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

Comparison of Artificial Intelligence Based Roundabout Entry Capacity Models

Ashish Kumar Patnaik[a], Rajrani Patra[b], Prasanta Kumar Bhuyan[c,*]

[a]*Research Scholar, Department of Civil Engineering, National Institute of Technology Rourkela, India. Email: ashishkumar.patnaik@yahoo.in*
[b]*Former M.Tech Student, Department of Civil Engineering, National Institute of Technology Rourkela, India. Email: rajranipatra25@gmail.com*
[c,*]*Assistant Professor, Department of Civil Engineering, National Institute of Technology Rourkela, India. Email:*pkbtrans@gmail.com

Abstract

This study focuses on the development of entry capacity models for roundabouts by using Genetic Algorithm (GA), Multi-variate Adaptive Regression spline (MARS) and Random Forest Regression (RFR) technique under heterogeneous traffic conditions. Required data were collected from 27 selected roundabouts spanning across 8 states of India by using high-defination video (HD) camera. To characterize the driver behavior, INAGA method is employed to find out the critical gap and follow up time. By employing Modified Rank Index (MRI), MARS based capacity model is found to be best fitted in this study. The coefficient of determination (R^2) and Nash-Sutcliffe model efficiency coefficient (E) are found to be (0.94, 0.88) and (0.94, 0.86) respectively. This indicates that the proposed model is statistically fit at 95% confidence level. To assess the contribution of each input variable in the developed model, sensitivity analysis is carried out in this study. It is found that variable like Entry width (E_w) contributes the most while the follow-up (T_f) time variable has less contribution in the proposed model. To assess the proposed model significance, the model is compared with internationally developed models like Brilon wu formula (Germany), Girabase Formula (France) and HCM 2010 respectively. These findings will be useful for traffic planners and designers in the capacity estimation of roundabouts under heterogeneous traffic conditions in developing countries with similar traffic characteristics as India.

Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY.

Keywords: Roundabout; GA; MARS; RFR; INAGA method; Critical gap; Follow up time; Artificial Intelligence (AI), Heterogeneous

1. Introduction

A roundabout is a modified form of an un-signalized intersection, where traffic movement is forced to move around a central circular island almost in one direction only. The circular traffic movement can follow either left hand rule or

* Corresponding author. Tel.: +91-661-246 2313; fax: +91-661-246 2313.
E-mail address: pkbtrans@gmail.com

2352-1465 © 2018 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY

right hand rule depending on the rules governed by their respective countries. Roundabouts have several advantages as compared to other intersections in terms of safety and operational efficiency. In terms of conflict points (safety), roundabouts have less number of conflict points as compared to signalized intersections. Roundabouts have 8 and 16 number of conflict points for one and two lane-based path respectively. Where signalized roundabouts have 32 number of conflict points. Apart from the reduction number of conflict points, other positive aspects of introducing roundabouts are speed reduction of vehicles, reduction of seriousness of fatal accidents, reduction in delay time, increasing the safety of the pedestrians and controlling the environmental pollution etc. With an increase in the traffic rate and road infrastructure day by day, it is essential to assess the performance of road infrastructure like roundabout under heterogeneous traffic conditions. Capacity is the important parameter under operational performance that describes about the present traffic condition. Capacity is the maximum entry flow that can be accommodated in the approach stream under prevailing traffic and geometric conditions (HCM, 2010).

In India, different kinds of vehicles are plying on roads with varying in terms of their physical structure, composition, size and operational performance. In addition, there is difference in driver behavior and geometric layout of intersection from place to place. It was ascertained that individual percentage contribution of vehicles is not more than 80% of total traffic, that indicates that traffic situation in India is highly heterogeneous in nature (Patnaik et al., 2017). The vehicle fleet consist of heavy vehicles (HV), light motor vehicles (LMV), motor cycles (MC), scooters (S) and non-motorized vehicle like bicycles. Under constraints of flow and space, the small sized vehicles merge into the circulatory area without following the lane discipline and heavy vehicles forcibly enter into the circulating stream of traffic. These facts indicate that the rule of priority is not accounted under heterogeneous traffic condition. Traffic flow is characterized as homogeneous in nature in developed counties like USA, UK, Australia and Germany etc. because uniform nature of driver behavior and vehicle characteristics. In India, two-wheeler percentage contribution is about (40-50) percentage of total traffic. Therefore, the gaps are randomly maintained between consecutive vehicles. Under these circumstances, the gap acceptance behavior is highly complex in nature. Thus, homogeneous traffic condition ignore some important characteristics of heterogeneous traffic condition.

2. Background literature

Broadly, two types such as empirical and gap acceptance based modeling approaches have been given by the previous studies for the capacity estimation of roundabouts. In empirical regression based models, entry capacity is taken as the dependent variable, whereas the circulating flow and other geometric variables are taken as independent variables (Al-Masaeid and Faddah 1997; Patnaik et al. 2017a, 2017b, 2018a, 2018b). The gap acceptance capacity model considers the driver's decision to move from the minor stream of traffic flow to the major stream of traffic flow. Critical gap and follow up gap are two important variables used for the development of this model. The HCM (2000) method of capacity estimation is one of the most popular method that follows the concept of gap acceptance. Further enhancements have given in the HCM (2010) in which detailed procedures for the determination of capacity and LOS are described. This model is a combination of gap acceptance and exponential regression concepts and calibrated by estimating critical gap and follow up gap.

Limited number of studies have been carried out to estimate capacity of roundabouts under the influence of heterogeneous traffic flow in Indian context. Ahmad and Rastogi (2016) developed roundabout capacity model and observed that the entry capacity varies negative exponentially with the increase in circulating flow. Arroju et al. (2015) developed a micro simulation model by using VISSIM software for roundabouts in which vehicles were classified and their critical gaps were evaluated by using Raff and Equilibrium of Probabilities method. Mathew *et al*. (2016) estimated the passenger car unit (PCU) for various categories of vehicles based on the concept of time occupancy at four-legged roundabouts. The critical gap and follow up time values are estimated to calibrate the HCM 2010 model equation and a multiplicative adjustment factor is recommended for the direct use of HCM 2010 model. Ahmad *et al.* (2015) proposed an iterative procedure based on the minimization of sum of absolute difference in critical gap for the determination of capacity at roundabouts.

