Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

Service Quality Analysis of Signalized Intersections from the Perspective of Bicycling

Sambit Kumar Beura[a], Kondamudi Vinod Kumar[a], Shakti Suman[a], Mahabir Panda[a], Prasanta Kumar Bhuyan[a,*]

[a]*Department of Civil Engineering, National Institute of Technilogy Rourkela, Odisha 769008, India*

Abstract

This article proposes reliable mathematical models for the assessment of signalized intersections from bicycling perspective under mixed traffic conditions. For investigation purposes, extensive data sets (related to the geometrical, operational and built-environmental characteristics) are collected from 70 intersection approaches located in seven Indian cities. Each approach has been rated by 200 on-site bicyclists based on their perceived satisfaction levels on a Likert scale of 1–6 (excellent–worst). Using these ratings as the array of dependent variable, Spearman's correlation analysis has been carried out and eight intersection attributes having significant influences on the bicycle service quality are identified (such as the effective approach width, peak hour volume, crossing pedestrian volume, and average delay, etc.). Subsequently, three efficient techniques namely, associativity functional network (FN), genetic programming (GP) and step-wise regression are utilized to develop highly reliable service prediction models, called bicycle level of service (BLOS) models. Of all, the FN technique has produced the most efficient BLOS model with coefficient of determination (R^2) values of 0.917 and 0.915 in its training and testing stages respectively. On the other hand, the regression technique has produced the least complex model. The service qualities of signalized intersections are classified into six levels A–F (excellent–worst) by using a letter-graded scale defined in this article. Results have shown that about 86% intersection approaches offering BLOS 'C' or inferior. Thus, the redesigning works supported by the outcomes of this research may be carried out for the improvement of existing signalized intersections.

Keywords: Signalized intersection; Heterogeneous traffic; Bicycle level of service; Artificial intelligence.

* Corresponding author. Tel.: +91-661-246 2349; fax: +91-661-246 2349.
E-mail address: pkbtrans@gmail.com

Nomenclature

a_i	coefficient of variable x at i^{th} degree of shape function
b	bias parameter
$BLOS_{max}$	maximum value of overall BLOS score
$BLOS_{min}$	minimum value of overall BLOS score
$BLOS_{SA}$	overall perceived BLOS score for a signalized approach
$BLOS_{SA, Norm}$	normalized value of overall BLOS score
C	cycle length
C'	set of constants
CPV	crossing pedestrian volume across the path of through bicyclists
D	average stopped time delay incurred by through bicyclists
d_{max}	maximum depth of GP tree
E	Nash–Sutcliffe efficiency coefficient
e_j	error in j^{th} data
E_n	sum of squared errors for n data sets
E_λ	auxiliary function
f	GP function
F	set of functional elements
g	green period
g/C	green time over cycle length
m	degree of shape functions
p	significance level
p_c	crossover probability
PHV	peak hour volume
p_m	mutation probability
p_r	reproduction probability
PT	on-street parking turnover
R^2	coefficient of determination
s	number of input variables
SDP	surrounding developmental pattern
v_1, v_2	random variables
V_{Turn}	volume of turning vehicular traffic across the path of through bicyclists
w	gene weight
W_{Eff}	effective approach width
X	vector of independent variables
$x_1, x_2, \ldots, x_s$	FN model inputs
x_{s+1}	FN model output
y	typical output variable
α_k, x_0	constant terms
ρ	Spearman's rho
ϕ	shape function
ϕ_i	shape function with degree i

1. Introduction

Road intersections are the at-grade junctions where two or more roads meet or cross. To safeguard the traffic and pedestrian movements at these hazardous and bottleneck points, one of the most effective attempts usually adopted is signalization. Pre-timed, partially actuated, and fully actuated are various signalization techniques used to control the traffic movements. The service quality of signalized intersections from bicycling perspective is largely determined by their geometrical properties, traffic flow characteristics, and built-environmental conditions. A thorough understanding of these variables is a matter of utmost concern for the planning and designing of bicycle-friendly intersections. Previous studies carried out in the state-of-the art have primarily focused on the service quality analyses of signalized intersections operating under homogeneous traffic flow conditions, which prevails in developed countries. This kind of traffic operation symbolizes a lane-disciplined flow of identical vehicles (cars), which is completely different from that in developing countries. The road traffic in the later context is heterogeneous in nature which symbolises a mixed flow of small and big vehicles (2-wheelers, 3-wheelers, 4-wheelers, and heavy vehicles, etc.) with a weak lane discipline. The unavailability of separate bicycle signal and bicycle lane facilities on the approaching legs at many intersections enforces the bicyclists to mobilize at their own risk. Such a loose structure of the regulatory system makes the operational characteristics of bicyclists substantially different and complex from that in developed countries. Thus, the efficiency of signalized intersections under heterogeneous traffic conditions cannot be accurately analysed with the help of heterogeneous traffic flow models.

In response to the above gap, the present study has been undertaken with the following two major objectives: (1) to figure out the variables having significant effects on bicycle through movements at signalized intersections under heterogeneous traffic conditions, and (2) to develop reliable bicycle level of service (BLOS) models for quantifying the efficiency of the mentioned facility. To accomplish these objectives, the bicycling environments persisting at 70 signalized intersection approaches are thoroughly analyzed. These intersections are identified from the diversified road environments of seven Indian cities. From these sites, a wide range of data including intersection attributes, personal characteristics of total 14,000 on-site bicyclists, and perceived satisfaction scores (BLOS scores) are collected. Subsequently, variables having significant influences on the intersection efficiency are identified, and modeled using the genetic programming (GP), associativity functional network (FN), and step-wise regression techniques. GP and FN are two highly reliable artificial intelligence (AI) techniques, which have numerous advantages over various other prediction tools available at present. These tools are problem-driven in nature, which do not assume or impose any particular kind of relationship (e.g., linear and exponential) among the input and output variables beforehand. This principle has made these techniques extremely suitable to deal with high dimensional problems with a large amount of input data. The goodness-of-fit and validation criteria of developed models are assessed in terms of various statistical parameters, and the best model is reported. The crucial outcomes of this study will largely help the traffic planners and engineers to quantify and enhance the bicycle-friendliness of urban signalized intersections carrying heterogeneous traffic.

2. Background studies

Davis (1987) proposed the first-ever mathematical model for the assessment of road intersections namely, Intersection Evaluation Index (IEI) model. This model is a function of main-street traffic volume, cross street traffic volume, type of signalization, and various geometric factors (presence of left- and right-turn lanes, number of through lanes, curb radii and sight distance). The 2000 version of American *Highway Capacity Manual* (HCM-2000) considered controlled bicycle delay (s/bicycle) as the only measure of effectiveness for estimating the BLOS of signalized intersections (TRB, 2000). However, Crider et al. (2001) reported that the safety and comfort levels of bicyclists traveling through road intersections are largely influenced by conflicts with turning vehicles, exposure to the conflicts with vehicular traffic, and the crossing delay.

Landis et al. (2003) applied the step-wise regression method to model the service levels experienced by through bicyclists at signalized intersections. This model included intersection crossing distance and three attributes of the subject approach namely, number of lanes, traffic volume, and sum total width of the outside lane and bicycle lane (if present). TRB (2010) incorporated a minor modification in this model and documented the revised model in HCM-2010. In this revised model, the traffic volume on the subject approach was replaced with the sum total volume of the through, left and right-turn demand flow rates. The Charlotte Department of Transportation (CDOT) developed a LOS methodology to assess the design features that affect the quality of bicycle crossings at signalized intersections (Steinman and Hines, 2004). The influencing variables included intersection crossing distance, roadway space allocation (i.e., crosswalks and bicycle), corner radius dimension, and traffic signal characteristics.

