ADVANCES IN TRANSPORTATION STUDIES
An International Journal

Editor in Chief: Alessandro Calvi

Vol. LXIV November 2024

Contents

aracne

Advances in Transportation Studies: an international Journal 64 (2024)

ISBN 979-12-218-1490-3 – ISSN 1824-5463 – DOI 10.53136/97912218149031 – pag. 3-24

Prediction of operating speed on horizontal curves of two-lane rural highways using artificial intelligence

K.K. Tottadi* A. Mehar

Department of Civil Engineering, National Institute of Technology, Warangal, India.
**corresponding author; email: kirant6666@student.nitw.ac.in; arpan@nitw.ac.in*

subm. 19th December 2023 approv. after rev. 3rd March 2024

Abstract

Horizontal geometric characteristics have significant impact on vehicle operating speed of vehicles on two-lane rural highways. Majority of the studies have used the conventional approach of modelling, found to be location specific and provide false judgement of determining operating speed. Thus, it becomes important to apply methods based on artificial intelligent for predicting operating speed of vehicles on two-lane roads under mixed traffic conditions. Field data was collected on 40 different locations (curves and tangent sections) that includes free speed of vehicles and geometric parameters. Geometric parameters such as curve radius, curve length, deflection angle, degree of curvature and preceding tangent length were measured in the field with total station, whereas free-flow speed data was collected using radar gun at the mid of the horizontal curves. The statistical analysis concluded that the curve radius, curve length, degree of curvature and preceding tangent length are found significant on the operating speed of vehicle type Car, Two-wheeler, Three-wheeler, Light commercial vehicle and Heavy commercial vehicle and developed MLR model. Further, the data driven soft computing methods such as Artificial Neural Network (AN), Adaptive Neuro-fuzzy Inference System (ANFIS) and Support Vector Regression (SVR) are applied to predict the operating speed of these vehicles and results were compared with MLR. The performance of the models evaluated using various goodness-of-fit measures indicates that the SVR model gives better results in prediction of operating speed in compared to other models. As for future research, further investigation could be conducted to explore uncertainties, and the model could be enhanced by utilizing other geometric and traffic parameters, and techniques like random forest, XGBoost etc.,

Keywords – operating speed, horizontal curves, MLR, ANN, ANFIS, SVR

1. Introduction

The operating speed is considered to be an essential factor for designing safer roadway and setting speed limits. Many researchers have identified various risk factors such as the driver, vehicle, road geometry configuration and environmental factors which can affect the operating speed of vehicles [1-3]. The horizontal curve is one of the elements of the roadway alignment involves higher probability of occurrence of crashes. According to the Ministry of Road Transport and Highways of India [4], 13% of the total 4,12,432 crashes takes place on horizontal curves. The main reason for crashes on horizontal curve is due to improper driver behaviour and poor coordination between geometric elements. Horizontal curves are usually designed based on the design speed concept with an assumption that the design speed of a vehicle attaining maximum speed on curve. But in reality, the vehicles may move on the curve with the same speed or at speed more or less than design speed. Fitzpatrick et al., [5] reported that the drivers choose their speeds based on the availability of road features rather than posted speed limits. The selection of speed of vehicle under free flow condition is defined as operating speed [6]. Some investigators recognized that the horizontal curves with a design speed less than 80 kmph [7-8] or less than 90 kmph [5, 9] had an operating speed that was more than the design speed and vice-versa. These studies clearly indicate that the design speed concept does not yield the actual speeds of vehicles as there is discrepancy between design speed and operating speed [10]. Hence, various researchers used and recommended an alternate approach i.e., operating speed concept for predicting actual speeds on the horizontal curves [6, 11-15].

Several studies have been conducted to investigate the effect of geometric characteristics on operating speed of vehicles at horizontal curves on two-lane rural highways. Majority of studies considered curve radius to be the most dominant factor for estimating operating speed on mid-point of the curve [16-20]. If the curve radius is more than 850m, it has no influence on the operating speed of vehicles [9]. Researchers found that curvature change rate [9, 19], degree of curvature [7, 20], super-elevation [21], deflection angle [18, 22-24], and length of the curve [18, 23, 25-26] impact the operating speed at the mid-point of horizontal curves. The independent variables (geometric parameters) used in past studies were significantly different from one study to another. Therefore, it is necessary to investigate the impact of geometric elements as the independent parameters for predicting operating speed of vehicles at mid-point of the horizontal curves.

In past studies, ordinary linear regression analysis was adopted for developing operating speed models under free flow conditions [27-28]. In recent years, several machine learning modeling techniques such as artificial neural networks (ANN), adaptive neuro-fuzzy inference systems (ANFIS), support vector machine (SVM), random forest (RF) etc., are few most effective applications being adopted to solve complex problems. The artificial neural network technique have been popularly used in operating speed studies conducted on horizontal curves [24, 29-30]. However, other modern techniques like ANFIS, or SVM have not been tested due to assuming their complexities in computation and interpretability while dealing with large data sets or high dimensional feature space. Hence, present study is an attempt to perform study by applying different machine learning techniques such as ANN, ANFIS and SVM along with conventional method MLR for perdition of operating speed of different vehicle types. The study also compares the result obtained from different methods in term of accuracy, precision and ease of use.

2. Literature review

Majority of researchers in the last four decades have employed multiple regression analysis for developing speed prediction models [6, 15-16, 22-24]. Maji et al. [29] used stepwise multiple regression technique for developing operating speed model for light Car, Light commercial vehicles and Heavy commercial vehicles. Authors used root mean square error for measuring the accuracy of the developed models. Sil et al., [17] deployed different multiple linear regression techniques like forward selection, backward elimination and stepwise regression for estimating the operating speed on curves and concluded that backward elimination is the best for such studies. Malaghan et al. [22] adopted all subset regression approach for developing operating speed models using data collected at two-lane highway sections. The different performance measures such as Akaike information criteria (AIC) and coefficient of determination are used for choosing the best model. Goyani et al. [15] used backward elimination stepwise regression method for predicting operating speeds on various points along the horizontal curve. The researchers considered different goodness-of-fit measures like mean absolute percentage error, root mean square error, and mean absolute deviation for validating the field data.

Some authors used advanced machine learning techniques like artificial neural network for developing operating speed models on highways. Najjar et al. [31] developed a relationship between geometric characteristics and 85th percentile speed using artificial neural network (ANN). The predicted operating speeds were used to establish posted limits on two-lane rural highways in Kansas. They considered different variables like shoulder width, shoulder type, average daily traffic and percentage of no passing zones as independent parameters. They concluded that the models developed were estimated operating speed with an accuracy of 96%. McFadden et al., [24] used back propagation artificial neural network (ANN) for developing operating speed models on horizontal curves. The ANN models were further compared with previously developed regression models with the same set of field data [7]. They reported that ANN models offered better predictive powers than regression models, and it may be considered as a possible alternative to linear regression technique. The researchers also suggested that the alternative architecture in addition to back propagation ANN i.e., competitive ANN learning technique determine the best prediction. Taylor et al. [32] developed models for estimating 15th, 50th, and 85th percentile speeds on high-speed highways using artificial neural network (ANN). They considered geometric features, traffic control measures and speed data for model development. Singh et al. [33] developed artificial neural network models for predicting operating speed as a function of geometric characteristics, traffic variables, pavement condition indices and crash data on two-lane highways. They concluded that V_{85} speed decreased with increase in crash rate, average daily traffic, skid number, and international roughness index whereas increases whereas it increased with carriageway width and shoulder width. Semeida [30] studied the effect of roadway characteristics and posted speed limit on operating speed. For this, they developed regression models and ANN models using various geometric variables, and a comparison was made between these models. The ANN model gives R^2 and RMSE equal to 0.978 and 3.11, and the regression model give R^2 and RMSE equal to 0.761 and 10.32. Finally, the author concluded that ANN models give better results than regression model. Semeida [34] developed operating speed prediction models for cars and trucks on horizontal curves using ANN. They found that the curve radius was most significant factor for predicting operating speed of car and the median width was considered as most influential factor for developing operating speed model for truck. Few authors applied ANFIS technique to estimate Passenger Car Unit (PCU) values at different level services and crowd speed based analysis [35-36]. Srikanth and Mehar [35] carried out the comparative study foe estimating PCUs at different level of services using MLR, ANN and ANFIS modeling techniques. They reported that the ANFIS model showed greater potential in predicting PCUs for all vehicle types than ANN and MLR models. Similarly, Yugender and Ravishankar [35] performed a study to find out the influencing parameters on crowd speed using MLR, ANN and ANFIS modeling techniques. They reported that the ANFIS model showed higher degree of accuracy than other models. Summarizing the literature, various modeling techniques are adopted in the applications of speed and hence, it is required to study the best prediction model among them.The objective of this study was to predict the operating speed of different vehicle types using MLR, ANN, ANFIS and SVR on the horizontal curves of two-lane rural highways.

3. Site selection

This study focuses on the two-lane rural highways in various parts of India. A total of 40 locations are selected for study from three different rural highways. The selected highway locations include 40 curved sections and 40 tangent sections, exhibiting diverse geometric features. Each location has a tangent section followed by a horizontal curve.

(a) (b)

Fig. 1 - Google map view (a) and camera view (b) of the study location at National highway NH 163

All selected sites are free from the influence of intersections, industrial and commercial activities, entry and egress etc. The pavement and shoulder conditions are good and measurements were taken in the dry and clear weather conditions. All roadway sections are composed of bituminous paved surfaces with well-maintained conditions of pavement as well as shoulders. The google map view of one of the selected sites along with its camera view picture is shown in Figures 1(a) and (b).

4. Data collection and preliminary analysis

The data collection part is split into two types: geometric data and free-flow speed data. The geometric data was measured in field using total station instrument during daytime in good weather conditions. For this, the local coordinates were found for every 10 m along the tangent to a curve using the total station at each horizontal curve. The coordinates of each location were converted into DXF format for processing data in AutoCAD (version 2021), and the roadway profiles were fitted using the best-fit curve tool. A "best fit curve" tool is typically a mathematical technique or software function that finds the most precise curve or line to fit a collection of surveyed data points. After that, the essential data related to roadway geometry such as curve radius (R), curve length (L_c), deflection angle (I_D), degree of curvature (D_c) and preceding tangent length (P_{tl}) obtained. The descriptive statistics of all geometric variables are presented in Table 1.

Tab. 1 - Descriptive statistics of geometric data

Geometric variable	Notation	Minimum	Maximum	Mean
Curve radius (m)	R	100	635	347
Curve length (m)	L_C	45	435	269
Deflection angle (°)	I_D	15	76	46
Degree of curvature (°)	D_C	3	18.2	11
Preceding tangent length (m)	P_{tl}	300	1700	1106
Carriageway width (m)	C_W	7	7	7
Shoulder width (m)	S_W	2.0	2.5	2.3

The free-flow speeds of Car, Two-wheeler (2W), Three-wheeler (3W), Light commercial vehicle (LCV) and Heavy commercial vehicle (HCV) were collected between 09:00 A.M. and 2:00 P.M. on weekdays during good weather conditions. Free-flow conditions on a highway denote a situation where traffic moves fluidly and encounters minimal obstructions. The speeds were measured using radar gun at mid-point of the horizontal curves keeping out of sight so as to be invisible to drivers. The sample size of all vehicles at each location is ranging from 650 to 800, and the average value is 695. The cumulative frequency curves were drawn for determining 85th percentile speed for all selected vehicle types. The descriptive data of 85th percentile speed as obtained from all sites are shown in Figure 3. It is observed that the vehicle type Car was moving with high speed while 3W and HCV were following the same trend. More standard deviation was observed in the case of LCV and HCV, and less for 2W.

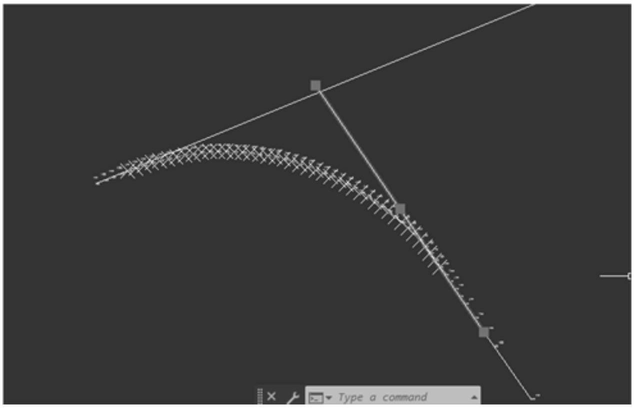


Fig. 2 - Schematic diagram of geometric data extraction

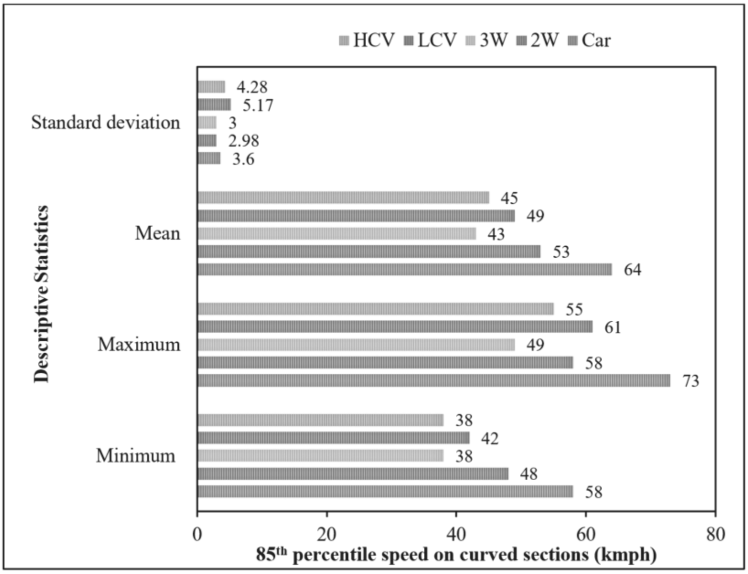


Fig. 3 - Descriptive statistics of 85th percentile speeds on curved sections

5. Methodology

In order to assess the most suitable model for predicting the operating speed on horizontal curve, multiple regression method and some of the machine learning techniques such as artificial neural networks (ANN), adaptive neuro-fuzzy inference systems (ANFIS), support vector machine (SVM) are applied. The detailed analysis was performed using each method and developed models by considering significant variables and parameters. The results obtained from each model have been compared using statistical measures of effectiveness. A best model was selected out of different methods which proved to be more accurate and precise for prediction of operating speed of vehicles under mixed traffic conditions. The various techniques applied for formulation of model are described as follows.

5.1. Multiple Linear Regression

Multiple linear regression is a widely used approach to describe a linear relationship between dependent and one or more independent variables. It facilitates to understand the level of variation among the variables. The optimum multiple linear regression model formulation is based on the higher correlation coefficient, low standard deviation and the value of F-ratio. The typical form of MLR equation is presented in Eq. 1.

$$Y_i = a_0 + a_1 b_1 + a_2 b_2 + - - - - - a_i b_i + \varepsilon_i \quad (1)$$

where Y_i is the dependent variable of i^{th} observed values, a_0 is the intercept, a_1, a_2, and a_i are the regression coefficients, b_1, b_2, and b_i are independent variables, ε_i is the regression model random error. In equation (1), Y represents the operating speed (V_{85}) as the dependent or response variables and (b_1, b_2, …… b_i) are the geometric variables such as curve radius, curve length, deflection angle, degree of curvature, preceding tangent length considered as independent or explanatory variables.

5.2. Artificial Neural Network

Artificial neural network is one of the most popular machine learning techniques for non-linear analysis, which has been inspired by the nervous system of the human brain. It can be used for applications including categorization, grouping, model development, forecasting, and prediction. The basic form of ANN is the perceptron, which has only one input layer and one output layer. Later, this perceptron extended to a Multi-layer perceptron (MLP) by adding one or more layers between the input and output layers, known as hidden layers. The ANN structure consists of neurons, weights, and activation function at various layers. The architecture of basic ANN processing unit with all elements is shown in Figure 4.

In the present study, a feed-forward with backpropagation neural network (BPNN) modeling technique was adopted for predicting operating speed on horizontal curves. The BPNN is a multi-layer perceptron, which consists of an input layer, a hidden layer and output layer are interconnected with neurons. Trial and error method was employed for selecting the required number of hidden layers and neurons in ANN architecture. The optimal network structure has a high correlation coefficient and low mean square error value between observed and predicted values. In Figure 4, each input (X_i) was connected to the neuron by a weighted link (W_i) and the total neuron input was determined using the Equation (2):

$$Net\ input\ (x) = \sum_{i=1}^{n} X_i W_i + K_i \quad (2)$$

where, X_i is the input value, W_i is the established weights and K_i is the related bias for the i^{th} layer of neuron.

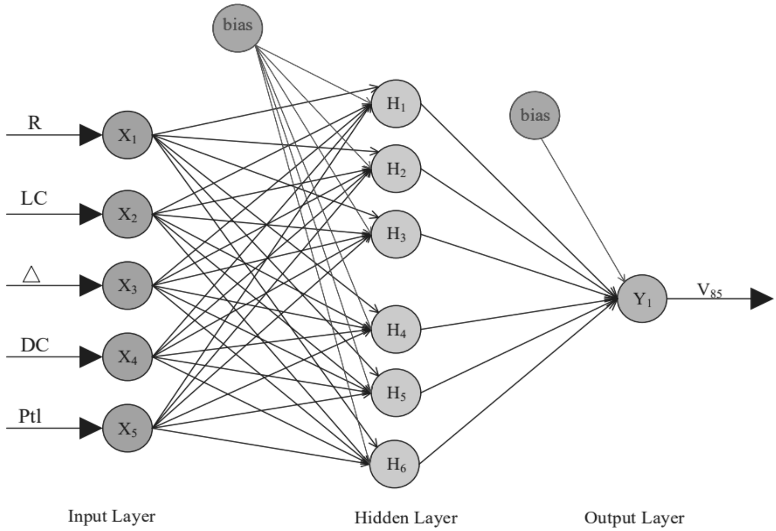


Fig. 4 - Basic architecture of ANN

Thereafter, the neuron output i.e., an activation function was calculated using Sigmoid function using Equation (3):

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

The most popular method for reducing error in feedforward networks is the backpropagation neural network (BPNN) algorithm. In BPNN, the gradient descent with momentum (GDM) algorithm is used to modify the weights for minimizing the sum of squared error loss. The most used equation of squared error function is given by Eq. (4)

$$Sum\ of\ squared\ error = \frac{(Target\ output - Predicted\ output)^2}{2} \quad (4)$$

5.3. Adaptive Neuro-fuzzy Inference System (ANFIS)

ANFIS demonstrates notable effectiveness in acquiring and modelling non-linear relationships inherent in data. Its adaptability enables it to comprehend intricate patterns and relationships that might be challenging to represent through conventional linear methodologies. ANFIS models typically offer greater interpretability when contrasted with intricate neural networks. The decision-making process is clearly represented by the fuzzy rules and membership functions employed in ANFIS, which facilitates user comprehension and interpretation of the model. Additionally, it is capable to represent intricate relationships and adjust to evolving conditions renders it well-suited for dynamic systems.

Adaptive neuro-fuzzy inference system (ANFIS) is a hybrid model that integrates the fuzzy logic and artificial neural network principles. The fuzzy inference systems are grouped into three categories based on the inference operations i.e., “IF-THEN” rule: (1) Mamdani’s system, (2) Tsukamoto’s system and (3) Takagi-Sugeno’s system. Among them, Takagi-Sugeno’s system is widely used because it is more effective and flexible. ANFIS is a Takagi-Sugeno’s type of fuzzy inference system that can evaluate nonlinear functions.

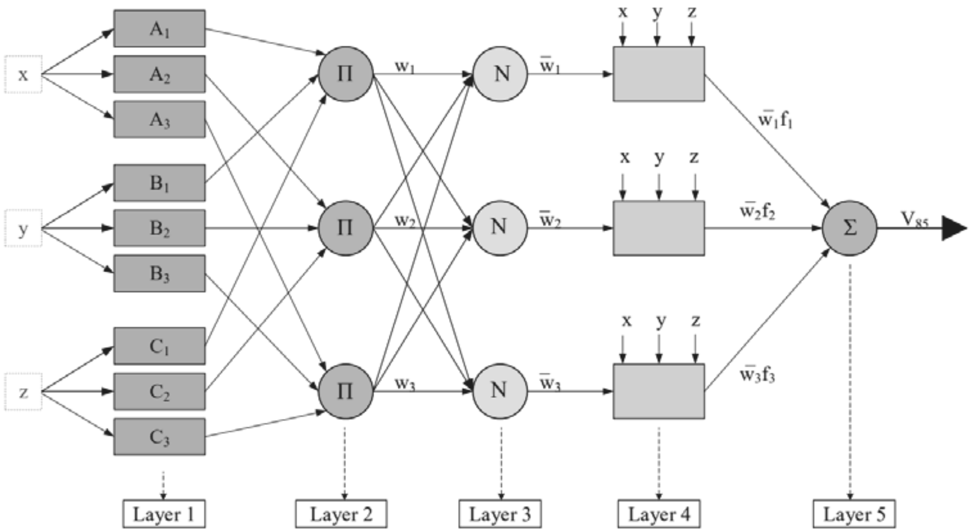


Fig. 5 - Typical architecture of ANFIS

Let us assume that the fuzzy inference system comprises three inputs x, y and z and one output f, the first-order Sugeno fuzzy model rules can be defined as follows:

Rule 1: IF x is A_1*, y is* B_1 *and z is* C_1 *THEN* $f_1=K_1x + L_1y + M_1z + N_1$ (5)

Rule 2: IF x is A_2*, y is* B_2 *and z is* C_2 *THEN* $f_2=K_2x + L_2y + M_2z + N_2$ (6)

Rule 3: IF x is A_3*, y is* B_3 *and z is* C_3 *THEN* $f_3=K_3x + L_3y + M_3z + N_3$ (7)

where x, y and z represent ANFIS inputs, A_1, A_2, A_3, B_1, B_2, B_3, C_1, C_2, C_3 are input membership functions, K_1, K_2, K_3, L_1, L_2, L_3, M_1, M_2, M_3, N_1, N_2, N_3 are the design parameters.

The typical architecture of adaptive neuro-fuzzy inference system consists of five layers and the structure is presented in Figure 5. Each circle represents the fixed node, and each box represents the adaptive node as shown in Figure 5. The explanation of each layer presented is as follows:

Layer 1: Fuzzification

The fuzzification layer consists of a number of membership functions (MF), which are converted into fuzzy inputs. The output of the i^{th} node in this layer is designated as O_{1i}. The output can be calculated as follows

$$O_{1i} = \mu_{Ai}(x) \text{ for } i = 1,2,3 \quad (8)$$

$$O_{1i} = \mu_{Bi}(y) \text{ for } i = 4,5,6 \quad (9)$$

$$O_{1i} = \mu_{Ci}(z) \text{ for } i = 7,8,9 \quad (10)$$

Layer 2: Rule layer

In this layer, calculate the weight of each membership function by multiplying ring layers with the output of first layer:

$$O_{2i} = w_i = \mu_{Ai}(x) * \mu_{Bi}(y) * \mu_{Ci}(z) \quad \text{for } i = 1,2,3 \quad (11)$$

Layer 3: Normalization layer

In this layer, the efficiency of the rule layer output w_i is calculated by normalization. The output of third layer is determined using the following equation:

$$O_{3i} = \overline{w_i} = \frac{w_i}{w_1 + w_2 + w_3} \qquad for\ i = 1, 2, 3 \quad (12)$$

Layer 4: Defuzzification

In this layer, the weighted output from the previous layer is multiplied with fuzzy model rules using Eq. (5, 6, 7). The outputs of layer are given by:

$$O_{4i} = \overline{w_i} = \overline{w_i} f_i = \overline{w_i}\,(K_1 x + L_1 y + M_1 z + N_1) \quad for\ i = 1,2,3 \tag{13}$$

Layer 5: Output layer

In this layer, calculate the sum of the weighted values obtained in the fourth layer to produce the system output. The overall output is determined by using the Eq. (14)

$$O_{i5} = \sum_{i=1}^{n} \overline{w_i} f_i = \frac{\sum_i \overline{w_i} f_i}{\sum_i w_i} \tag{14}$$

5.4. Support Vector Regression (SVR)

Support vector machine was used to carry out classification tasks, but it may also be used to solve continuous outcomes or regression problems with the inclusion of a few additional principles, called support vector regression (SVR). It demonstrates scenarios involving high-dimensional spaces, rendering it suitable for datasets characterized by a significant number of features. It has the capability to manage concerned relationships and patterns within these environments. SVR exhibits lower sensitivity to outliers in the training data when contrasted with certain other regression techniques. The motive of SVR is to identify a hyperplane that optimally fits the data while minimizing errors. The additional parameters in SVR are epsilon (Ɛ) and slack (ξ). Epsilon is the width on both sides of hyperplane and the points within this width are considered as no error or exact prediction. Slack is the distance between epsilon and extreme points away from the hyperplane as shown Figure 6. Slack indicates the degree of tolerance for points away from the epsilon.

For a given data set {(x_1, y_1), (x_2, y_2), (x_3, y_3),….(x_m, y_m)}, which are each xi ∈ R_m indicates the input space of the sample and m is the number of samples. Consider a linear regression function which can be written as:

$$y = f(x) = a x_i + b \tag{15}$$

where, a is the coefficient vector corresponding to x and b is the bias.

In SVR, Ɛ-insensitive loss is to be minimized and Vapnik's linear function can be written as follows:

$$L_\varepsilon = \begin{cases} 0, & if\ |y_i - (a x_i + b)| \le \varepsilon \\ |y_i - (a x_i + b)| - \varepsilon & otherwise \end{cases} \tag{16}$$

In the above equation, Ɛ is a tuning parameter and it can be re-written for optimization as follows:

$$\text{minimize} \quad \frac{1}{2}\|a\|^2 \tag{17}$$

$$subject\ to \begin{cases} y_i - a x_i - b \le \varepsilon \\ -y_i + a x_i + b \le \varepsilon \end{cases} \tag{18}$$

If data points do not lie within Ɛ on both sides of regression line or hyperplane, there is no solution to the problem. The slack variables i.e., ξi, ξ_i^* introduced to allow error levels greater than Ɛ and the optimization function for minimizing the error are given below:

$$\text{minimize} \quad \frac{1}{2}\|a\|^2 + C\sum_{i=1}^{m}(\xi i + \xi_i^*) \tag{19}$$

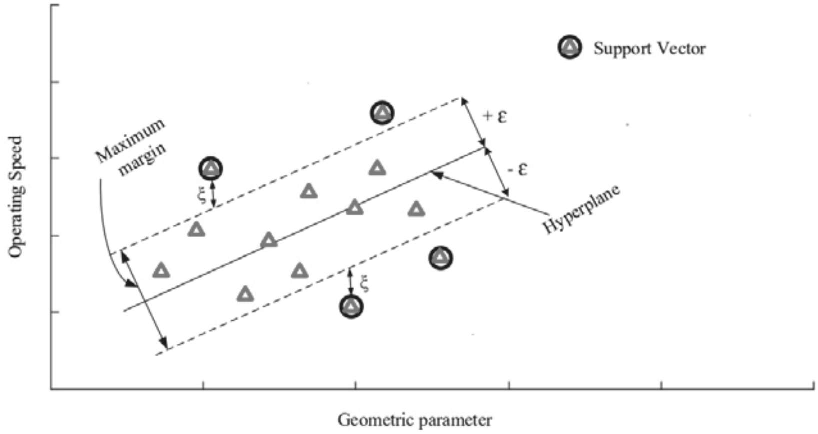


Fig. 6 - Schematic diagram of Support Vector machine

$$subject\ to \begin{cases} y_i - ax_i - b & \leq \varepsilon + \xi i \\ -y_i + ax_i + b & \leq \varepsilon + \xi_i^* \\ \xi i, \xi_i^* & \geq 0 \end{cases} \quad (20)$$

where C is the correction parameter and $\xi i, \xi_i^*$ are slack variables. In order to get the ideal constrained function's optimum and produce the appropriate weight vector for the regression function, the aforementioned constrained optimization problem is solved utilising the Lagrangia theory and Karush-Kuhn-Tucker conditions. The proposed formulation is given as follows:

$$a = \sum_{i=1}^{m} (\beta_i - \beta_i^*)\ \varphi(x_i) \quad (21)$$

$$y = f(x) = \sum_{i=1}^{m} (\beta_i - \beta_i^*) K(x_{i,} x) + b \quad (22)$$

where β_i, β_i^* are dual variables, $K(x_{i,}x)$ is called kernel function and it indicates the multiplication between $\varphi(x_i)$ and $\varphi(x)$. The above function gives the solution for original regression problem.

5.5. Statistical performance measures

The prediction performance of each model is evaluated using various goodness-of-fit measures, namely, root mean square error (RMSE) and mean absolute error (MAE).

RMSE is a measurement for the difference between observed and predicted data in addition to an indicator of how accurately the model fits the data in its entirety. The lower value of RMSE shows good fitting of data. The value of RMSE is calculated using the following equation:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (Observed_i - Predicted_i)^2}{n}} \quad (23)$$

MAE was used to compare the different series and scale of data. It was determined using the following equation:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |Predicted_i - Observed_i| \quad (24)$$

6. Analysis of field data, results and discussion

6.1. Correlation analysis

Correlation analysis was performed among 85th percentile speed of various vehicle type and geometric features of each horizontal curve. The Pearson correlation analysis was conducted in R studio and values are obtained as presented in Figure 7. It can be observed that the four geometric parameters are significantly correlated with 85th percentile speed. These parameters are identified as radius, curve length, degree of curvature and preceding tangent length. Among them, three variables namely, radius, curve length and preceding tangent length are positively correlated, and the only degree of curvature is negatively correlated. Four significant variables were adopted for model development in MLR, ANN, ANFIS and SVR.

6.2. MLR analysis

Regression analysis was carried out to develop linear relationship between operating speed (V_{85}) and geometric characteristics of roadway for five vehicle types at a significance level of 0.05. Four explanatory variables, namely, curve radius, curve length, degree of curvature and preceding tangent length were used to predict the operating speed on horizontal curves. It is observed that the corresponding *p*-value of all variables were approximately equal to zero as the *t*-stat was very less than critical value. Moreover, the Variable Inflation Factor (VIF) was less than 4 and the standard error was also less for all variables. The coefficient of determination (R^2) observed maximum for Car as 0.782 and minimum for 3W as 0.575. The operating speed decreased with increase in the coefficient value of variable D_C and it increased with increase in coefficient value of variables R, L_C and P_{tl}. It is noted that the R and D_C have more influence on the operating speed of vehicles than other parameters at horizontal curve sections. The high coefficient value of D_C was observed for HCV vehicle type and indicates clearly that the HCV is very sensitive to D_C of horizontal curve. The results for prediction of V_{85} using MLR are presented in Table 2. The models developed are validated using observed and predicted value in terms of RMSE and MAE are discussed in the comparison section.

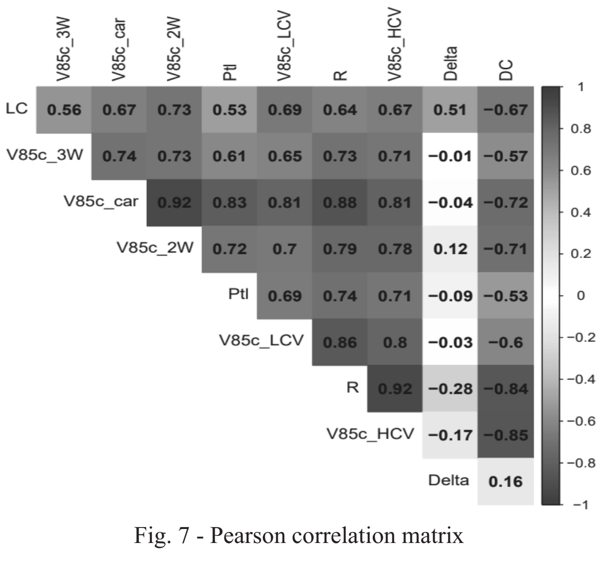


Fig. 7 - Pearson correlation matrix

Tab. 2 - Results of multiple linear regression analysis

Model No.	Model	EV	VIF	SE	*t*-stat	*p*-value
1	V_{85c}_Car $= 56.80 + 0.012R + 0.003L_c - 0.27D_c + 0.006P_{tl}$	R	1.65	0.00	4.31	0.000
		L_C	1.33	0.00	1.82	0.001
		D_c	1.21	0.02	3.55	0.000
		P_{tl}	1.56	0.00	3.11	0.000
2	V_{85c}_2W $= 43.26 + 0.00392R + 0.008L_c - 0.19D_c + 0.003P_{tl}$	R	2.32	0.00	2.95	0.000
		L_C	1.31	0.00	3.48	0.000
		D_c	1.22	0.01	-1.38	0.001
		P_{tl}	1.65	0.00	3.26	0.001
3	V_{85c}_3W $= 35.14 + 0.013R + 0.004L_c - 0.188 + 0.001P_{tl}$	R	1.26	0.02	2.89	0.000
		L_C	1.49	0.00	3.05	0.001
		D_c	1.43	0.00	-2.28	0.000
		P_{tl}	1.95	0.06	3.32	0.001
4	V_{85c}_LCV $= 44.26 + 0.022R + 0.0084_c - 0.24D_c + 0.004P_{tl}$	R	1.55	0.00	4.06	0.000
		L_C	1.36	0.00	3.94	0.001
		D_c	1.88	0.05	-2.44	0.000
		P_{tl}	1.93	0.00	3.15	0.002
5	V_{85c}_HCV $= 36.74 + 0.018R + 0.002L_c - 0.44D_c + 0.002P_{tl}$	R	1.41	0.00	3.56	0.000
		L_C	1.21	0.04	2.56	0.000
		D_c	2.21	0.02	-2.78	0.001
		P_{tl}	1.61	0.00	2.48	0.004

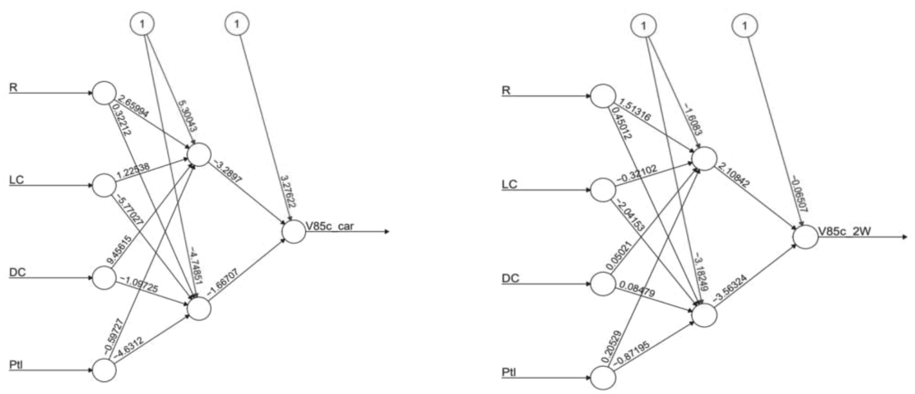
6.3. Artificial Neural Network

In this study, ANN based operating speed model was built using R Studio "neuralnet" package. As a result of correlation analysis, four explanatory variables (R, L_C, D_C and P_{tl}) highly correlated with V_{85} of all vehicle types are used in the input layer. For network development, the curve data was divided into training data set (70%) and testing data set (30%). To develop an optimal model, the number of neurons in the hidden layer was changing in accordance with the performance of maximum correlation coefficient and minimum error. Several trials were done to reach the optimum model performance and the result of some of them are presented in Table 3.

It is noticed that the error value decreases by increasing number of neurons in the hidden layer in all model structures, but the correlation coefficient randomly changes. The best ANN architectures for all vehicle type models are presented in Figure 8. In addition, it can be observed that the highest performance model architecture for Car consists of 4 input neurons, 2 hidden neurons and one output neuron. The structure having highest correlation coefficient and error value as 0.9537 and 0.2678, respectively. The best ANN structure for 2W was 4-2-1 selected with high correlation coefficient and least error value as 0.8121 and 0.8683, respectively. High error value and lowest correlation coefficient were observed in 3W models among all best model architectures i.e., 1.4160 and 0.7758, respectively. In case of LCV, only one neuron was found in the hidden layer for optimum model, in which the correlation coefficient is 0.8154 and error value is 0.8745. The best ANN architecture is found for HCV is 4-2-1 with correlation coefficient and error value as 0.8562 and 0.6206, respectively.

Tab. 3 - ANN architecture with correlation coefficient and error value

Vehicle type	Architecture	Maximum correlation coefficient	Error (%)
Car	4-1-1	0.9359	1.5353
	4-2-1	**0.9537**	**0.2678**
	4-3-1	0.7945	0.1888
	4-4-1	0.8934	0.0074
	4-5-1	0.9004	0.0068
2W	4-1-1	0.8045	2.3571
	4-2-1	**0.8121**	**0.8683**
	4-3-1	0.5479	0.6639
	4-4-1	0.4450	0.0338
	4-5-1	0.6529	0.1435
3W	4-1-1	0.4574	4.1601
	4-2-1	**0.7758**	**1.4160**
	4-3-1	0.0026	1.3260
	4-4-1	0.6726	1.9137
	4-5-1	0.4503	0.0369
LCV	**4-1-1**	**0.8154**	**0.8745**
	4-2-1	0.2338	0.6941
	4-3-1	0.7004	0.2652
	4-4-1	0.4657	0.1973
	4-5-1	0.0548	0.1133
HCV	4-1-1	0.6062	1.2919
	4-2-1	**0.8562**	**0.6206**
	4-3-1	0.7340	0.3117
	4-4-1	0.5918	0.1528
	4-5-1	0.4946	0.1186



(a) (b)

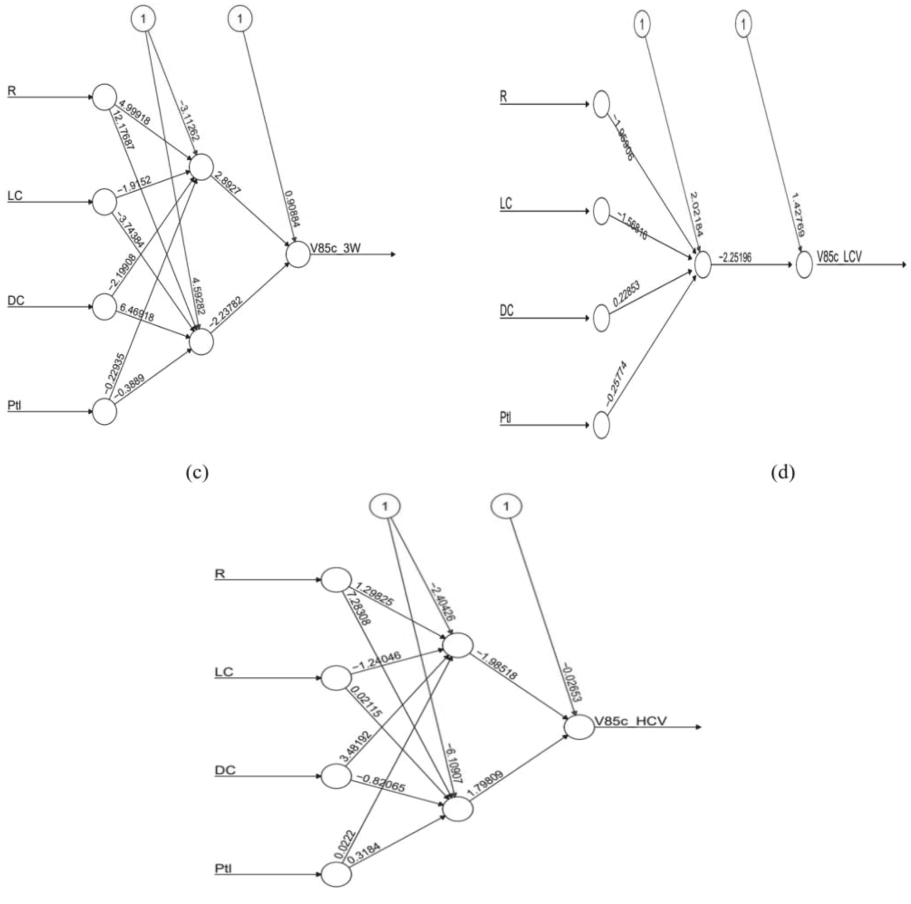


Fig. 8 - Best performance ANN model architecture (a) Car (b) 2W (c) 3W (d) LCV (e) HCV

6.4. Adaptive Neuro-fuzzy Inference System

ANFIS employs a hybrid learning algorithm for developing relationship between operating speed (V_{85}) and geometric characteristics on horizontal curves in the present study. The optimal learning variables were trained using triangular membership function (trimf) as it gives low error value and better performance than other membership functions. During data training, the learning process will be resumed when data error falls below the threshold or the maximum number of epochs had been reached. The maximum number of epochs used to train the data in the present study are 1000. For model development, 70% of data was used for calibration and the remaining 30% data was used for testing. The best performance model was chosen based on low error value for both calibration and testing datasets.

The correlation analysis was performed considering the curve radius, curve length, degree of curvature and preceding tangent length as input variables and V_{85} as output variable to develop fuzzy model for all vehicle types. Figure 9 shows the domain of each input variable covered with membership functions.

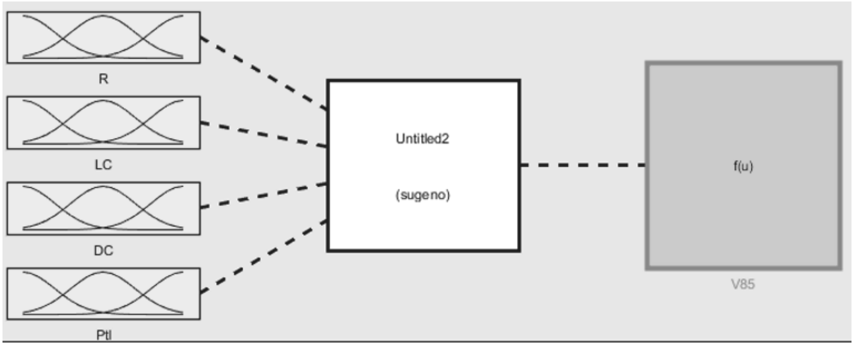


Fig. 9 - Fuzzy model construction

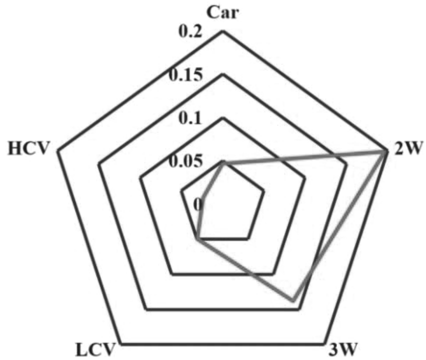


Fig. 10 - Error value of each vehicle at 1000 epochs

Thereafter, the data was trained at different epochs, and it can be observed that the minimum error value was obtained at 1000 epochs for all vehicle types. Figure 10 shows the error values at 1000 epochs for different vehicles. The HCV vehicle type shows minimum error value as 0.0236 among all other vehicle types. Car and LCV have approximately equal error values of 0.0468 and 0.0492, respectively. The highest error value observed in 2W was 0.1967 and a moderate error value of 0.1398 was observed in case of 3W. The coefficient of determination of better performance models are obtained as 0.842, 0.716, 0.602, 0.798, and 0.635 for Car, 2W, 3W, LCV and HCV, respectively.

The surface diagram functions of "radius and preceding tangent length" were estimated using ANFIS models for Car, 2W, 3W, LCV and HCV, as shown in Figure 11(a)-(e). These figures describe the relationships between variables with each other as well as with operating speed. From figure 11(a), it is observed that the long preceding tangent length followed by a small radius of curve leads to decreasing speed on the horizontal curve and vice-versa. Similarly, the same phenomenon was found in other vehicle types, but more reduction was found in speed of HCV. The other relationships like radius and degree of curvature, preceding tangent length and length of curve etc., were also developed for all types of vehicles. It can be found the degree of curvature increases while other parameters remain constant and thus results in decreasing operating speed. Moreover, it is observed that the HCVs are moving with low speed whereas Cars are moving with high speed while traversing horizontal curve.

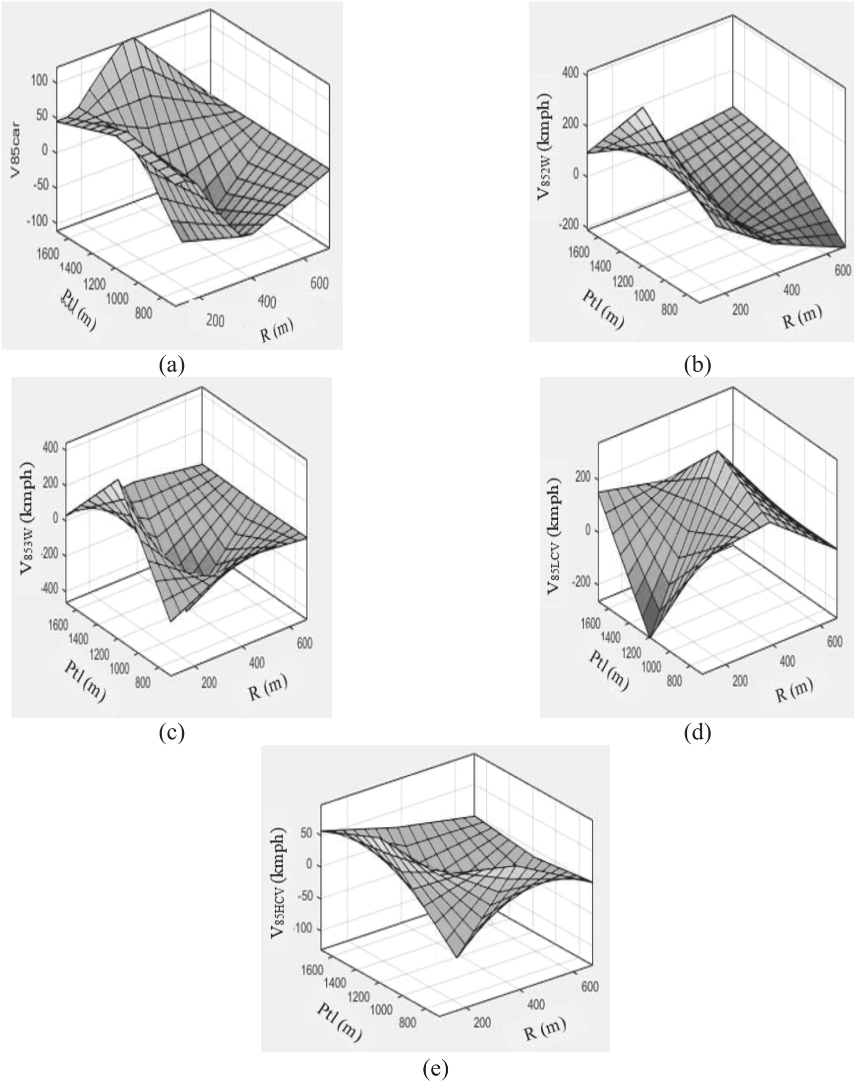


Fig. 11 - Surface diagrams (a) Car (b) 2W (c) 3W (d) LCV (e) HCV

6.5. Support Vector Regression

The observations in earlier correlation analysis, such as curve radius, curve length, degree of curvature, preceding tangent length and 85th percentile speeds (V_{85}) of different vehicle types were used in training and test data set. The training and test data set contains 70% and 30% of observations, respectively. In this study, several trials were done to find the better performance model with different folds. It was found that using 5-fold cross validation gave best performance for SVR model, as shown Figure 12.

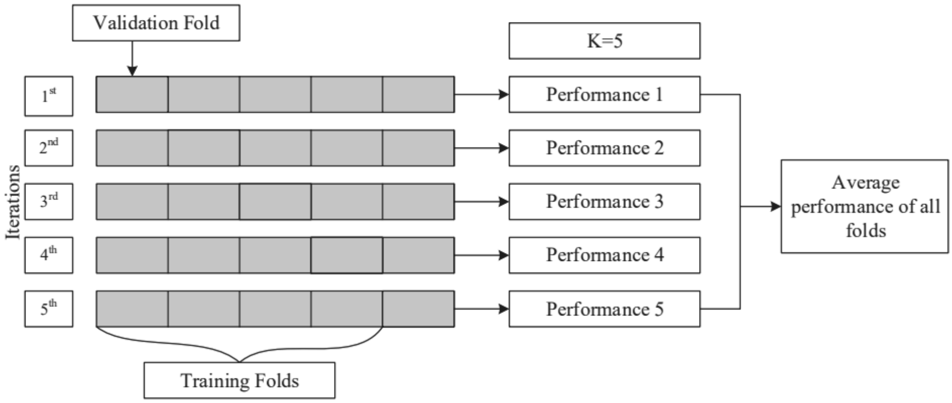


Fig. 12 - Cross-validation

Initially, the training data was split into five folds and the model was developed using four training folds in 1st iteration. This process was repeated for all 5 folds and finally, the accuracy measure was calculated by taking average of all 5 performances. The best performance model was selected based on the low error value for each individual vehicle type.

Optimum models obtained from cross-validation were evaluated using test dataset in terms of RMSE and MAE. Support vector regression models were evaluated using various kernel functions, namely, polynomial, Gaussian, Radial basis function and sigmoid functions. Among them, Gaussian kernel function gives better performance in terms of R^2 of 0.864, 0.735, 0.658, 0.803, and 0.662 for Car, 2W, 3W, LCV and HCV, respectively.

7. Model comparison with previous studies

In the present study, authors adopted MLR, ANN, ANFIS, and SVR modeling techniques to estimate operating speed. In previous research, the predominant approach for model development was MLR, with only one study focusing on the ANN technique. Notably, there have been no reported studies on predicting operating speeds on curves using ANFIS and SVR modeling techniques. Table 4 provides a comparison between the suggested MLR and ANN models and the pre-existing models. It is evident from the results that the proposed MLR model exhibits the lowest RMSE values for car, 2W, and HCV compared to other existing models. Similarly, the proposed ANN model demonstrates a high degree of accuracy than the preceding model.

8. Comparison of MLR, ANN, ANFIS and SVR

The goodness-of-fit measures like RMSE and MAE are used to compare the accuracy of MLR, ANN, ANFIS and SVR in predicting the operating speed of different vehicle types on horizontal curves. It may be observed that estimated values of RMSE and MAE are lower with SVR method, shows it performs better with high precision and accuracy for predicting operating speed for all vehicle types in compared to conventional and other machine learning methods. MLR showed highest error values compared to ANN, ANFIS and SVR in all the cases. It clearly indicates that the soft computing modeling techniques predicted the operating speeds with high accuracy. Hence, these techniques are found to be better alternatives for processing continuous variable dataset such as speed. Among them, this study recommends SVR modeling technique as best for predicting the operating speed of Car, 2W, 3W, LCV and HCV on complex location such as horizontal curves. The performance measures applied with respect to different models and vehicle types are presented in Figures 13 (a) and (b).

Tab. 4 - Comparison of proposed models with existing models

S.No.	Author(s)	Vehicle type	Model	RMSE	MAE
Multiple Linear Regression					
	Proposed Study	Car	$V_{85c} = 56.80 + 0.012R + 0.003L_c - 0.27D_c + 0.006P_{tl}$	3.662	26.69
1.	Malaghan et al. [15]	Car	$V_{85c} = 72.10 + 0.002R - 0.01L_c - 1.14D_c$	6.52	28.26
2	Jacob and Anjaneyulu [1]	Car	$V_{85c} = 69.00 - (0.012/R) + 0.065L_c$	9.11	19.12
3	Misaghi and Hassan [2]	Car	$V_{85c} = 94.30 + 8.67x10^{-6}R^2$	11.86	8.23
4	Perco [37]	Car	$V_{85c} = 121.78 - (544.78/\sqrt{R})$	15.63	11.26
5	Zuriaga et al. [38]	Car	$V_{85c} = 97.43 - (3310.94/R)$	13.52	10.22
6	Ottesen and Krammes [39]	Car	$V_{85c} = 103.66 - 1.95D_c$	14.66	9.86
7	Passetti and Fambro [40]	Car	$V_{85c} = 103.90 - (3020.50/R)$	22.68	17.23
8	Castro et al. [24]	Car	$V_{85c} = 91.13 + 0.03L_c - 0.48\Delta$	21.96	15.56
9	Voigt [20]	Car	$V_{85c} = 99.61 - (2951.37/R\ )$	18.65	14.23
10	Taragin and Leisch [41]	Car	$V_{85c} = 88.87 - (2554.76/R)$	29.69	26.89
11	Glennon et al. [42]	Car	$V_{85c} = 103.9 - (4524.9/R)$	28.11	21.35
12	Lamm and Choueiri [18]	Car	$V_{85c} = 94.39 - (3189.94/R)$	31.23	22.12
13	Krammes et al. [7]	Car	$V_{85c} = 103.64 - (3400.73/R)$	22.56	19.23
14	Kanellaidis et al. [43]	Car	$V_{85c} = 129.88 - (623.13/\sqrt{R}$	24.36	15.56
	Proposed Study	2W	$V_{85c\_}2W = 43.26 + 0.00392R + 0.008L_c - 0.19D_c + 0.003P_{tl}$	2.982	19.41
1	Jacob and Anjaneyulu [1]	2W	$V_{85c} = 67.00 - (1105.72/R) + 0.069L_c$	6.33	24.63
	Proposed Study	HCV	$V_{85c\_}HCV = 36.74 + 0.018R + 0.002L_c - 0.44D_c + 0.002P_{tl}$	5.86	36.42
1	Jacob and Anjaneyulu [1]	HCV	$V_{85c} = 63.20 - (1063.82/R) + 0.061L_c$	9.26	28.63
2	Castello et al. [12]	HCV	$V_{85c} = 73.76 - (1063.82/e^{0.0072-R})$	6.33	26.33
Artificial Neural Network					
S.No.	**Author(s)**	**Vehicle type**		**RMSE**	**MAE**
	Proposed Study	Car		2.631	18.96
1	Singh et al. [33]	Car		5.1	22.15
**V_{85C}=Operating speed on mid-point of the curve, R=curve radius, D_C=Degree of curvature, Δ =Deflection angle, L_C=Curve length, P_{tl}=Preceding tangent length.