In order to develop the roundabout entry capacity models, Rodegerdts et al.(2007) observed that prior knowledge regarding the relationship between dependent and independent variables are required and found difficulties while investigating the relationship due to scatter type of data. In the recent past, researchers found that Artificial Intelligence (AI) based methods predict the output more accurately as compared to regression based models. Gupta et al. (2017) measured the performance of plate fin heat exchanger by using the different AI techniques and found that Adaptive Neuro Fuzzy Inference Systems (ANFIS) model predicts the output more accurately in comparison to Artificial Neural Network (ANN), Genetic Algorithms (GA) and Simulated Annealing (SA) models. Ozuysal *et al.* (2009) estimated the roundabout capacity of Turkish roundabouts by employing ANN model and observed that the model produced better results than regression and gap acceptance models. Faris (2013) developed cutting machine temperature model via genetic programming symbolic regression approach. It was observed that, GP model is capable of predicting data more accurately in comparison to the other models developed by Least Squares regression (LS), Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) respectively.

It could be summarized from the available literatures that the geometry of a roundabout plays a crucial role in determining its performance. Many researches have been carried out based on this aspect, which resulted in rigid models for different nations. When compared with other models, these models based on empiricism showed better results in some cases, while underestimate at cases, especially, when used in different countries. Models formulated based on geometry of roundabouts along with flow characteristics give better explanation on performance prediction. Drivers' behaviour and attitude to accept/reject gap, which contribute to flow characteristics at roundabouts, are assumed consistent under homogeneous traffic flow conditions. Critical gap is an important parameter for the development of gap acceptance based roundabout capacity model. Inaccurate estimation of critical gap leads to biased capacity value. Existing methods regarding critical gap estimation do not give consistent results when no rejection of gaps found in data sets. However, under the heterogeneous traffic flow condition, behaviour of driver varies depending upon the dimension of vehicles plying on roads, adequacy of road geometric, information signposts and road markings provided at roundabouts. Hence, to overcome this limitation, recently a new method known as INAGA is developed for the estimation of the critical gap (Patnaik et al. 2017). In this study, trapezoidal shape of Influence area is assumed in the conflicting zone where the conflict between entry flow and circulating flow is maximum. Then, gap acceptance capacity model is developed by accounting critical gap, follow up time and circulating flow as explanatory variables. In the recent past, Artificial intelligence based models have gained wide recognition regarding prediction problems. Hence, the present study furnishes strong background regarding modelling approaches usually followed using Genetic Algorithm (GA), Multi-variate Adaptive Regression spline (MARS) and Random Forest Regression (RFR) techniques. These methods help in predicting capacity more accurately in comparison to gap acceptance and empirical models applied to a new data set.

3. Study Objectives and Organization of the Paper

The prime objectives of this study are to develop the entry capacity models for an un-signalized roundabout. With an overview of heterogeneous traffic conditions, the following study objectives are provided below.

- Development of capacity prediction models by employing Artificial Intelligence based methods such as GA, MARS and RFR.
- To analyze the relative importance of input variables in the model developments through the application of sensitivity analysis.
- To assess the significance of developed models, compare the models and select the most suitable one by employing error estimation techniques.

The content of the paper is categorized as follows. The section "Methodology" describes about the critical gap and follow up time estimation by using newly developed INAGA method. A brief introduction about modelling techniques such as Genetic Algorithm (GA), Multi-variate Adaptive Regression spline (MARS) and Random Forest Regression (RFR) is added in that section. In addition, the study area and data collection procedure is included in that section. Development of three roundabout entry capacity models by using GA, MARS, and RFR techniques, ranking of proposed models, percentage contribution of each input variables by sensitivity analysis and comparison of models are included in the "Results and Discussion" part. The last section reports about the "Conclusion" in which the application, limitation and direction for further research are also discussed.

4. Methodology

The study is categorized into three parts. The first part deals about the study area and data collection process. Determination of gap acceptance variables like critical gap and follow up time by using INAGA method is given in second part. Then the details about modelling techniques like GA, MARS and RFR is included in the last part.

4.1. Study areas and Data collection procedure

In order to develop the roundabout entry capacity models, it is essential to have a proper and adequate dataset. Therefore, 27 roundabouts were considered for the purpose of reconnaissance. These roundabouts were situated in twelve different cities of eastern, western, northern, southern and central India in the details given in Fig. 1. Roundabouts were chosen based on location in the city, geometrics and traffic composition. Roundabouts were taken for study area based on certain conditions like it should be free from bus stops, road side parking and side friction. The impacts of pedestrians were ignored in this study. Roundabouts were at grade intersections having 4 numbers of approaches which are perpendicular to each other. Selected roundabouts are uncontrolled in nature. Also, Roundabouts were chosen on criteria like 2 numbers of entry and circulating lanes respectively. These above mentioned criteria have taken into considerations for selecting the study sites.