Carter et al. (2007) proposed a Bicyclist Intersection Safety Index (Bike ISI) model for the assessment of bicycle through movement at road intersections. This model considered several parameters such as main-street and cross-street traffic volumes, main-street speed limit, presence of turning vehicles across the path of through bicyclists, number of right turn lanes, presence of bicycle lane, presence of signalization (yes or no), and on-street parking activities on the main approach. Strauss et al. (2013) revealed that the presence wider crosswalks and bus stops increases bicyclists' injury occurrence at signalized intersections, while the provision of raised medians decreases the same. Jensen (2013) proposed a cumulative logit-based BLOS model for predicting the riding quality of bicyclists moving straight ahead at signalized intersections. This model considered various parameters such as width of bicycle facility at stop line, type of crossing facility for bicyclists, and type of bicycle facility before the intersection.

However, no standard BLOS model is available for the reference of road authorities in developing countries. To date, only Beura et al. (2017) have developed a step-wise regression-based BLOS model for the assessment of bicycle through movement at signalized intersections carrying heterogeneous traffic. Input variables of this model are specific to the intersection approach of interest and include approach width per direction (m), peak hour traffic volume (Passenger Car Units per hour or PCUs/h), pavement condition index, average stopped time delay (min/bicycle), land use pattern, and on-street parking turnover. Thus, the effects of other parameters like cross-street traffic volume, turning vehicular volume, crossing pedestrian volume, and the green time over cycle length, etc. need further investigation. Another limitation of this model is that limited quantity of data set is used for its training and testing, i.e., 25 and 10 approaches respectively. In light of this, the present study has carried out extensive data collections and an in-depth investigation of the intersection BLOS under said conditions.

3. Brief descriptions of Modelling Tools

Of the three modelling tools used in this study, the step-wise regression analysis is a well-known statistical tool, whose analytical procedure follows three major steps. First, several combinations and/or transformations (e.g., square, square root, inverse, exponential, logarithmic, etc.) of input variables are carried out. Second, each variable is included in the analysis either in its original form or its modified form (combined, transformed or both), and its coefficient is estimated. Third, the significance (*p*-value) of each coefficient is tested. These steps are repeated until the best combination of all variables is obtained. Subsequently, the best model is selected for implementations. The principles of GP and associativity FN techniques are discussed below.

3.1. Genetic programming (GP)

GP, introduced by Koza (1992), is an evolutionary approach that automatically evolves computer programs to develop predictive models without specifying their structures beforehand (McPhee et al., 2008). It applies the Darwin's theory of natural selection to select and reproduce non-linear models. The modelling procedure starts with the formation of an initial population of a large number of individuals generated at random. Each individual resembles a tree structure, called GP tree. The nodes of a GP tree are constructed by selecting suitable elements either from a functional set (arithmetic operators, mathematical functions, Boolean operators, logical expressions or any other user-defined functions) or a terminal set (variables, constants or both).

Once the initial population is created, the objective function (root mean square error or 'RMSE' in this study) assesses the fitness of each individual. Unless the problem is so small and simple, all most all individuals of the initial population exhibit extremely poor fitness. Thus, GP attempts to create an offspring population of better-fitted individuals by implementing various genetic operations (reproduction, crossover and mutation) on them. These three operations are carried out at their respective frequencies p_r, p_c and p_m respectively, where $p_r + p_c + p_m = 1$. The "reproduction" process involves the direct duplication of high-fitness individuals of the initial population to the offspring population. The "crossover" operation involves the production of better-fitted individuals by exchanging the randomly chosen parts of two parental individuals. Fig. 1a illustrates a typical crossover of two parents $\ln\{v_1^2/(v_2+2)\}$ and $\exp(v_1/v_2^{0.5})$ to produce the offspring $\ln(v_1^2/v_2^{0.5})$ and $\exp\{v_1/(v_2-5)\}$. The "mutation" operation involves the production of a better-fitted individual by randomly replacing a node of the parent with another element of the functional or terminal set. Fig. 1b illustrates a typical mutation operation on a parent $\exp(v_1+v_2)$ to produce an offspring $\exp(v_1-v_2)$.

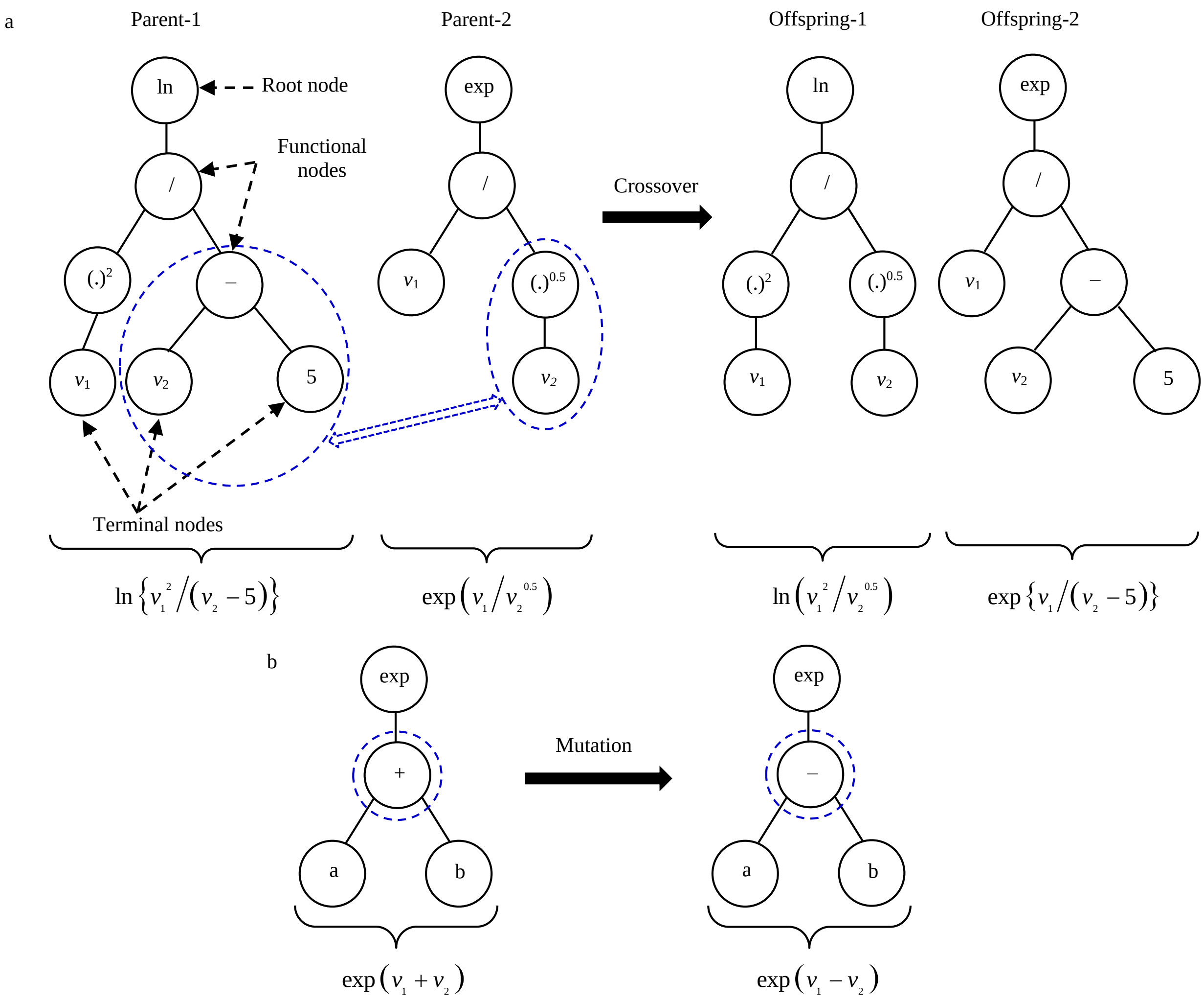


Fig. 1. Typical evolutionary operations in GP: (a) crossover; (b) mutation.

