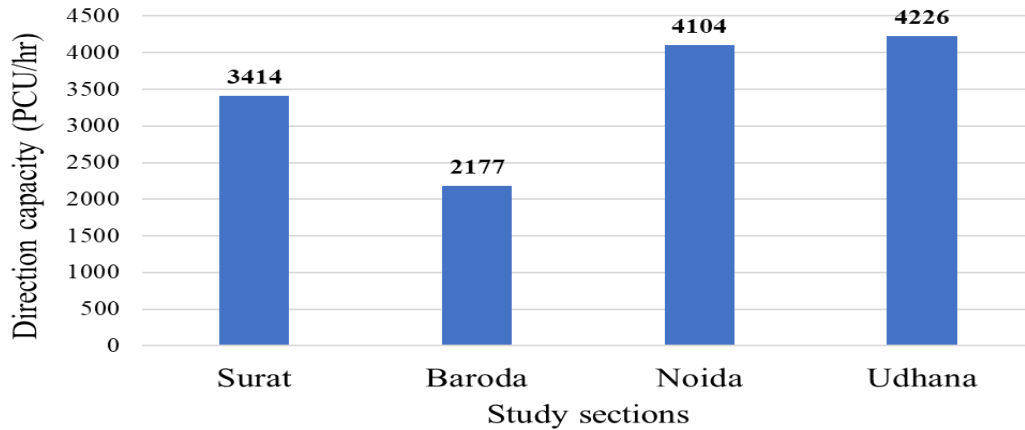


Figure 5: Capacity variation at various study sections

All of the non-base sections have a significant reduction in capacity, as seen in Figure 5. The reason for this is that when there is pedestrian crossflow at a section, the traffic volume reduces. As part of the sensitivity analysis, the variation in capacity with varied degrees of pedestrian crossflow is calculated. For each of the non-base sections, the Figure 6 shows how capacity changes with varied degrees of pedestrian crossflow.