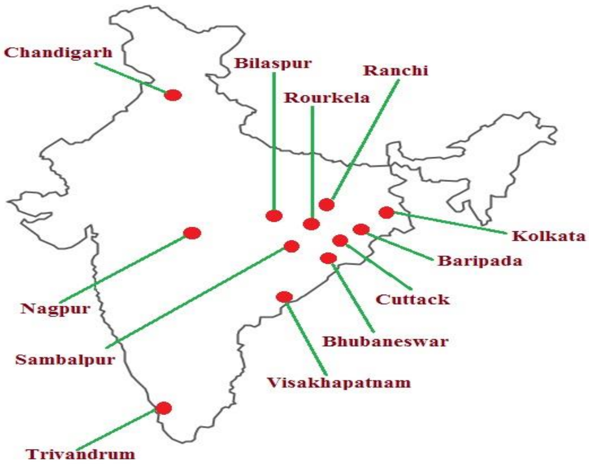


Fig. 1. Location of cities on Indian map

Three types of data were collected such as gap acceptance, traffic flow and geometric details for the development of roundabout entry capacity models. Gap acceptance variables like critical gap and follow up time are estimated by using INAGA method. Geometric data include weaving length (W_l), weaving width (W_w), entry width (E_w), Diameter (D), entry angle (ϕ) and approach width (A_w) which were measured by measuring tape during off peak hours of traffic flow. The data collection was carried out in the months of September and October of 2014. These months are considered as normal months for which the traffic flow is least affected by environmental influences during this period. The videos were captured either from morning 9 am to 11 am or from evening 5 pm to 7 pm on a typical clear weekday. Sixty hours of videos were recorded for this study and these videos were displayed in office computer to extract the desired data. After getting the gap acceptance, road geometrics and traffic flow data, Artificial Intelligence technique (AI) like GA, MARS and RFR are applied to develop the roundabout entry capacity models.

To reflect the heterogeneity, vehicles were classified into five types such as heavy vehicles (HV), light motor vehicles (LMV), motorcycles/scooters (MC/S), bicycles and animal/human drawn vehicles (ADV/HDV). Then, the total number of entry flow and circulating flow are converted into static PCU as given on IRC: 65-1976. PCU factors are generally found out by considering the vehicle categories and their composition, manoeuvrability and also the speed of the vehicles, which varies from one country to another (Mauro, 2010). Composition of traffic flow is given in Table 1. below.

Table 1. Investigation of dependent and explanatory variables in this Study

Variables	Units	Minimum	Maximum	Mean	SD.
Observed entry capacity(Q_e)	PCU/h	250	3415	1819.6	774.85
Circulating flow (q_c)	PCU/h	219	3688	1071.29	721.52
Weaving length (W_l)	m	28.96	58.62	44.49	6.15
Weaving width (W_w)	m	8.2	20.05	17.90	5.69
Entry width (E_w)	m	5	20.12	15.18	3.90
Diameter of central island (D)	m	10.76	60.12	43.63	11.78
Critical gap (T_c)	seconds	0.54	2.87	1.83	0.59
Follow up time (T_f)	seconds	0.92	2.99	1.58	0.53

Note: SD=Standard deviation of the sample size

The traffic composition varies from (0.26 to 61) %, (10 to 79.5) %, (14.55 to 79) %, (0.05 to 33.1) %, (0.1 to 7) % for HV, LMV, (MC/S), Bicycle, (HDV/ADV) respectively. It is observed from the analysis that individual percentage (%) of each category of vehicles is not exceeding 80%. Hence, it is assumed that the traffic flow is a heterogeneous traffic in developing countries (Arasan and Krishnamurthy, 2008).

4.2 INAGA method

Gap acceptance variables such as critical gap and follow up time play key role for the development of roundabout entry capacity models. These variables help in determining the local driver behaviour under a given condition. Existing methods such as the latest developed equilibrium probabilities and traditionally followed Raff method assumed that driver behaviour is consistent and traffic flow is homogeneous for estimation of critical gap. But these methods do not give reliable results under no rejection of gap data. Hence, this fact is treated as a major issue under heterogeneous traffic flow conditions. Generally, no gaps are rejected for heavy and free flow traffic under heterogeneous traffic flow conditions. Hence, to address such problem and heterogeneity, INAGA method is used for the determination of critical gap and follow up time in this study (Patnaik et al., 2017). By taking this concept, the critical zone (Trepezoidal shape) is assumed in the most conflicting path of roundabout where the interactions among the entry flow and circulating flow is maximum.

4.3 GA technique

Genetic algorithm (GA) is a metaheuristic inspired by the process of natural selection that belongs to the larger class of evolutionary algorithms (EA). The finite length of binary code strings of one (1) and zero (0) are used to describe the parameters for each solution. Initially an unknown function is given to get an optimal solution. The optimal solution consisting of terminal set and functional set, which is known as a tree based hierarchical structure. Fitness function is represented as an absolute error in GA regression to reduce the total error of data samples. During each successive generation, a portion of the existing population is selected to breed new generation. Individual solution are selected through a fitness-based process, where the best-fit individuals are more likely to be selected. In the study, tournament selector is used for selection of data points. The next step is to generate second generation population of solutions from those selected through a combination of genetic operators like mutation and crossover. Elite transfer is used to allow a relatively small number of the fittest programs in order to keep the best solution found so far. As a result, a new population of trees of the same size as the original one is created, but it has a higher average fitness value. The whole process converges to a population where the 'weak' individuals have eliminated and the 'fittest' ones survive. The process is repeated until the best individual in population converges to an optimal solution.

4.4 MARS technique

Multivariate adaptive regression spline (MARS) is an artificial intelligence (AI) technique for modelling the data where some local sections of each independent variable are considered as a discrete region. For each independent variable, some break points are defined where the relationship between the response and predictor variable changes, commonly known as 'knot'. MARS technique is generally run in the two computations such as forward pass (FP) and backward pass (BP) by using the training data sets. FP is performed on the training data sets with only intercept by using the help of least square error. At each subsequent steps, that basis function pair which produces the maximum error reduction in the training error is added. Process of adding pair is continued until the change in residual error is too small or maximum number of iterations has reached. In this process for each independent variable, functions are added in the form of pair irrespective of their fitness to the model, hence FP leads to form an over-fitted model. After the completion of FP, BP is allowed to run on the training data. It removes the less significant terms one by one at each step until it finds best sub-model. Model subsets are compared using Generalized Cross Validation (GCV) criterion (Hastie et al., 2009), a weighted mean squared error criterion to prevent over-fitting. After development of optimal model, Statistical parameters such as RMSE, R^2 are calculated using training and testing data.