The GP formalism continues over several "generations", and at each generation, the existing population is replaced by a new population of better-fitted individuals. This process terminates when either the maximum number of generations is reached or a threshold fitness value is achieved. Finally, the best-fit individual appeared at any generation is defined as the output of GP analysis. The mathematical expression of a BLOS model developed through the GP formalism can be presented as follow:

$$\mathrm{BLOS}_{\mathrm{Pred}} = b + w \times f\left[X, F\left(X\right), C\right] \tag{1}$$

Where, $\mathrm{BLOS}_{\mathrm{Pred}}$ (predicted BLOS score) represents the model output, *b* is the bias parameter, *w* is the weight of gene *f*[*X*, *F*(*X*), *C*], *f* represents the GP function, *X* represents the vector of independent variables, *F* represents the set of functional elements, and *C'* represents the set of constants. Here, *w* and *b* are estimated by the ordinary least squares method.

3.1.1 Optimum values of GP algorithm parameters: The GP formalism is carried out and regulated by the optimum values of various algorithm parameters such as population size, number of generations, and the maximum depth of

GP tree (d_{max}), etc. As each problem has its own curvature of the solution space, the optimum values of GP algorithm parameters are specific to the problem at hand (Searson et al. 2010). However, no concrete guideline is available at present estimate the same. Thus, a stepwise selection procedure is adopted and the parameter to be optimised is scanned thoroughly in a wide range (Table 1), while keeping the values of all other parameters unchanged. At each trial, the model performance is tested, and the best value of the corresponding parameter is identified. Higher ranges of the parameters (as compared to Table 1) could also be adopted elsewhere keeping in mind that the computational cost and model complexity would be higher. The condition used to terminate the program execution is the maximum number of generations or a fitness value less than 0.0001, whichever is earlier. Further details on the GP formalism could be found in Koza (1992), McPhee et al. (2008), and Searson et al. (2010).

Table 1. Implemented ranges of GP algorithm parameters.

Parameter	Range
Population size	100–1500
Number of generations	100–500
d_{max}	2–10
p_r	0.01–0.07
p_c	0.75–0.9
p_m	0.05–0.15

3.2. Associativity functional network (FN)

In the recent past, FN has evolved as a powerful alternative to the artificial neural networks (ANN) technique (Castillo et al., 2000a, 2000b). Unlike in ANN, where the best-fit model is developed by selecting the number of hidden layers and the number of neurons in each hidden layer through numerous trials, FN derives its initial topology based on the domain knowledge of the problem. Once the initial topology is established, the FN algorithm attempts to simplify it through the concepts of functional equations (Castillo et al., 2004). The simplified network is called as the equivalent FN of the initial network. Further, FN uses the data knowledge to estimate unknown neuron functions using which the predictive model of interest is derived.

The present study has applied the concepts of associativity FN, which is the most recent and simplest form of FN. This approach uses the basic theory of functional equations to convert any multi-input network with s inputs (x_1, x_2,…, x_s) and one output (x_{s+1}) to an associative network of equal prediction ability (Castillo et al., 2000a). Fig. 2a presents a typical FN structure of three inputs (x_1, x_2, and x_3) and one output (y), and its associative FN is presented in Fig. 2b. In Fig. 2b, ϕ represents the shape function (SF) of input variables, which can be any mathematical function.

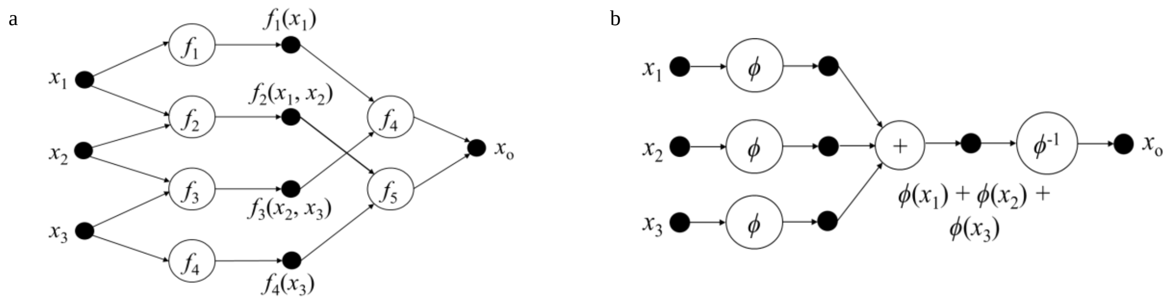


Fig. 2. Representation of (a) a typical FN structure; and (b) its associativity FN structure.

The mathematical expression of an associative FN, $f(x)$, built for any independent variable x is represented as:

$$f\left(x\right)=\sum_{i=1}^{m}a_i\phi_i\left(x\right) \tag{2}$$

Where, m is the degree of SFs used, a_i is the coefficient of variable x at i^{th} degree, and ϕ_i is the SF with degree i.

For a given problem with s inputs, all input functions $f_1(x_1)$, $f_2(x_2)$, $f_3(x_3)$,… $f_s(x_s)$ are derived using Equation 2 and integrated to estimate the output function $f_{s+1}(x_{s+1})$ as follow:

$$f_{s+1}\left(x_{s+1}\right) = f_1\left(x_1\right) + f_2\left(x_2\right) + \ldots + f_s\left(x_s\right) \tag{3}$$

Accordingly, the error in j^{th} observation (e_j) is obtained as:

$$e_j = f_1\left(x_{1j}\right) + f_2\left(x_2 j\right) + \ldots + f_s\left(x_{sj}\right) - f_{s+1}\left(x_{s+1,j}\right) \tag{4}$$

To estimate the values of a_i, the sum of squared error E_n for all n data sets is minimized as follows:

$$E_n = \sum_{j=1}^{n} e_j^{\,2} = \sum_{j=1}^{n}\left[\sum_{i=1}^{m} a_i \left\{\varphi_i\left(x_{1j}\right) + \varphi_i\left(x_{2j}\right) + \ldots + \varphi_i\left(x_{sj}\right) - \varphi_i\left(x_{s+1,j}\right)\right\}\right]^2 \tag{5}$$

subject to

$$f_k\left(x_0\right) = \sum_{i=1}^{m_k} a_{ki}\varphi_{ki}\left(x_0\right) = \alpha_k; \quad k = 1, 2, \ldots, s+1 \tag{6}$$

Where, x_0 and α_k are constant terms.