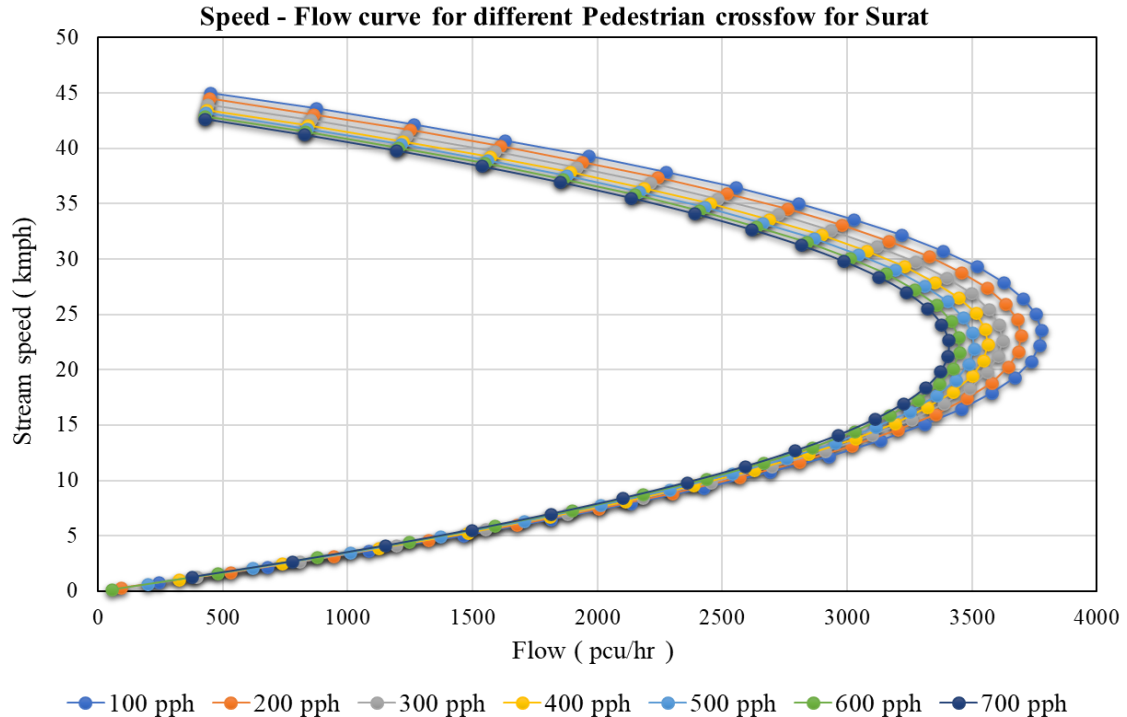


Figure 6: Variation in capacity with varying pedestrian crossflow for Surat

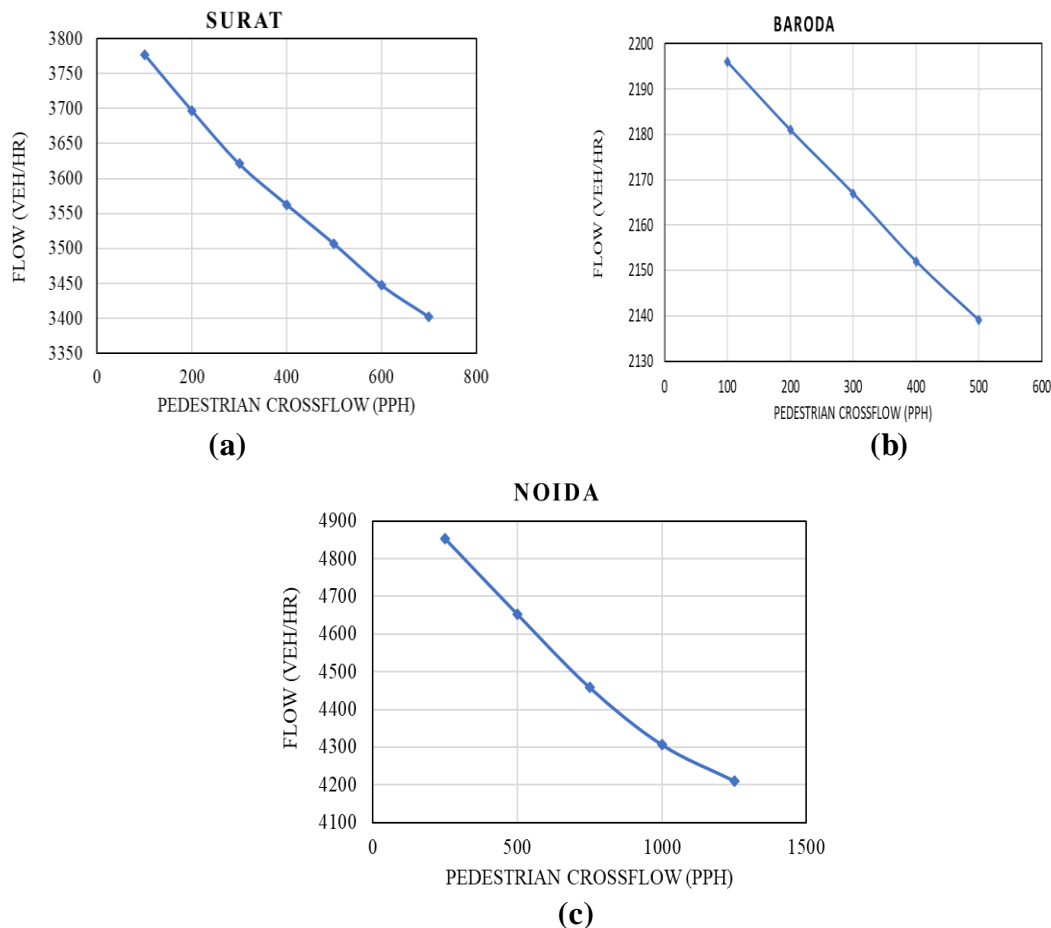


Figure 7: Reduction in capacity with pedestrian crossflow at non-base sections

With steady traffic volume, pedestrian crossflow in Surat varied from 100 to 700 people per hour (pph). Likewise, pedestrian flows in Baroda and Noida ranged from 100 to 500 pph. Figure 7(a), (b), and (c) shows how capacity is reduced. Because pedestrians impede vehicle movement, capacity falls as pedestrian crossflow increases. Reduced pedestrian cross-flow has little effect on traffic volume and flow. But when the number of pedestrians rises, cars could have to stop and yield, which would reduce the capacity of a road segment.