4.5 RFR technique

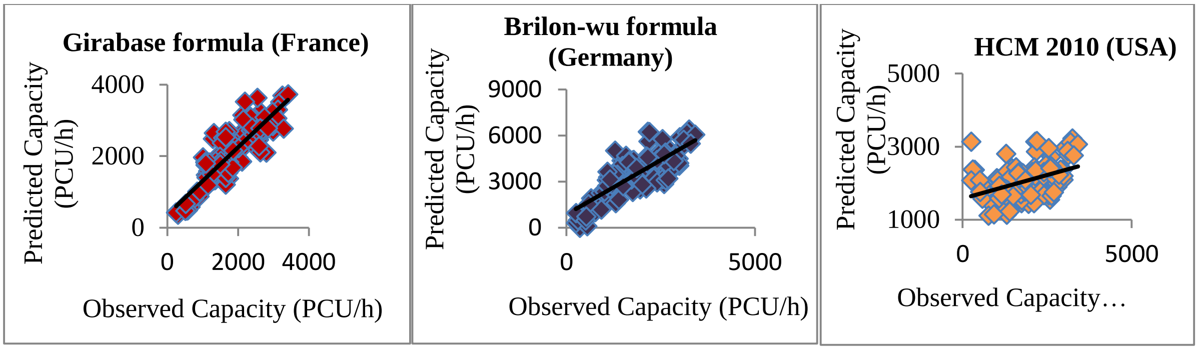
Random forest regression (RFR) is an another technique in the field of AI in which each decision tree is guarded by some features. The final prediction is given by the sum of the feature contribution and bias (the mean given by the most effective variables which almost covers the training data set). In this technique, the prediction in a forest is the average prediction of its decision trees. Calculation of variance is very much essential for a regression tree in order to choose the best split. The split with lower variance is selected as the criteria to split the population. Bootstrap Aggregating or Bagging is a best method of ensemble. This method is used to reduce the variance of the final prediction model by combining the results of multiple models derived from different sub-samples of the same data set. Sampling is done on the original training data sets to form a new data set by taking points at random with replacement. So, each bag carries an equal amount of data points. Out of "n" number of bags, "n" number of split nodes are selected based on lower variance and higher bias. Using the best split nodes, regression model is formed by containing the mean value of their respective independent variables. In order to get the final model, the predictions of all the bags are combined using their mean values. As the final model is the combined model, hence it is generally known to be more robust than a single model.

5. Results and discussion

Initially the existing models are evaluated to ascertain the prediction ability of developed models. The development of three entry capacity models for roundabouts by utilizing GA, MARS and RFR techniques, selection of best model by employing modified rank index (MRI) and sensitivity analysis of proposed model under heterogeneous traffic conditions are depicted in this section.

5.1 Existing model evaluation

To assess the prediction ability of existing entry capacity models; Girabase formula (France), Brilon-wu formula (Germany) and HCM 2010 (USA) models are applied to a new set of roundabouts. Mauro (2010) examined that there is possible to establish the relationship between capacity, other geometrics and gap acceptance variables as explanatory variables. Hence the data collected from field observations are compared with the predicted capacity obtained by above methods which are shown in Fig. 2 below.



(a) **(b)** **(c)**

Fig. 2. Prediction of Capacity by using International Capacity Models

It is observed from the Fig. 2 (a) and (b) that, the predicted entry capacities of Girabase and Brilon-Wu methods are comparatively higher to capacities observed from field conditions. However, from the Fig. 2 (c) it is observed that the predicted capacities are relatively lower compared to the observed from field conditions. This difference in capacity values can be attributed because of two factors. The first contributing factor is due to the heterogeneous traffic flow conditions on Indian roads. Driving behaviour of Indian road users is totally different in comparison to other developed countries assumed as the second factor. Under heterogeneous traffic flow conditions, traffic comprise different kinds of vehicles with different manoeuvrability. Indian traffic is predominantly shared by two wheelers rather than four wheelers and large sized vehicles. Two wheelers in the approaching stream forcibly create a gap towards the circulating stream of traffic flow, which violates the principle of priority. But under homogeneous traffic, vehicles flow in a stream line manner and maintain gap consistently. Models developed under homogeneous traffic flow conditions have over prediction of entry capacity under high circulating flows which leads to inaccurate fitting of these models when applied with heterogeneous traffic flow data sets. Difference in capacity predictions can be minimized by measuring the driver behaviour variables such as critical gap and follow up gap at local conditions. Also it is acknowledged that geometric variables are important that should be considered for the entry capacity prediction models. Recently developed Artificial Intelligence (AI) techniques acquired wider appreciation over regression-based models for prediction of values. Hence, GA, MARS and RFR models are developed via integrating both geometric and gap acceptance variables as explanatory variables under heterogeneous traffic flow conditions.

5.2 Development of GA, MARS and RFR models

Before developing the model, the preliminary assessment is carried out to identify the dependency of individual variables w.r.t. dependent variable. Therefore, Pearson-correlation test is conducted in this study, in details given in Table 2 below. Pearson co-relation defines the degree of linear dependence between two variables.

Table 2. Pearson correlation among variables

Model variables	Q_e	W_l	T_c	W_w	E_w	q_c	T_f	D
Q_e	1	….	….	….	….	….	….	….
W_l	0.87	1	....	….	….	….	….	….
T_c	0.883	-.150	1	….	….	….	….	….
W_w	0.918	-.165	-.436	1	….	….	….	….
E_w	0.893	-.331	.003	-.329	1	….	….	….
q_c	-0.888	.082	.088	.260	.083	1	….	….
T_f	-0.866	.055	.124	.006	.201	-.515	1	….
D	-0.613	-.123	.075	-.091	.225	.116	-.077	1

It is observed from the Table 2 that the linear dependence between the explanatory variables and dependent variable varies as 61% to 91.8% that defines the good correlation ship established among dependent and explanatory variables. In order to develop the Genetic programming based models, it is assuming that the variables are free from the effect of multi- collinearity. Means the explanatory variables are not establishing significant relationship (Highly correlated) among them.