Further, an auxiliary function E_λ is defined by using the Lagrangian multipliers as:

$$E_\lambda = \sum_{j=1}^{n}\left[\sum_{k=1}^{s+1}\sum_{i=1}^{m} a_{ki}\varphi_{ki}\left(x_{kj}\right)\right]^2 + \sum_{k=1}^{s+1}\lambda_k\left[\sum_{i=1}^{m} a_{ki}\varphi_{ki}\left(x_0\right) - \alpha_k\right] \tag{7}$$

The minimum value of E_λ is then estimated by using the following two equations.

$$\frac{\partial E_\lambda}{\partial a_{kr}} = 2 \times \sum_{j=1}^{n}\left[\sum_{k=1}^{s+1}\sum_{i=1}^{m} a_{ki}\varphi_{ki}\left(x_{kj}\right)\right] \times \varphi_{kr}\left(x_{jk}\right) + \lambda_k\varphi_{kr}\left(x_0\right) = 0 \tag{8}$$

Where, $k = 1, 2, \ldots s + 1$, and $k = 1, 2, \ldots m$.

$$\frac{\partial E_\lambda}{\partial \lambda_k} = \sum_{i=1}^{m} a_{ki}\varphi_{ki}\left(x_0\right) - \alpha_k = 0; \quad k = 1, 2, \ldots, s+1 \tag{9}$$

Finally, the above system of linear equations with $k\times(m+1)$ equations and $k\times(m+1)$ unknowns is solved to get the values of model coefficients.

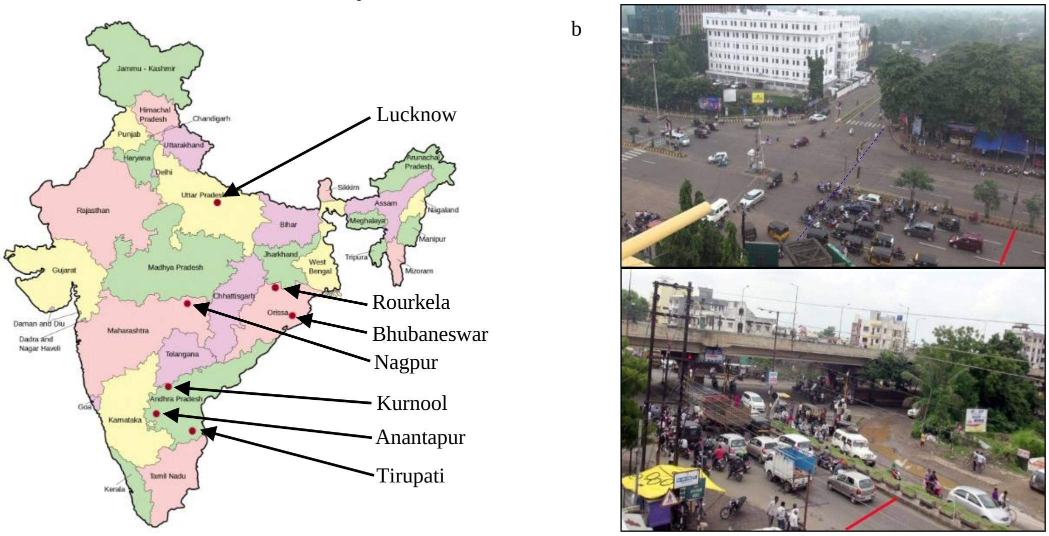
4. Site Selection and Data Collection

The collection of sufficient amount of data from diversified road conditions is the key requisite for the appropriate analysis of bicycling environments. In India, the urban development along with the infrastructural growth unevenly differ from one city to another. The locations of roadways inside a city also largely affect its traffic flow and built-environmental characteristics. Thus, as many as 70 intersection approaches located in the central parts, suburbs and outskirts of seven Indian cities were considered for data collection purposes, which include:

1. 25 approaches from Bhubaneswar city, the capital of Odisha state
2. 14 approaches from Lucknow city, the capital of Uttar Pradesh state
3. 7 approaches from Nagpur city of Maharashtra state, the largest city in central India
4. 6 approaches from Rourkela, the third largest city in Odisha state
5. 3 approaches from Kurnool, the seventh most populous city in Andhra Pradesh state
6. 12 approaches from Tirupati, the ninth most populous city in Andhra Pradesh state
7. 3 approaches from Anantapur, the tenth most populous city in Andhra Pradesh state

Fig. 3a shows the geographic locations of investigated cities, and Fig. 3b shows the conditions of two typical signalized intersection approaches in Indian context. Selected approaches are the legs of T-type or 4-legged isolated urban intersections, where number of lanes on the major and minor approaches vary in the ranges of 1–4 and 1–3 respectively. The peak hour volume on the major and minor approaches of these intersections vary in the ranges of 395–4085 and 208–2110 PCUs/h respectively. Considered approaches are characterised by 3–14 m carriageway, nonexistence to 5 m median, and non-existence to 3.5 m shoulders. Various other statistics of these sites are given in Table 2 of the “Variable Selection, Model Development, and Results” section. Following sub-sections discuss about the various data sets collected in this study.

a b



(The red colored solid line represents the end of queue, and blue colored dotted line represents the stop line at signalized intersection)

Fig. 3. (a) Location of studied cities; (b) typical intersection conditions prevailing in India.

4.1. Geometrical data

Listed below, several geometrical parameters are collected from each intersection, where all parameters are specific to the subject approach unless otherwise stated, and the direction of turning lanes (left or right) corresponds to the left-hand drive conditions prevailing in India:

- Number of lanes on the subject, opposing and conflicting approaches
- One-way or two-way (1 or 2)
- Effective width of the approach (m)
- Presence (1 = yes, 0 = no) and widths (m) of shared-use path, paved shoulder, bicycle lane, sidewalk, parking lane, crosswalk, median, curb and gutter

- Intersection crossing distance (m)
- Left-turn curb radius (m)
- Sight distance (1 = adequate, 0 = inadequate)
- Presence of exclusive left-turn lane (1 = yes, 0 = no)
- Allowance for right-turn on red (1 = yes, 0 = no)
- Number of legs at the intersection

Of the above parameters, the left-turn curb radius was estimated from google earth, while all others were collected through field observations and geometrical measurements. The effective width of each approach was estimated as the total width of the approach including travel lanes, paved shoulder, and the width of paving between outermost lane stripe and outer edge of pavement minus the average width reduction due to encroachments, if any.

4.2. Traffic flow and signalisation data

The videography technique was implemented to collect various information on the signalization and traffic flow characteristics at studied intersection approaches. At a particular time, one approach of an intersection was selected for the data collection. As demonstrated in Fig. 3b, a longitudinal trap was made on the selected approach from the stop line of the intersection to the end of the queue to notify the entry and exit of vehicles at intersections. This aided the accurate estimation of several parameters including bicycle delay and platoon ratio (the ratio of percentage of vehicle arriving during green and percentage of time green). The length of queue was decided by visually observing queue lengths in a few randomly selected cycles. The video camera was mounted on a nearby high-rise building, foot-over bridge, or any other rigid structure to record the unobstructed view of the trap on subject approach simultaneously with signal timings (if operated with traffic signals) and the traffic flows from conflicting and opposing approaches.