9. Model validation

To validate the formulated models, traffic data from a new location influenced by crossing pedestrians (non-base section) in Surat city was collected. The geometric features of this location are similar to those of other locations. The speed, traffic volume, and pedestrian crossing flow were observed at the site. The data was used to predict speed using the LOESS method and the mathematical models proposed by Golakiya and Dhamaniya (2019). The results of the speed predictions from the two approaches are presented in Table 6.

Table 6: RMSE and MAPE value comparison between LOESS and mathematical models

Type of Vehicle	LOESS		Mathematical model		% Improvement (RMSE)	% Improvement (MAPE)
	RMSE	MAPE	RMSE	MAPE		

2w	0.36	3.25	0.62	6.12	42	47
3w	0.32	3.02	1.52	8.15	79	63
SC	0.6	4.78	0.92	9.25	35	48
BC	0.91	7.02	1.11	8.87	18	21
HV	0.78	8.15	2.4	9.56	68	15

Table 6 presents a comparison of speed predictions using the LOESS method at the independent location in Surat. The table indicates that the results from LOESS are superior to those from the mathematical models, with lower error values. There is an approximate 43% improvement in the results.

10. Conclusions

With the goal to estimate the speeds of vehicle types (2w, 3w, SC, BC, HV), this study compares data-driven methods (LOESS, SVR, GPR) with theoretical pedestrian models, mainly linear regression. The results show that data-driven models perform better than theoretical models, especially LOESS, with lower RMSE and MAPE values. The study highlights the potential for precise outcomes of data-driven modelling approaches for pedestrians, adding to our understanding of their performance. Furthermore, the findings of this study show that all data-driven techniques, particularly LOESS, are extremely promising approaches for speed and pedestrian modelling, as they are likely to deliver accurate results. Sensitivity analysis of speed with respect to change in pedestrian crossflow was done. The findings revealed that when the amount of pedestrian crossflow increases, all vehicle speeds decrease. The acquired traffic data on various metropolitan midblock portions impacted by crossing pedestrians was investigated further to determine the capacity of such non-base sections. For 6-lane highways in four separate locations, average base section capacity is 4226 PCU/hr for one direction, while for non-base sections it ranged from 2177 PCU/h to 4104 PCU/h. The discrepancy in capacity is due to variations in operating speed at various sections in various towns and cities of various sizes. Sensitivity analysis of capacity also was done to check the effect of pedestrian crossflow on it. Results showed that, as pedestrian crossflow increases, at non-base sections, capacity tend to decrease. The datasets used in this work are previously used to calibrate linear regression models. Although this approach has yielded useful insights, it may not comprehensively measure the relative performance of the models. In order to overcome this constraint, forthcoming research will integrate an autonomous third dataset, which was not utilised in the calibration process of either model. This will facilitate a fairer comparison and a thorough sensitivity analysis.

References

- Antoniou, C., Koutsopoulos, H. N., and Yannis, G. (2013). Dynamic data-driven local traffic state estimation and prediction. *Transportation Research Part C*. 34: 89–107
- Awad, M., Khanna, R. (2015). Support Vector Regression. *Efficient Learning Machines*. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4302-5990-9_4
- Cantillo, V., Arellana, J., and Rolong, M. (2015). Modelling pedestrian crossing behavior in urban roads: A latent variable approach. *Transportation Research Part F Traffic Psychology and Behavior*. 32:56-67

- Chattaraj, Ujjal. (2021). Development of a Microscopic Fuzzy Inference Based Cellular Automata Model for Pedestrian Flow. *European Transport/Trasporti Europei*. 1-17. 10.48295/ET.2021.83.3.
- Cleveland, W. S., & Devlin, S. J. (1988). Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association*, 83(403), 596–610. <https://doi.org/10.1080/01621459.1988.10478639>
- Dhamaniya, A., and Chandra, S. (2013). Speed Prediction Models for Urban Arterials under Mixed Traffic Conditions. *Procedia - Social and Behavioral Sciences*. 104: 342 – 351
- Gao, L., Liu, M., and Feng, J. (2012). Delay Modeling of Ped-Veh System Based on Pedestrian Crossing at Signalized Intersection. *Procedia - Social and Behavioral Sciences*. 43: 530 – 539
- Golakiya, H.D., and Dhamaniya, A. (2019). Modeling Speed and Capacity Estimation at Urban Midblock Sections under the Influence of Crossing Pedestrians. *Journal of Transportation Engineering, Part A: Systems*. 145: 9
- Golakiya, H.D., and Dhamaniya, A. (2021). Reexamining Pedestrian Crossing Warrants based on Vehicular Delay at Urban Arterial Midblock Sections under Mixed Traffic Conditions. *Journal of Transportation Engineering, Part A: Systems*, ASCE.147 (8)
- Golakiya, H.D., Patkar, M., and Dhamaniya, A. (2019). Impact of Midblock Pedestrian Crossing on Speed Characteristics and Capacity of Urban Arterials. *Arabian Journal for Science and Engineering*. 44: 8675–8689
- Golakiya, Haresh & Chauhan, Ritvik & Dhamaniya, Ashish. (2020). Evaluating Safe Distance for Pedestrians on Urban Midblock Sections Using Trajectory Plots. *European Transport/Trasporti Europei*. No. 75, Paper no. 2. ISSN 1825- 3997.
- Indo-HCM (2017). *Indian Highway Capacity Manual (Indo-HCM): CSRI-Central Road Research Institute, New Delhi*.
- J. Wang, (2023). An Intuitive Tutorial to Gaussian Process Regression, *Computing in Science & Engineering*, vol. 25, no. 4, pp. 4-11, doi: 10.1109/MCSE.2023.3342149.
- Kadali, R. B., Rathi, N., and Vedagiri, P. (2014). Evaluation of Pedestrian Mid-block Road Crossing Behavior Using Artificial Neural Network (ANN). *Journal of Traffic and Transportation Engineering (English Edition)*. 1(2): 1911-1922
- Kadali, R., and Vedagiri, P. (2013). Modelling pedestrian road crossing behaviour under mixed traffic condition: *European Transport - Trasporti Europei*, No. 55, pp. 1– 17.
- Kouskoulis, G., Spyropoulou, I., and Antoniou, C. (2018). Pedestrian simulation: Theoretical models vs. data driven techniques. *International Journal of Transportation Science and Technology*. 7: 241–253
- Lyons, G., Hunt, J., and McLeod, F. (2001). A neural network model for enhanced operation of midblock signaled pedestrian crossings. *European Journal of Operational Research*. 129(2):346-354
- Marisamynathan, S., and Vedagiri, P. (2013). Modeling Pedestrian Delay at Signalized Intersection Crosswalks under Mixed Traffic Condition. *Procedia - Social and Behavioral Sciences ASCE*. 104: 708-717
- Papathanasopoulou and Antoniou (2015). Towards data-driven car-following models. *Transportation Research Part C*. 55: 496-509
- Sharifi, F., and Burris, M.W. (2019). Application of machine learning to characterize uneconomical managed lane choice behavior. *Case Studies on Transport Policy*. 7(4): 781-789

- Ujjal, C., Das, J.B., and Suresh, B. (2016). Impacts of Vehicle Pedestrian Interaction on Traffic Flow: Midblock and Intersection. *European Transport\Trasporti Europei*, No. 60, pp. 1–13.
- Xie, Y., and Huynh, N. (2010) Kernel-Based Machine Learning Models for Predicting Daily Truck Volume at Seaport Terminals. *Journal of Transportation Engineering*, ASCE. 136: 12
- Yoon, B., and Chang, H. (2014). Potentialities of Data-Driven Nonparametric Regression in Urban Signalized Traffic Flow Forecasting. *Journal of Transportation Engineering*. 140 (7): 1943-5436