It is observed from the above mentioned Table 2 that, the explanatory variables are not establishing relationship among them, which indicates that the model holds good under heterogeneous traffic condition. Hence by considering the above seven variables, GA, MARS and RFR regression modeling were carried out in MATLAB software. In this study, a total of 110 number of data points (observations from each approach leg of roundabouts) were used. Out of 110 data points, 77 data points (representing 70% of the data) were used for training the model and the rest were used for testing.

GA model

GA regression model is developed by formulating mathematical expressions and tree based structure. These expressions or tree based structures match perfectly with the dependent and independent variables to get the desired output. The solution functions are represented by the tree based hierarchical structured computer programs. The function comprises basic algorithm functions which are represented as +, -, *, / etc. The size and shape of the trees can change in every successive iteration. The iteration process continues until the error is minimal. Hence, several iterations were carried out in GA symbolic regression to get the best model. The best GA model was obtained at the mutation probability of 5%. The variables and values are shown in the Table 3 below. Also, the model was obtained at the maximum generation of 1000 cycles for which the mathematical expression of the model and the tree based structure is represented in Equation 1 and Fig. 3. respectively.

Table 3. Variables and values of GA model

Variables	**Value**
Mutation probability	5%
Population size	100
Maximum generations	1000
Selector	Proportional selector
Elites	1
Operators	+, -, *, /, power, log, root
Analyzer	Multi analyzer
Cross over	Subtree swapping cross over
Mutator	null

$$Q_e = \left(\left(\left(c_0.E_w.c_1.E_w + \left(\frac{\left(c_2.T_c + \left(c_3.T_c + c_4.T_c\right)\right)}{c_5.T_f} + c_6.E_w \right) .c_7.W_l \right) .c_8 + c_9 \right) \right) \quad (1)$$

$[R^2 = 0.89, RMSE = 243.45]$

Note: *RMSE*= Root mean square error

Where, c_0 = 0.37545, c_1 = 0.37545, c_2 = 1.8597, c_3 = 1.8597, c_4 = 1.8597, c_5 = 1.941, c_6 = 0.37545, c_7 = 0.70293, c_8 = 5.2837, c_9 = 0.4668.

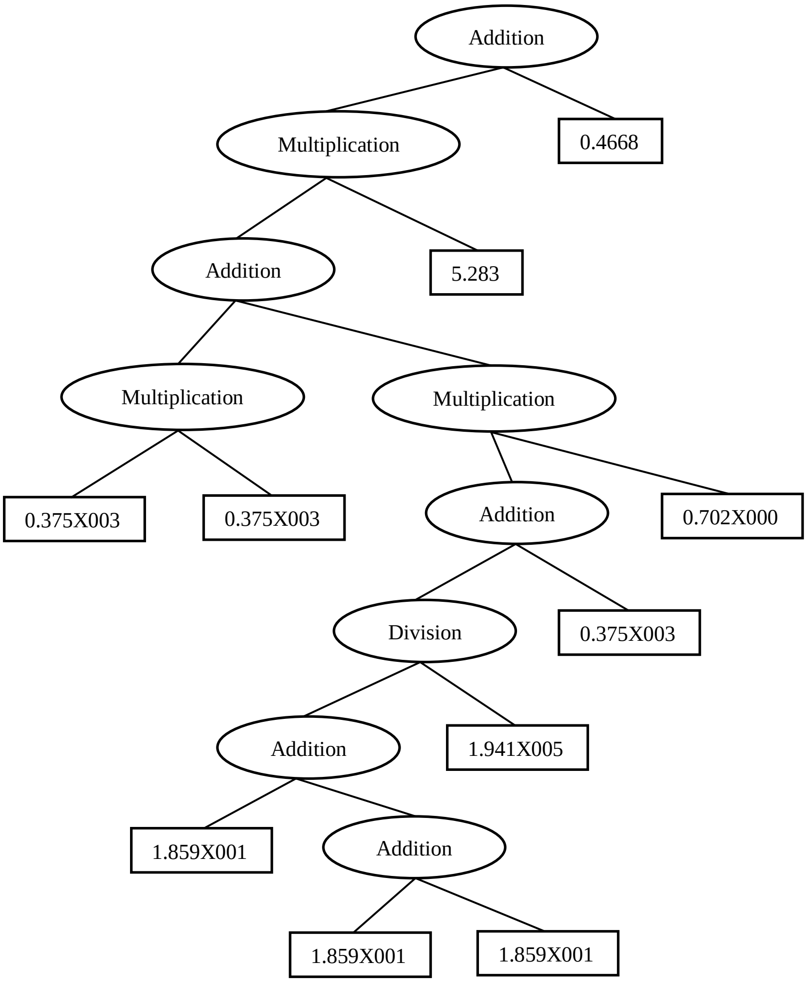


Fig. 3. Tree based structure of GP model

NOTE:-

X000= W_l = Weaving length
X001= T_c = Critical gap
X002= W_w = Weaving width
X003= E_w = Entry width
X004= q_c = Circulating flow
X005= T_f = Follow up time
X006= D = Diameter of central island
X007= Q_e = Entry capacity

MARS model

MARS model is obtained by using the output as dependent variable and independent variables as predictor variables respectively. MARS input variables are taken as based on “basis function”. There are seven independent variables for the development of roundabout entry capacity model. Each variable is associated with two basis function. The basis functions are selected by using the combination of following parameters.