The video recording was carried out for a period of two hours either during the morning or evening peak periods of Indian traffic (8:30–10:30 A.M. or 4:30–6:30 P.M.). Recorded videos were played on a large screen, and manual traffic volume count was carried out. The traffic volume on the subject approach was converted to PCUs/h by using the conversion factors given in *Indian Road Congress* (IRC) code of practice-106 (IRC, 1990). Subsequently, the peak one hour at each site was determined through the running average method. This hour was chosen as the analysis period as bicyclists encounter the most complex operational conditions during this time. Subsequently, various other parameters listed below were extracted from the peak one-hour video of each site, where all parameters are specific to the subject approach unless otherwise stated and the direction of turning movements correspond to the left-hand drive conditions prevailing in India:

- Green period g (sec), cycle length C (sec), and the green time over cycle length g/C
- Compositions and volumes of the through, left-turning, right-turning, and total traffic (PCUs/h) on all approaches
- Volume of turning vehicular traffic across the path of through bicyclists (PCUs/h)
- Conflicting traffic volume (PCUs/h)
- Average platoon ratio
- Average stopped time delay incurred by through bicyclists (s/bicycle)
- Crossing pedestrian volume across the path of through bicyclists (ped/h)

4.3. Built-environmental and other data

As the built-environmental attributes are spatial in nature, those were observed within at least 100 m upstream of the stop line at an intersection approach. The list of all built-environmental and other data sets collected in this study include:

- Visibility of signaling devices (1 = good, 2 = moderate, 3 = poor)
- Surrounding developmental pattern (1 = highly commercial, 0.5 = moderately commercial, 0 = minimal commercial)
- On-street parking turnover (1 = high, 0.5= moderate, and 0 = minimal)

- Pavement Condition Index (varies from 1 = excellent through 5 = worst)
- Street-lighting condition (1 = good, 2 = moderate, 3 = poor)
- Number of driveways

4.4. Opinion survey of on-site bicyclists

The face-to-face interaction method was implemented to gather various information on the socio-demographic details, travel-related characteristics, and perceived BLOS scores of on-site bicyclists. At each site, effective responses were obtained from as high as 200 bicyclists (14,000 in total) comprising of employees, homemakers, students and retirees. Bicyclists those who recently crossed the intersections participated in this survey. To obtain reliable responses from the participants, following two criteria were laid: (1) the exclusion of children under the age of 14 from participation, as they may lead to improper responses, and (2) the inclusion of participants having at least one year of bicycling experience in urban areas. The percentage variations of various important characteristics of surveyed bicyclists observed across all sites are as follows:

- Gender: 47–49% females, and 51–53% males
- Age: 31–33% young (14–25 years old), 56–59% middle-aged (26–60 years old), and 9–11% elderly (above 60 years old)
- Bicycling experience: 19–28%, 42–56% and 18–27% have 1–5, 5–10 and over 10 years of experience respectively
- Daily average bicycling distance: 24–32%, 46–59%, 11–17% and 1–3% ride 1–5, 5–10, 11–20, and more than 20 km/day respectively
- User type: 39–44% regular and 55–61% occasional bicyclists

As per census 2011, 48% of the total Indian population are females, while 52% are males. And, the total Indian population of 14 years old and over includes approximately 33% young, 57% middle-aged and 10% elderly persons. Thus, there was a considerable similarity between the survey sample demographics and national demographic statistics. Necessary information regarding the socio-demographic details and travel-related characteristics of participants were collected with the help of a simple questionnaire sheet. Participants were asked to rate the intersection approaches on a 6-point Likert scale where, 1 represents that the participants were extremely satisfied while crossing the intersection and 6 represents that they were extremely dissatisfied with it. These ratings were denoted as perceived BLOS scores. Total 14,000 BLOS scores were collected in this study, which had the mean and standard deviation of 3.64 and 1.02 respectively. The Cochran's sample size formula (Cochran, 1997) was applied on this database to calculate the allowed errors in the estimation of mean perceived BLOS score. Results obtained at 95% confidence level concluded that the anxious error was considerably minimal and was well below 1%. This indicated that the amount of BLOS ratings collected in this study were sufficient for statistical analyses and BLOS model development.

5. Variable Selection, Model Development, and Results

The preliminary step of data analyses involves the identification of variables having significant influences on the perceived BLOS scores at investigated facilities. In the model development process, only significant variables are used as the set of dependent variables, while the insignificant variables are excluded from further considerations. The database included both continuous and ordinal (or categorical) variables. Thus, the Spearman's correlation analysis was preferred for the identification of significant variables as it is suitable to deal with both continuous and ordinal variables. During this analysis, a wide range of variables related to the characteristics of intersection approaches and surveyed bicyclists was used as the set of independent variables, while perceived BLOS scores were used as the set of dependent variables. The variables related to bicyclists' socio-demographic and travel characteristics were subjective in nature. Thus, suitable ordinal or categorical scales were defined to assess the influences of these variables on the perceived BLOS. For instance, the gender was defined as 0 = female and 1 = male. The age group was defined as 1 = young age, 2 = middle age, and 3 = elder age. The bicycling experience was defined as 1 = less than 10 years, 2 = 11–20 years, and 3 = more than 20 years. The daily average bicycling distance was defined as 1 = less than 20 km/day, 2 = 20–50 km/day, 3 = 50–80 km/day, and 4 = more than 80 km/day. The user type was defined as 0 = regular bicyclist and 1 = occasional bicyclist.

The Spearman's correlation analysis revealed that, seven several attributes of the interaction approaches have significant ($p < 0.001$) influences on the facility BLOS. These variables include effective width of the intersection approach (W_{Eff}), peak hour volume on the approach (PHV), crossing pedestrian volume across the path of through bicyclists (CPV), volume of turning vehicular traffic across the path of through bicyclists (V_{Turn}), average stopped time delay incurred by through bicyclists (D), on-street parking turnover (PT), and surrounding developmental pattern (SDP). On the other hand, the socio-demographic and travel-related characteristics of bicyclists did not have significant correlation with the perceived BLOS score. This concluded that the bicycle users irrespective their personal and travel-related characteristics perceived similar kind of service levels in the present context. Thus, the respective variables were excluded from further considerations. The descriptive statistics (range, mean and standard deviation) alongside the Spearman's ρ and associated p-values of all significant variables are summarized in Table 2. These descriptive statistics ensure that the identified parameters are collected from widely diversified road environments, and the database is capable of developing a well-generalized BLOS model.

Table 2. Important attributes of signalized approaches and their statistics.

Sl. No.	Variable	Unit or scale	Range	Mean	Standard deviation	Spearman's ρ	p-value
1	W_{Eff}	m	3–14	8.84	2.73	-0.483	< 0.001
2	PHV	PCUs/h	395–4086	1609.75	831.74	0.466	< 0.001
3	CPV	ped/h	33–1700	409.15	370.81	0.676	< 0.001
4	V_{Turn}	PCUs/h	69–703	288.8	140.05	0.462	< 0.001
5	D	s/bicycle	15–52.2	27.29	8.79	0.569	< 0.001
6	PT	Scale-1*	0–1	0.46	0.34	0.625	< 0.001
7	SDP	Scale-2**	0–1	0.51	0.43	0.563	< 0.001
8	$BLOS_{SA}$***	Varies in the range of 1–6	1.30–5.70	3.64	1.02	–	–

Note: *Scale-1 varies as 1 = high, 0.5= moderate and 0 = minimal, **Scale-2 varies as 1 = highly commercial, 0.5 = moderately commercial, 0 = minimal commercial, and ***$BLOS_{SA}$ represents the overall perceived BLOS for a signalized approach.

A positive ρ-value indicates that the perceived BLOS score increases with increasing values of the concerned variable (i.e., service quality decreases); while a negative value indicates the reverse. Thus, it can be concluded from Table 2 that only W_{Eff} has a positive influence on the bicycle service quality, while all other parameters adversely influence the same. CPV, PT, and D are by far the most influencing variables in the present context as they have produced the highest three ρ-values. Two other variables namely, g/C and opposing traffic volume also had significant correlation with perceived BLOS score. However, these two were highly correlated with D with ρ-values of above 0.8. Thus, D was only included in the model building process as it had the higher correlation with perceived BLOS score as compared to others ($\rho = 0.569$).

5.1. Model development

As the socio-demographic and travel-related characteristics of bicyclists did not have significant correlation with the perceived BLOS score, it was more meaningful to use the overall perceived scores ($BLOS_{SA}$) obtained for individual approaches as the array of output variable instead of the perceived BLOS scores of individual participants. This reduced the initial database of 14,000 observations to 70 observations in total. Here, one observation was retained for each intersection approach. Of these observations, 70% were used for model training, and the reaming 30% were used for model testing. The splitting of total data into training and testing groups was done in such a way to cover wide variations in both groups. In other words, care was taken to retain similar statistical properties (range, mean and standard deviation) of each variable in both groups. To accomplish this, a MATLAB (MathWorks, Inc., 2016) function namely, dividerand was used in this study. Subsequently, model development processes were carried out and the obtained results are presented below.

5.1.1 Regression-based BLOS model: The best regression model was obtained with the following configuration: PHV/W_{Eff}, ln($V_{\text{Turn,}}$)×(1+SDP), CPV×(1+PT), and D^2. The numerical values, statistics and significance of the coefficients obtained for these terms are given in Table 3. As observed, all terms (including the constant parameter) are associated with minimal standard errors, and their *t*-statistics are highly significant at the 95% confidence level (p-value < 0.05). The mathematical expression of the evolved BLOS model is shown in Equation 10. This model has produced a satisfactory R^2 value of 0.821 with training data sets.

$$\begin{aligned}\text{BLOS}_{\text{SA}} &= 1.1344 + 0.0019\times\left(\frac{\text{PHV}}{W_{\text{Eff}}}\right) + 0.1226\times\ln\left(V_{\text{Turn}}\right)\times\left(1+\text{SDP}\right)\\ &+ 0.00064\times\text{CPV}\times\left(1+\text{PT}\right) + 0.00085\times D^2\end{aligned} \tag{10}$$

Table 3. Regression model terms and statistics.

Model term	Coefficient	Standard error	*t*-statistic	*p*-value
Constant	1.1344	0.2798	4.0533	0.0002
PHV/W_{Eff}	0.0019	0.0006	3.0409	0.0039
ln(V_{Turn})× (1+SDP)	0.1226	0.0313	3.9109	0.0003
CPV×(1+PT)	0.00064	0.0001	4.6502	0.0000
*D*2	0.00085	0.0001	6.0759	0.0000

5.1.2 GP-based BLOS model: Optimum results for the GP analysis were obtained with a population size of 1400 individual generated for 300 times, $d_{\max}$ of 8, tournament size of 7, and p_{r}, p_{c} and p_{m} values of 0.02, 0.84 and 0.14 respectively. The gene structure and its mathematical expression obtained for the optimal BLOS model is presented in Fig. 4. The values of gene weight (*w*) and the bias parameter (*b*) estimated at 95% significance level ($p < 0.05$) are 0.0051 and -1.64 respectively. Thus, Equation 1 was used to derive the mathematical expression of the optimum BLOS model as follow, which produced a high R^2 value of 0.906 with the training data:

$$\begin{aligned}\text{BLOS}_{\text{SA}} &= -1.64 + 0.0051\times\ln\left\{\text{PHV}\times\text{CPV}^4\times D^8\times\left(V_{\text{Turn}}+D\right)^2\times\left(\text{W}_{\text{Eff}}-D\right)^4\right\}\\ &\times\left[\left\{\left(\text{PHV}+D\right)^{0.5}\times\left(\text{CPV}-4.3396\right)^{0.5} - V_{\text{Turn}}\times\text{SDP}^3 + V_{\text{Turn}}\times\text{SDP}\times\exp\left(\text{PT}\right)\right\}^2\right]\end{aligned} \tag{11}$$

5.1.3 FN-based BLOS model: All input and output variables were normalized within the range of [0, 1] to minimize their dimensional effect in the model building process. Further, an associative FN model, shown below, was derived from Equations 2 and 3 to predict the normalized values of overall BLOS scores ($\text{BLOS}_{\text{SA, Norm}}$) within the range of [0, 1].

$$\text{BLOS}_{\text{SA, Norm}} = \sum_{j=1}^{s}\left(\sum_{i=1}^{m} a_{ji}\phi_{ji}\left(x\right)\right) \tag{12}$$

Where, *s* is the number of input variables, and all other notations are as defined earlier.

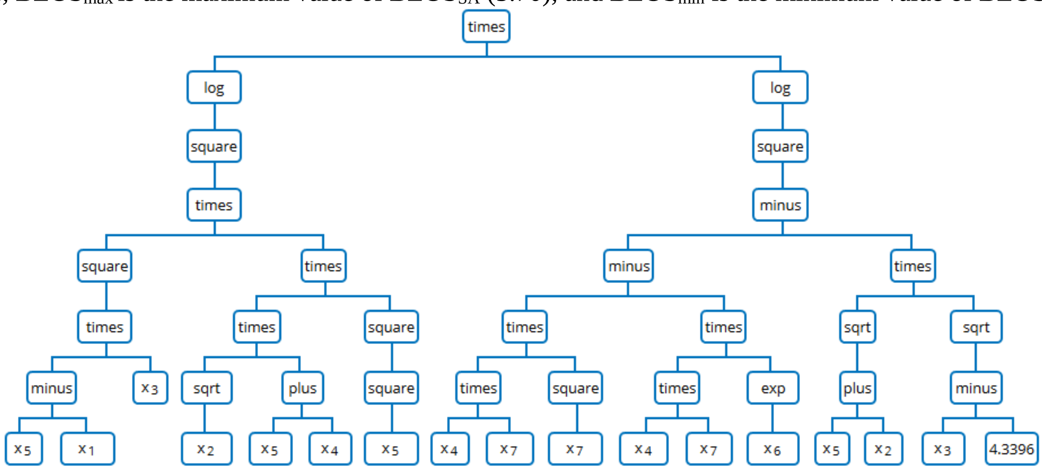
For the present problem, the prediction performances of five SFs were investigated namely, polynomial, exponential, sin(.), cos(.) and tan(.). The maximum acceptable degree of SF was limited to ‘two’ to keep the structural expression of the BLOS model as simple as possible. At this degree, the listed SFs produced R^2-values of 0.917, 0.905, 0.899, 0.880, and 0.878 respectively. Thus, the polynomial function was selected for model development. The values of constant parameter and coefficients of each input variable at degrees one and two were estimated and inputted in Equation 12 to obtain the following BLOS model, which has produced a high R^2 value of 0.917 with training data sets:

$$\begin{aligned} \mathrm{BLOS}_{\mathrm{SA,\,Norm}} &= 0.1704 - 0.3488 \times W_{\mathrm{Eff}} + 0.2222 \times W_{\mathrm{Eff}}^{\;2} + 0.5452 \times \mathrm{PHV} - 0.4125 \times \mathrm{PHV}^2 \\ &+ 0.8535 \times \mathrm{CPV} - 0.5001 \times \mathrm{CPV}^2 - 0.2062 \times V_{\mathrm{Turn}} + 0.4492 \times V_{\mathrm{Turn}}^{\;2} + 0.5412 \times D \\ &- 0.1865 \times D^2 - 0.0366 \times \mathrm{PT} + 0.1035 \times \mathrm{PT}^2 + 0.1477 \times \mathrm{SDP} - 0.0662 \times \mathrm{SDP}^2 \end{aligned} \tag{13}$$