Maximum number -15
GCV (Generalized Cross Validation) - 7
MSE (Mean Squared Error) - 9
DOF (Degree of Freedom) penalty – 3

In Forward pass (FP) each bias functions are found in pairs as it doesn’t involve any short of best fit operation. So FP takes into account all 7 pairs of basis functions. But in Backward Pass (BP), MARS accounts for error terms (RMSE and MSE) and selects only those basis functions with lesser error and better fitting to the model. Individual basis functions which don’t fit into the model are discarded. Hence finally there are nine basis functions in the above mathematical expression of MARS model (some of them are in pair and some are not) in details given below.

$$y = 339.68 + 35.87 \times BF_1 + 19.86 \times BF_4 - 0.127 \times BF_5 + 0.902 \times BF_6 + 55.50 \times BF_7 + 247.97 \times BF_8 + 18.32 \times BF_9 - 330.80 \times BF_{12} + 332.68 \times BF_{14} \quad (2)$$

Where $y = \text{Entry capacity } (Q_e)$

Note: BF = Basis function
$BF_1 = \max(0, \mathrm{D} - 45.91);$
$\mathrm{BF}_4 = \max(0, 21.44 - W_w);$
$BF_5 = \max(0, q_c - 624);$
$\mathrm{BF}_6 = \max(0, 624 - q_c);$
$\mathrm{BF}_7 = \max(0, E_w - 4.43);$
$BF_8 = \max(0, T_c - 0.54);$
$BF_9 = \max(0, W_l - 28.96);$
$BF_{12} = \max(0, W_w - 27.9);$
$BF_{14} = \max(0, W_w - 27);$

Model $Q_e = BF_1 BF_4 BF_5 BF_6 BF_7 BF_8 BF_9 BF_{12} BF_{14}$ (3)

[R^2 = 0.94, Adj. R^2 = 0.93, GCV R^2 = 0.88]

RFR model

A RFR technique is used to develop the roundabout entry capacity model. The technique is optimized by combination of following three parameters such as ntree, mtry, nodesize.

ntree: Number of trees produced from bootstrapped sample (ntree = 3000).

mtry: Predictor numbers tested randomly at each node (mtry = 7).

nodesize: The minimal size of terminal node (nodesize = 1).

The dependent variable is log-transformed to develop the superior model (Death and Fabricius, 2008). The model produces an unbiased error estimate via bootstrapping. The following procedures have been followed to develop the capacity model. For "nth" regression tree, the training data is selected by leaving one third of the cases out, that is out of bag (OOB) data. In each case, the constructed tree is evaluated for the "ith" case.

For each "i" an average value predicted in "i" test by employing the mean square evaluation.

$$MSE_{OOB} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}_{iOOB})^2 \qquad (4)$$

By using the same procedure, several iterations are carried out to develop the roundabout entry capacity model for which the co-efficient of determination (R^2) and root mean square error (RMSE) is obtained as 0.94 and 188.24 respectively. The limitation of the RFR model is that the model is capable of predicting values, but the equation based model cannot be built up due to ''black box'' system (Giustolisi et al., 2007).

Fig. 4. gives the scatter plot between observed capacity and predicted capacity of GA, MARS and RFR models respectively. It is observed that the line of equality is within the 95% prediction limit for roundabout entry capacity models. The coefficient of correlation (R^2) is observed to be (0.89, 0.90), (0.94, 0.93) and (0.94, 0.88) in both training and testing stage for GA, MARS and RFR models respectively. These indicate a strong correlation (R^2>0.80) exists in the datasets (Smith, 1986). However, it is also known that R^2 is the biased estimate (Das and sivakugan, 2010). Hence to determine the efficiency (significance test) of these models and to select the best model among three "modified rank index" is applied in this study.

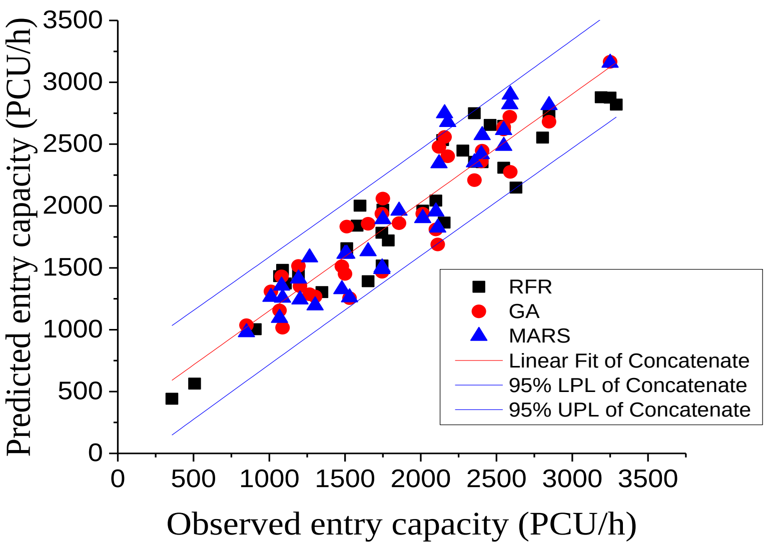


Fig. 4. Scatter plot between observed entry capacity and predicted entry capacity

5.3 Selection of suitable model by using modified rank index

In this study, the overall prediction assessment of three proposed entry capacity models by utilizing GA, MARS and RFR techniques are carried out using several statistical parameters. The results are summarized in Appendix 1. Subsequently, the models are ranked by using a "modified rank index" (MRI) parameter (Beura and Bhuyan, 2017). The Equation 5 shows the mathematical expression of MRI.

$$MRI = R_1 + R_2 + R_3 + R_4 + R_5 \tag{5}$$

R_1 = ranking of a model based on best-fit calculations such as: coefficient of determination (R^2) and Nash–Sutcliffe model efficiency coefficient 'E'.
R_2 = ranking of a model based on error measuring parameters such as average absolute error (AAE), maximum absolute error (MAE) and root mean square error (RMSE).
R_3 = ranking of a model based on arithmetic calculations (mean 'μ' and standard deviation 'σ') of ratio of predicted and observed entry capacities.
R_4 = ranking of a model based on 50% and 90% cumulative probability (P_{50} and P_{90}) values obtained from cumulative probability plot of C_P/C_O (ratio of predicted and observed entry capacity) values, and
R_5 = ranking of a model based on prediction of entry capacity within ±20% accuracy level, calculated using the histogram and lognormal probability distribution of C_P/C_O (ratio of predicted and observed entry capacity) values.