In the above equation, both input and output variables have their normalised values in the range of [0, 1]. The de-normalised value of $\mathrm{BLOS}_{\mathrm{SA}}$ can be estimated as:

$$\mathrm{BLOS}_{\mathrm{SA}} = \mathrm{BLOS}_{\mathrm{SA,\,Norm}} \times \left(\mathrm{BLOS}_{\mathrm{Max}} - \mathrm{BLOS}_{\mathrm{Min}}\right) + \mathrm{BLOS}_{\mathrm{Min}} \tag{14}$$

Where, $\mathrm{BLOS}_{\mathrm{max}}$ is the maximum value of $\mathrm{BLOS}_{\mathrm{SA}}$ (5.70), and $\mathrm{BLOS}_{\mathrm{min}}$ is the minimum value of $\mathrm{BLOS}_{\mathrm{SA}}$ (1.30).



Where, $x_1 = W_{\mathrm{Eff}}$, $x_2 = \mathrm{PHV}$, $x_3 = \mathrm{CPV}$, $x_4 = V_{\mathrm{Turn}}$, $x_5 = D$, $x_6 = \mathrm{PT}$, and $x_7 = \mathrm{SDP}$

$$\ln\left\{\mathrm{PHV} \times \mathrm{CPV}^4 \times D^8 \times \left(V_{\mathrm{Turn}} + D\right)^2 \times \left(\mathrm{W}_{\mathrm{Eff}} - D\right)^4\right\}$$
$$\times \left[\left\{\left(\mathrm{PHV} + D\right)^{0.5} \times \left(\mathrm{CPV} - 4.3396\right)^{0.5} - V_{\mathrm{Turn}} \times \mathrm{SDP}^3 + V_{\mathrm{Turn}} \times \mathrm{SDP} \times \exp\left(\mathrm{PT}\right)\right\}^2\right]$$

Fig. 4. Gene structure of the GP-based BLOS model.

5.2. Goodness-of-fit, testing, and ranking of developed models

The prediction performances of developed BLOS models are presented in Figure 5, 5a for training data and 5b for testing data. As observed, the $\mathrm{BLOS}_{\mathrm{SA}}$ values predicted by GP and FN models are very much close to the ideal line of fit (linear fit of concatenate), while the regression model has comparatively inferior performance. Various statistical parameters are applied to assess and rank the developed models in terms of their goodness-of-fit and testing results. These statistical parameters include R^2, Nash–Sutcliffe efficiency coefficient (E), average absolute error (AAE), mean absolute percentage error (MAPE), and root mean square error (RMSE). R^2 and E measure the best-fit criteria of the model. Values of these parameters close to 'one' signify that the predicted and observed values of the output variable are very much close to each other. The maximum values of these parameters are ‘one’, which are attained under ideal conditions only. On the other hand, AAE, MAPE and RMSE are the error measuring parameters, which measure various types of errors produced between the predicted and observed values of the output variable. Thus, their values are expected to be minimal for a good prediction model. Table 4 summarizes the prediction performances of all BLOS models in terms of these statistical parameters. As observed, both GP and FN models have over-performed the regression model with higher values of R^2 and E, and lower values of error measuring parameters. The performances of both regression and GP models are good in training, but not as good in testing. However, the FN model has shown excellent performances with both data sets. Thus, the traffic planners and engineers may use this model to assess the efficiency of signalized intersections. Other two models may also be used for simpler calculations keeping in mind that the prediction results would be marginally inferior.

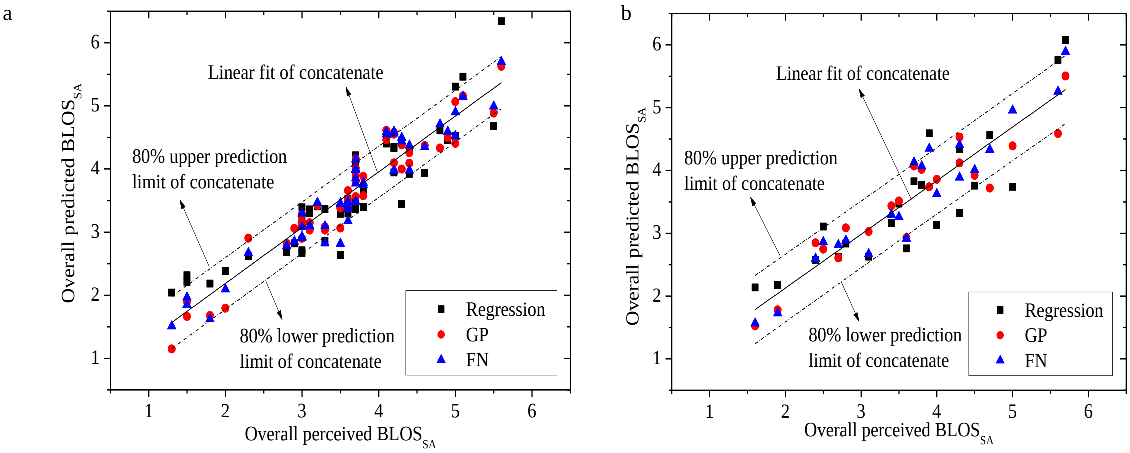


Fig. 5. Prediction performances of signalized intersection models with (a) training data; (b) testing data.

Table 4. Prediction precision and ranking of BLOS models.

Model	Data	R^2	E	AAE	MAPE (%)	RMSE
Regression-based	Training	0.821	0.821	0.350	11.478	0.414
	Testing	0.767	0.747	0.414	11.967	0.547
GP-based	Training	0.906	0.906	0.234	6.959	0.284
	Testing	0.869	0.844	0.321	8.385	0.429
FN-based	Training	0.917	0.917	0.227	7.005	0.282
	Testing	0.915	0.909	0.281	7.851	0.327

5.3. Comparison with existing model

The prediction performances of developed models are compared with a regression-based BLOS model previously developed by Beura et al. (2017). Other existing models discussed under “Background studies” section are not included in the comparison process as those are developed for the homogeneous traffic flow environment and it is obvious that those models will perform poorly in the present context. The prediction results of Beura et al. (2017) model with all data sets are as follows: R^2 = 0.761, E = 0.679, AAE = 0.432, MAPE = 13.469 %, and RMSE = 0.557. Thus, all newly developed models have over-performed this model in terms of higher R^2 and E, and lower prediction errors. Beura et al. (2017) model is trained and tested with limited quantity of data sets (25 and 10 observations respectively). Thus, the generalization ability of the model to the vast urban parts of a developing county is in doubt. On the other hand, the BLOS models proposed in this study are trained and tested with as high as 49 and 21 observations. These data sets are collected from widely diversified road conditions prevailing in India. Thus, the analyses of bicycling environment carried out in the present study are more in-depth. Two new parameters namely, turning vehicular volume and crossing pedestrian volume are identified to have considerable influences on the BLOS of signalized intersection approaches, which are not considered in the Beura et al. (2017) model. Thus, the major advantages of newly developed models include the better prediction ability and the consideration of all essential parameters.

5.4. Estimation of BLOS classes (A–F)

A service scale has been defined to convert the values of $BLOS_{SA}$ to letter-graded service classes A–F. The BLOS criteria have been classified into six levels (A–F) in order to remain consistent with various previous studies including TRB (2000, 2010). The linguistic descriptions of all BLOS classes are presented in Table 8. As the perceived BLOS ratings were collected on a scale of 1–6, the values of $BLOS_{SA}$ vary in the range of 1–6. Thus, the mean value of this scale (3.5) corresponds to the boundary between BLOS classes ‘C’ and ‘D’. This means a $BLOS_{SA}$ value below 3.5 corresponds to one of the service classes ‘A–C’, and a $BLOS_{SA}$ value above 3.5 corresponds to one of the service classes ‘D–F’. Hence, symmetrical cutoffs were made on both sides of 3.5 to define the ranges of BLOS classes ‘A’ through ‘F’. Obtained results are presented in Table 8. Similar stratification concepts are also previously documented in Beura and Bhuyan (2017) and several other studies.

Table 5. Prediction precision and ranking of BLOS models.

BLOS class	Service level	Ranges of $BLOS_{SA}$
A	Excellent	$\leqslant$ 1.5
B	Very good	1.5 – 2.5
C	Good	2.5 – 3.5
D	Fair	3.5 – 4.5
E	Poor	4.5 – 5.5
F	Very poor	> 5.5

6. Conclusions

Road intersections carrying heterogeneous traffic are probably the most complex points of the bicycle operation. The service levels of these critical points from bicycling perspective are assessed by using suitable BLOS models. However, all existing BLOS models are solely based on homogeneous traffic conditions and are not transferable to developing countries where the road traffic is heterogeneous in nature. To fill this gap partially, this study has analyzed the quality of bicycle through movement at signalized intersections. The Spearman’s correlation analysis has revealed that the BLOS of this facility is determined by seven quantitative attributes, while personal characteristics of bicyclists do not have significant influences on the same. The observed list of significant attributes includes W_{Eff}, PHV, CPV, V_{Turn}, D, PT, and SDP. Of these variables, W_{Eff} is observed to have positive influence on the facility BLOS while all other variables adversely influence the same. CPV, PT and D (having Spearman’s ρ values of 0.676, 0.625 and 0.569 respectively) are observed to be by far the most important indicators of signalized intersection BLOS.

By incorporating the aforementioned variables as model inputs and the array of $BLOS_{SA}$ scores obtained at individual sites as model output, highly reliable BLOS models are developed using three highly efficient techniques namely, step-wise regression, GP and associativity FN. For the present problem, FN has produced the best prediction results with the highest R^2 values of 0.917 and 0.915 with training and testing data sets respectively. On the other hand, the regression technique has produced the least efficient but simplest and easiest to implement BLOS model. Thus, the FN-based model is highly desirable for field applications, while prediction accuracies are of utmost concerns and the regression model is highly desirable while computational efforts are of utmost concern. These models would largely help the traffic planners and engineers to quantify and enhance the service qualities urban signalized intersections operating under heterogeneous traffic conditions. The utilization of widely diversified data sets (Table 2) for the development and testing of these models also ensure that those are well-transferable to the vast majority of developing cities around the world.

Field implementations of the FN-based BLOS model (the most efficient one) along with the service scale defined in Table 5 has revealed that above 86% of all investigated approaches are offering BLOS ‘C’ or inferior. Thus, essential cares should be taken by the city authorities to achieve better service levels, and to satisfy the future demands. The most influencing variables, mentioned earlier, could be largely prioritized in the planning process to achieve better service levels effortlessly. Amongst the limitations of this study, proposed models may not provide expected prediction precision in highly crowded metropolitan cities, as the scenarios are more complex. In these cases, either the proposed models could be used with required calibrations or similar new models could be developed if the choice of significant variables alters.

References

Beura, S.K., Bhuyan, P.K., 2017. Development of a bicycle level of service model for urban street segments in mid-sized cities carrying heterogeneous traffic: a functional networks approach. Journal of Traffic and Transportation Engineering (English Edition) 4.6, 503–521.

Beura, S.K., Kumar, N.K., Bhuyan, P.K., 2017. Level of service for bicycle through movement at signalized intersections operating under heterogeneous traffic flow conditions. Transportation in Developing Economies 3.2.

Carter, D., Hunter, W., Zegeer, C., Stewart, J., Huang, H., 2007. Bicyclist intersection safety index. Transportation Research Record: Journal of the Transportation Research Board 2031, 18–24.

Castillo, E., Cobo, A., Gómez-Nesterkin, R., Hadi, A.S., 2000a. A general framework for functional networks. Networks 35.1, 70–82.

Castillo, E., Cobo, A., Gutiérrez, J.M., Pruneda, E., 2000b. Functional networks: A new network-based methodology. Computer-Aided Civil and Infrastructure Engineering 15.2, 90–106.

Castillo, E., Iglesias, A., Ruiz-Cobo, R., 2004. Functional equations in applied sciences. Mathematics in science and engineering 199. Elsevier, Amsterdam.

Cochran, W.G., 1997. Sampling techniques. John Wiley & Sons, Inc., New York.

Crider L. B., Burden, J., Han, F., 2001. Multimodal LOS-point level of service project, final report. Department of Urban and Regional Planning, University of Florida, Gainesville, Florida.

Davis, J., 1987. Bicycle safety evaluation. Auburn University, City of Chattanooga and Chattanooga-Hamilton County Regional Planning Commission, TN.

IRC (Indian Road Congress)., 1990. Guidelines for capacity of urban roads in plain areas. Indian Road Congress-106, New Delhi.

Jensen, S.U., 2013. Pedestrian and bicycle level of service at intersections, roundabouts and other crossings. Transportation Research Board’s 92nd Annual Meeting, Washington, D.C., 19p.

Koza, J.R., 1992. Genetic programming: On the programming of computers by means of natural selection, MIT Press.

Landis, B., Vattikuti, V., Ottenberg, R., Petritsch, T., Guttenplan, M., Crider, L., 2003. Intersection level of service for the bicycle through movement. Transportation Research Record: Journal of the Transportation Research Board 1828, 101–106.

MathWorks, Inc., 2016. MATLAB user’s manual 9.0. MathWorks, Inc., Natick.

McPhee, N.F., Poli, R., Langdon, W.B., 2008. A field guide to genetic programming, Lulu Enterprises, UK Ltd.

Searson, D.P., Leahy, D.E., Willis, M.J., 2010. GPTIPS: An open source genetic programming toolbox from multi-gene symbolic regression. International Multi-Conference of Engineers and Computer Scientists, I, Hong Kong.

Steinman, N., Hines, K., 2004. Methodology to assess design features for pedestrian and bicyclist crossings at signalized intersections. Transportation Research Record: Journal of the Transportation Research Board 1878, 42–50.

Strauss, J., Miranda-Moreno, L.F., Morency, P., 2013. Cyclist activity and injury risk analysis at signalized intersections: a Bayesian modelling approach. Accident Analysis & Prevention 59, 9–17.

TRB (Transportation Research Board)., 2000. Highway Capacity Manual (HCM), Transportation Research Board of the National Academies, Washington, D.C.

TRB., 2010, Highway Capacity Manual (HCM), Transportation Research Board of the National Academies, Washington, D.C.