From the literatures, it is ascertained that when the predicted entry capacity is exactly equal to the field observed entry capacity, the mean 'μ' and standard deviation 'σ' of C_P/C_O become 1 and 0 respectively. Therefore, these values closer to 1 and 0 respectively indicate better performance of a predictive model. P_{50} value below and above '1' implies under and over predictions respectively. P_{90} parameter reflects the variation in C_P/C_O values for total observations. Thus, values of both P_{50} and P_{90} parameters closer to '1', indicate better performance of a predictive model. MRI value as obtained for each entry capacity model developed in this study is shown in the Appendix 1. It was observed from the analysis that the MARS model has MRI value of "8" and showing the final rank of "1" that reflects the best performance in this study. Otherwise, RFR (rank-2) and GA (rank-3) models have shown relatively inferior performance under the given data set.

5.4 Sensitivity analysis of MARS model

It is observed from the above analysis that, MARS model is found to be best fitted under heterogeneous traffic flow conditions. Hence, in order to assess the percentage contribution of each input variable in MARS model, Garson's algorithm (Gandomi et al., 2013) is applied in this study. The ranking of each input variable is based upon percentage of contribution in the model development, which is given in Table 4 below.

Table 4. Percentage contribution of each variable in MARS model

Sl. No.	Variables	Sensitivity (%)	Rank
1.	Weaving length (W_l)	15.26	4
2.	Critical gap (T_c)	16.17	3
3.	Diameter of central island (D)	14.27	5
4.	Weaving width (W_w)	8.74	6
5.	Entry width (E_w)	24.37	1
6.	Circulating flow (q_c)	21.19	2
7.	Follow-up gap (T_f)	0	Not affecting

It is observed from the above Table 4 that, three input variables such as entry width (E_w), circulating flow (q_c) and critical gap (T_c) are contributing 24.37 %, 21.19 % and 16.17 % respectively in the model fitting. These three variables primarily representing the behaviour of traffic under heterogeneous traffic flow conditions are all-together contributing approximately 60% to the entry capacity prediction. The remaining four variables (D, W_l, W_w and T_f) which are related to the geometric condition of the roundabout altogether contribute approximately 40% to the capacity prediction developed in this study. It is also observed that follow up gap (T_f) acted as non-contributing variable in the model development by the application of Garson algorithm respectively. Conclusion part of this study is addressed in the next section.

6. Conclusions

- At selected roundabouts, traffic flow is observed to be highly heterogeneous, in which percentage share of Bicycles, Two Wheelers, Light Motor Vehicles (Car, Three Wheelers), Animal Drawn Vehicles and Heavy Vehicles varies from 0.06% to 36.51%, (6.84-77.9)%, (10.52 to 78.9) %, (0.04-7.49) % and (0.14-15.07) % respectively.
- To reflect the actual driver behavioural habits under heterogeneous traffic conditions, gap acceptance variables such as critical gap and follow up time are determined by using the INAGA method developed in this study. The critical gap values are varying from 0.54 seconds to 2.87 seconds, which are nearly half the values of (4-4.6) seconds as mentioned for developed nations like USA and European countries. These differences in critical gap values are due to the large percentage share of two-wheelers in the prevailing conditions. Actually, two wheelers require small gap as compared to other vehicles to merge into the major stream of traffic flow.
- The GA model is obtained at the mutation probability of 5% with maximum generation of 1000 cycles. MARS model considered a default penalty value of 3 and GCV value of 7. RFR model did not give any mathematical expression though it gives the highest coefficient of determination value that is 0.94. Due to non-equation based model, the practical application of RFR model seems to be poor.
- The co-efficient of determination (R^2) and Nash-Sutcliffe co-efficient (E) values of proposed models like GA, MARS and RFR are found to be (0.90, 0.90), (0.94, 0.94) and (0.94, 0.94) that signifies that the proposed model is statistically significant at 95% confidence level.
- Based on the MRI index, MARS model is found to be best fitted under heterogeneous traffic flow conditions. The other two models such as RFR and GA are placed in 2nd and 3rd rank respectively in this study.
- Sensitivity analysis reports that the variables like entry width is the prominent variable in the study and almost contributing 25% to the developed model. In addition, it is observed that follow up gap variable is found to be non-contributing variable in the developed model.

The proposed MARS model provides versatile information to practitioners, traffic engineers as an effective tool for estimating roundabout entry capacity with similar conditions of traffic in other developing countries. The proposed MARS model is easy to use and manageable from a practical point of view. Impact of pedestrian crossing was not taken into consideration for the development of the roundabout entry capacity model. Also, the entry capacity models were developed by using the static PCU factors mentioned in IRC: 65-1976. Hence future studies can be considered for the development of roundabout entry capacity models by incorporating the impact of pedestrian crossing and dynamic PCU values respectively.

Advantages of MARS model

- It performs in efficient manner for large number of predictor variables.
- Automatically establish the relationship between variables.
- It is an efficient and fast algorithm, despite its complexity
- Robust to outliers

References

Ahmad, A., Rastogi, R., 2016. Regression Model for Entry Capacity of a Roundabout under Mixed Traffic Condition– an Indian Case Study. Transportation Letters. doi: http://doi.org/10.1080/19427867.2016.1203603.

Ahmad, A., Rastogi, R., Chandra, S., 2015. Estimating of Critical gap on a Roundabout by Minimizing the Sum of Absolute Difference in Accepted gap data. Canadian Journal of Civil Engineering 42.12, 1011-1018.

Al-Masaeid, H, Faddah, M., 1997. Capacity of Roundabouts in Jordan. Transportation Research Record: Journal of the Transportation Research Board 1572: 76–85, https://doi.org/10.3141/1572-10.

Arasan, V.T., Krishnamurthy, K., 2008. Effect of Traffic Volume on PCU of Vehicles under Heterogeneous Traffic Conditions. Road Transportation Research ARRB 17. 1, 32–49.

Arroju, R., Gaddam, H.K., Vanumu, L.D., Rao, K.R., 2015. Comparative Evaluation of Roundabout Capacities under Heterogeneous Traffic Conditions. Journal of Modern Transportation 23.4, 310–324. DOI 10.1007/s40534-015-0089-8.

Beura, S.K., Bhuyan, P.K., 2017. Modelling Quality of Bicycle Accommodations on Urban Road Segments using Functional Networks and Multivariate Adaptive Regression Splines Techniques. Transportation Research Board's 96th Annual Meeting, Washington D.C., Compendium of Papers CD-ROM, Paper Number: 17-01262.

Das, S.K., Sivakugan, N., 2010. Discussion of Intelligent Computing for Modeling Axial Capacity of Pile foundations. Canadian Geotechnical Journal 47.8, 928–930. https://doi.org/10.1139/T10-048.

De'ath, G., Fabricius, K.A., 2000. Classification and Regression Trees: A Powerful yet Simple Technique for Ecological Data Analysis. Ecology 81, 3178–3192.

Faris, H., 2013. A Symbolic Regression Approach for Modeling the Temperature of Metal Cutting Tool, International Journal of Control and Automation, Vol.6 (4).

Gandomi, A.H., Yun, G.J., Alavi, A.H., 2013. An Evolutionary Approach for Modeling of Shear Strength of RC deep Beams. Materials and Structures 46.12, 2109–2119. doi:10.1617/s11527-013-0039-z.

Giustolisi, O.A., Doglioni, D.A., Savic, B.W., 2007. A Multimodel Approach to Analysis of Environmental Phenomena. Environmental Modelling and Software 22.5, 674–682. https://doi.org/10.1016/ j.envsoft.2005.12.026.

Gupta, A.K., Kumar, P., Sahoo, R.K., Sahu, A.K., Sarangi, S.K., 2017. Performance Measurement of Plate Fin Heat Exchanger by Exploration: ANN, ANFIS, GA, and SA. Journal of Computational Design and Engineering 4.1, 60-68. https://doi.org/10.1016/j.jcde.2016.07.002.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data mining, Inference and Prediction. 2nd edition, Springer.

HCM 2000. Highway capacity manual. Transportation Research Board, Special report 209, Washington, D.C.

HCM 2000. Highway capacity manual. Transportation Research Board, Special report 209, Washington, D.C.

IRC 65:1976: Recommended Practice for Traffic Rotaries. Indian Road Congress.

Mauro, R., 2010. Calculation of Roundabouts: Problem Definition. In: Calculation of Roundabouts. Springer, Berlin, Heidelberg, DOI: 10.1007/978-3-642-04551-6_1.

Ozuysal, M, Caliskanelli, S.P., Tanyel, S., Baran, T., 2009. Capacity Prediction for Traffic Circles: Applicability of ANN. Proceedings of the Institution of Civil Engineers – Transport 162.4, 195-206. https://doi.org/10.1680/tran.2009.162.4.195.

Patnaik, A.K, Krishna, Y., Rao, S., Bhuyan, P.K., 2017. Development of Roundabout Entry Capacity Model Using INAGA Method for Heterogeneous Traffic Flow Conditions. Arabian Journal for Science and Engineering 42.9, 4181–4199. doi:10.1007/s13369-017-2677-x.

Patnaik, A.K., Chaulia, S., Bhuyan, P.K., 2018. Roundabout Entry Capacity Models: Genetic Programming Approach. Proceedings of the Institution of Civil Engineers – Transport. https://doi.org/10.1680/jtran.17.00089.

Patnaik, A.K., Ranjan, A.R., Bhuyan, P.K,. 2018. Investigating Entry Capacity Models of Roundabouts under Heterogeneous Traffic Conditions. Transportation Research Record: Journal of the Transportation Research Board, https://doi.org/10.1177/0361198118777603.

Patnaik, A.K., Rao, S., Krishna, Y., Bhuyan, P.K., 2017. Empirical Capacity Model for Roundabouts under Heterogeneous Traffic Flow Conditions. Transportation Letters 9.3**,** 152-165, http://dx.doi.org/10.1080/19427867.2016.1203583.

Rodegerdts, L., Blogg, M., Wemple, E., Myers, E., Kyte, M., Dixon, M., List, G., Flannery, A., Troutbeck, R., Brilon, W., Wu, N., Persaud, B., Lyon, C., Harkey, D., Carter, D., 2007. Roundabouts in the United States”. National Cooperative Highway Research Program Report 572. Transportation Research Board of the National Academic. Washington. D.C., USA.

Smith, G.N., 1986. Probability and statistics in civil engineering: an introduction. Collins, London.

Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

Appendix 1. Significance test and ranking of proposed models

Model	Data	**Best Fit Calculations**			**Error Measuring Parameters**				**Arithmetic Calculations of P/O**			**Cumulative Probability of P/O**			**±20% Accuracy (%)**			**Overall Rank**	
		R^2	E	R1	AAE	MAE	RMSE	R2	μ	σ	R3	Ratio at P_{50}	Ratio at P_{90}	R4	Log-normal	Histogram	R5	MRI	Final Rank
GA	Training	0.90	0.90	3	203.01	637.77	254.3274	3	1.00	0.15	1	0.98	1.22	2	80.51	80.52	3	12	3
	Testing	0.86	0.86		183.04	423.42	221.286		1.04	0.13		1.02	1.24		84.36	81.82			
MARS	Training	0.94	0.94	2	170.35	517.63	204.01	2	1.01	0.13	2	0.99	1.16	1	85.31	89.61	1	8	1
	Testing	0.88	0.86		175.05	591.40	218.97		1.04	0.12		1.03	1.25		85.40	84.85			
RFR	Training	0.94	0.94	1	151.38	476.5	188.12	1	1.04	0.14	3	1.01	1.24	3	84.21	88.31	2	10	2
	Testing	0.88	0.88		215.66	480.92	256.49		1.04	0.14		1.02	1.24		80.37	81.82			

Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY

Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

R_1, R_2, $R3$, $R4$, R_5 = Ranking variables of a model
E = Nash–Sutcliffe model efficiency coefficient
AAE = Average absolute error
MAE = Maximum absolute error
$RMSE$ = Root mean square error
C_P = Predicted capacity
C_O = Observed capacity MRI = Modified Rank Index

Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY