

RESEARCH

Open Access



# Developing operating speed models for elevated multilane urban arterials using artificial neural networks

Dalia Said<sup>1\*</sup> , Fady Reyad<sup>2</sup> and Hoda Talaat<sup>1</sup>

\*Correspondence:  
daliasaid@eng.cu.edu.eg

<sup>1</sup> Public Works Department,  
Traffic and Highway Engineering,  
Cairo University, Giza, Egypt

<sup>2</sup> Civil Engineering Department,  
The British University in Egypt,  
Cairo, Egypt

## Abstract

Operating speed models help assess and evaluate geometric design consistency along successive road segments. The development of operating speed models has been mostly focused on rural two-lane two-way highways, where horizontal curvature plays a dominant role in speed prediction. The need to enhance the prediction power of operating speed models and ability to capture more complex relationships within an urban setting have motivated this investigation. This research investigates the use of artificial neural networks, to develop operating speed models for multilane urban elevated arterial roads. Variables investigated in this study included geometric/operational features of a road segment in addition to the residual impact of the characteristics of upstream segments. A data collection exercise was undertaken on two major urban elevated arterial roads in Greater Cairo Region, Egypt: 6th of October and Saft Al-Laban corridors. Speed data was extracted from Google Distance Matrix Application Programming Interface and validated using test vehicle speed data. A regression-based modeling exercise was undertaken in the preliminary investigation phase to serve as a benchmark for the intended machine learning modeling exercise. Results showed that the prediction power of the developed ANN models — capturing the residual effect of upstream speeds — outperformed regression-based ones. The best-performing model used operating speeds of two upstream segments in addition to geometric/operational features of the segment under investigation to predict the segment operating speed (the model reported MAPE of 6.7%). Outputs of this model were used in a design consistency evaluation and potential transferability exercises to further investigate the model practicality.

**Keywords:** Elevated urban arterials, Operating speed models, Artificial neural networks, Geometric design consistency

## Introduction

Road safety and design consistency have been of eminent concern in the last few decades. The hefty price that comes with post construction evaluations and design improvements has motivated efforts for pre-construction design consistency evaluation (during the design phase). Speed profile along a given road is a key measure in road safety and design consistency assessments. Variations in operating speeds are directly influenced

by variations in road geometry. Limiting expected variations in operating speeds, by design, has become a worldwide necessity for sustainable road development. Thus, understanding the impact of road geometry on operating speeds and enhancing the prediction power of operating speed models have been of crucial concern to researchers as well as practitioners.

Immense efforts have been dedicated to modeling operating speeds, most of which were focused on freeways and two-lane two-way highways [1]. Studying the relationships between geometric/operational features of a given segment (such as horizontal curve radii, cross-section elements, and posted speeds) and its operating speed, using ordinary least-square regression techniques, has been the core of many developed operating speed models. Such modeling efforts have significantly contributed to the understanding of this complex relationship.

The need for enhancing the prediction power of operating speed models and modeling more complex relationships within an urban setting has motivated this investigation. Three main ideas were of interest to investigate: (1) widening the scope of modeling inputs to include road features of upstream segments; (2) generalizing the modeling attempt through adopting a case study of an elevated urban arterial and not the conventional two-way two-lane case study, and (3) using a machine learning modeling approach in addition to the conventional regression-based approach to enable the comparative assessment of both approaches. In the following section, examples of previous modeling efforts are described to set the stage for the current investigation.

## **Background**

Developing operating speed models has been the focus of several research endeavors. Ordinary least-square regression has been the conventional modeling approach for a long time, with elements of horizontal curves serving as the key input to most of the developed models. On the other side, the use of machine learning in operating speed modeling has been gaining recent interest. Models were developed using artificial neural networks (ANN) in an attempt to compare their performance to that of conventional regression-based models. This section presents selected models from both sides.

### **Regression-based operating speed models**

Table 1 presents a summary of previous regression-based operating speed models developed to predict the driver speed on two-lane two-way highways or arterials. For two-lane two-way highways, the models developed were mostly for curves with a few developed for tangents. Predictors usually represent the geometric characteristics of the alignment, such as horizontal curve radius (or its variations), deflection angle, length of curve, vertical grade of curve, approach tangent length, and operating speed of the previous element (approaching speed). Predictors of arterial roads on the other hand are mainly based on speeds such as posted speed limit and inferred design speed or based on urban segment characteristics such as access density, land uses, number of lanes, and the presence of sidewalks. This highlights that predicting operating speeds on arterial segments is more complicated than two-lane two-way highways due to various factors affecting speed other than the geometry of the alignment.

**Table 1** Previously developed regression-based operating speed models

Author (s)	Model	R2	Eq.
Two-lane two-way highways			
Lamm et al. [2]	$V_{85C} = 34.70 - 1.00DC + 2.081LW + 0.174SW + 0.0004AADT$	0.842	1
Krammes et al. [3]	$V_{85C} = 102.45 - 1.57DC + 0.0037Lc - 0.1\Delta$	0.82	2
	$V_{85C} = 41.62 - 1.29DC + 0.0049Lc - 0.12\Delta + 0.95V_{85T}$	0.9	3
Voigt [4]	$V_{85} = 102 - 2.08DC + 40.33e$	0.81	4
	$V_{85C} = 99.6 - 1.69DC + 0.014L - 0.13\Delta + 71.82e$	0.84	5
Gong and Stamatiadis [5]	For inside lane:	0.647	6
	$V_{85IL} = 51.520 + 1.567ST - 2.795MT - 4.001PT - 2.15AG + 2.22ln(LC)$	0.432	7
	For outside lane $V_{85OL} = 60.779 + 1.804ST - 2.521MT - 1.071AG - 1.519FC$ $+ 0.00047R + 2.408 \frac{LC}{R}$		
Zuriaga et al. [6]	For all radii values:	0.76	8
	$V_{85C} = 97.4254 - \frac{3310.94}{R}$	0.79	9
	$V_{85C} = \frac{1}{0.00948323 + (0.00001501 * CCR)}$	0.84	10
	For radii > 400 m $V_{85C} = 102.048 - \frac{3990.26}{R}$	0.52	11
	For Tangents $V_{85T} = V_{85C} + (1 - e^{-\lambda L}) \times (V_{des} - V_{85C})$		
Passetti and Fambro [7]	$V_{85C} = 103.9 - 3020.5/R$	0.68	12
Abbas et al. [8]	$V_{85MC} = 75.344 - \frac{368.14}{\sqrt{R}} + 0.307V_{85AT}$	0.532	13
Mahmoud et al. [9]	Two-dimensional model:	0.69	14
	$V_{85C} = 102.466 - \frac{5003.426}{R}$	0.732	15
	Three-dimensional model: $V_{85C} = 103.559 - \frac{4951.51}{R} - 3012.795G^2$		
Islam and Seneviratne [10]	$V_{85PC} = 95.41 - 1.48DC - 0.012DC^2$	0.99	16
	$V_{85MC} = 103.3 - 2.41DC - 0.029DC^2$	0.98	17
	$V_{85PT} = 96.11 - 1.07DC$	0.98	18
Bird and Hashim [11]	$V_{85C} = 119.073 - \frac{518.275}{\sqrt{R}} - \frac{125440}{ATL^2} + \frac{413.181}{\Delta^2}$	0.88	19
Hashim et al. [12]	For point of curvature:	0.704	20
	$V_{85} = 99.885 - \frac{3880.21}{R}$	0.730	21
	For center of the curve: $V_{85} = 101.564 - 3480.88/R$	0.728	22
	For point of tangency: $V_{85} = 101.18 - \frac{3969.9}{R}$	0.851	23
	For tangents $V_{85T} = 84.34 + 0.593\sqrt{LT}$		
Arterials			
Fitzpatrick et al. [13]	For all roadways	0.904	24
	$V_{85C} = 7.675 + 0.98PSL$	0.86	25
	For urban/suburban arterials $V_{85C} = 8.666 + 0.963PSL$	0.81	26
	For rural arterials $V_{85} = 36.453 + 0.517PSL$		
Fitzpatrick et al. [14]	$V_{85C} = 43.5 + 0.38IDS$	0.83	27
	$V_{85T} = 74.91 + 22.29/AD$	0.71	28
Wang et al. [15]	$V_{85T} = 31.565 + 6.491lane.num - 0.101roadside - 0.051driveway$ $- 0.082intersection + 3.01curb - 4.265sidewalk - 3.189parking$ $+ 3.312land.use1 + 3.273land.use2$	0.67	29

Where:

$V_{85}$  = predicted operating speed (mph or km/h)

$V_{85C}$  = predicted operating speed on horizontal curve (mph or km/h)

$V_{85T}$  = predicted operating speed on tangent (mph or km/h)

$V_{85PC}$  = predicted operating speed at point of curvature (km/h)

$V_{85MC}$  = predicted operating speed at midpoint of horizontal curve (km/h)

$V_{85PT}$  = predicted operating speed at point of tangency (km/h)

$V_{85IL}$  = predicted operating speed on horizontal curve inside lane (mph)

$V_{85OL}$  = predicted operating speed on horizontal curve outside lane (mph)

**Table 1** (continued)

---

$R$  = horizontal curve radius (m)  
 $DC$  = degree of curvature ( $^{\circ}$ ) (equals to  $1746.38/R$ )  
 $CCR$  = curvature change rate  
 $LC$  = length of horizontal curve (ft)  
 $LW$  = lane width (ft)  
 $SW$  = shoulder width (ft)  
 $AADT$  = average annual daily traffic (vehicles/day)  
 $L_T$  = tangent length (m)  
 $\Delta$  = deflection angle ( $^{\circ}$ )  
 $e$  = superelevation rate (m/m)  
 $G$  = Grade on Horizontal Curve  
 $ST$  = shoulder type indicator (1 if the shoulder type is surfaced; 0 otherwise)  
 $MT$  = median type indicator (1 if no median barrier present; 0 otherwise)  
 $PT$  = pavement type indicator (1 if pavement type is concrete; 0 if asphalt)  
 $AG$  = approaching section grade indicator (1 if the absolute grade  $\geq 0.5\%$ ; 0 otherwise)  
 $FC$  = front curve indicator (1 if the approaching section is a curve; 0 otherwise)  
 $V_{des}$  = desired speed (km/h)  
 $PSL$  = Posted speed limit (km/h)  
 $\lambda$  = calibration factor to minimize mean square error (MSE)  
 $ATL$  = the average length of the preceding and following tangents of the curve  
 $AD$  = approach access density (number of access points per km)  
 $IDS$  = inferred design speed (km/h)  
 $roadside$  = density of roadside objects per mile / average offsets from roadside (ft)  
 $driveway$  = density of driveways (number of driveways per mile)  
 $lane.num$  = number of lanes  
 $curb$  = 0 if there is no curb; otherwise, curb = 1  
 $sidewalk$  = 0 if there is no sidewalk; otherwise, sidewalk = 1  
 $parking$  = 0 if there is no on-street parking; otherwise, parking = 1  
 $land.use1 = 0$  and  $land.use2 = 0$  if land use is commercial  
 $Intersection$  = intersection density per mile

### Artificial neural networks models

McFadden et al. [16] conducted a comparative study by developing two ANN models using the same predictors from the same dataset collected by Krammes et al. [3]. The accuracy of ANN models was examined by comparing the coefficients of determination ( $R^2$ ). The performance of ANN model 1 was compared to the performance of the regression model shown previously in Eq. 4 where ANN model 1 reported  $R^2$  of 0.761 compared to 0.81 (regression based). The performance of ANN model 2 was compared with the performance of the regression model shown previously in Eq. 5 where ANN model 2 reported  $R^2$  of 0.79 compared to 0.84 (regression based).

Semeida [17] investigated the accuracy of ANN compared to regression by developing models using same inputs for two-way multilane highways in Egypt. Posted speed limit and geometric features of road segments were the independent variables for the regression-based model to predict operating speed on any road segment. The model with the best accuracy reported a  $RMSE$  of 10.32. Semeida [17] stated that although the regression model had an acceptable accuracy, the model did not consider all the predictors. Therefore, he used multilayer perceptron feedforward ANN to develop a model which considers all the independent variables of the dataset as predictors. ANN model training set had  $RMSE = 2.9$  and  $R^2 = 0.982$ , while for validation set, the  $RMSE = 4.12$  and  $R^2 = 0.84$ .

In a separate study, Smeida [18] developed two regression models (one model for passenger cars and the other for trucks) and two ANN models to predict the operating speed on horizontal curves for multilane rural highways in Egypt. Only two independent variables were included in the regression-based model (median width and deflection angle). There were four other variables (shoulder width, curve length, curve radius, superelevation rate) that were significantly correlated with the operating speed that were not included in the regression models. Therefore, two ANN models were developed to examine the effect of using all correlated variables as inputs. The performance of the developed ANN models reported a *RMSE* of 5.77 and 4.3 and  $R^2$  of 0.932 and 0.95, for passenger cars and trucks, respectively.

Based on the presented literature, it is apparent that using ANN in modeling operating speeds is still evolving. Most of reviewed models were developed for freeways and two-lane two-way highways. Almost all considered predictors are segment specific (curvature radii, deflection angles, cross section elements, posted speed, etc.), which implies that changes in driver's speed are only triggered by immediate changes in road features. This observation motivated this research to investigate the residual impact of upstream segments' features on operating speeds. In other words, are driver's speed-related decisions function only of the features of the current segment, or is there some built-up knowledge that comes with travelling on upstream road segments?

The objective of this research is to investigate the performance of ANN in modeling operating speeds on multilane elevated urban arterial roads. In doing so, the accuracy of the developed ANN model is compared to another model developed using multiple linear regression (MLR) as a benchmark. The effect of adding the operational and geometric features of the upstream segments as a variable on the accuracy of the developed model was investigated. In addition, to facilitate speed data collection, speed data collected from google maps (using its application programming interface (API) was investigated.

## Methods

To realize the objective of this research, four main steps were adopted according to the research methodology chart, given in Fig. 1. First, a thorough review was performed to understand and select the main factors (variables) affecting operating speeds on multilane elevated urban arterial roads. Two corridors were selected for the case study: 6th of October Bridge and Saft Al-Laban Corridor, in Greater Cairo, Egypt. An experiment was set up to collect geometric and speed data from the two corridors using a GPS device in a test vehicle. The second step was concerned with data collection. While geometric features data was collected by creating AutoCAD Civil 3D<sup>®</sup> alignments, speed data was collected from Google API<sup>®</sup> distance matrix and then validated using field measured speeds.

The third step was concerned with model development. Two types of operating models were developed: MLR and ANN models. All models were developed to predict operating speeds of a roadway segment given its geometric/operational features in addition to features from upstream segments. A stepwise regression technique was used to develop MLR operating speed models. On the other hand, several ANN architectures were investigated. While different architectures vary in number of inputs, number of hidden

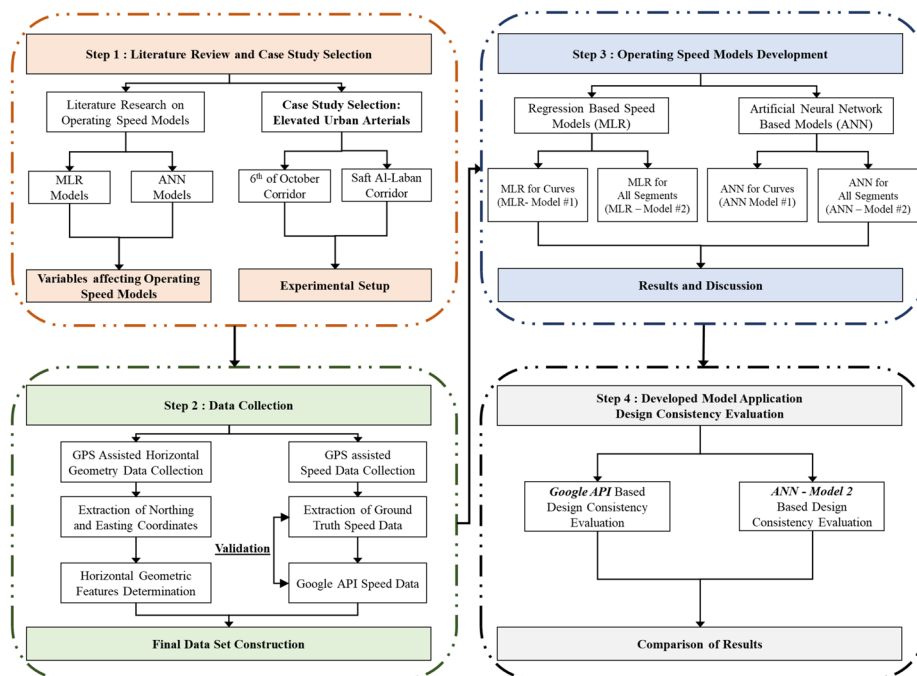


Fig. 1 Research methodology

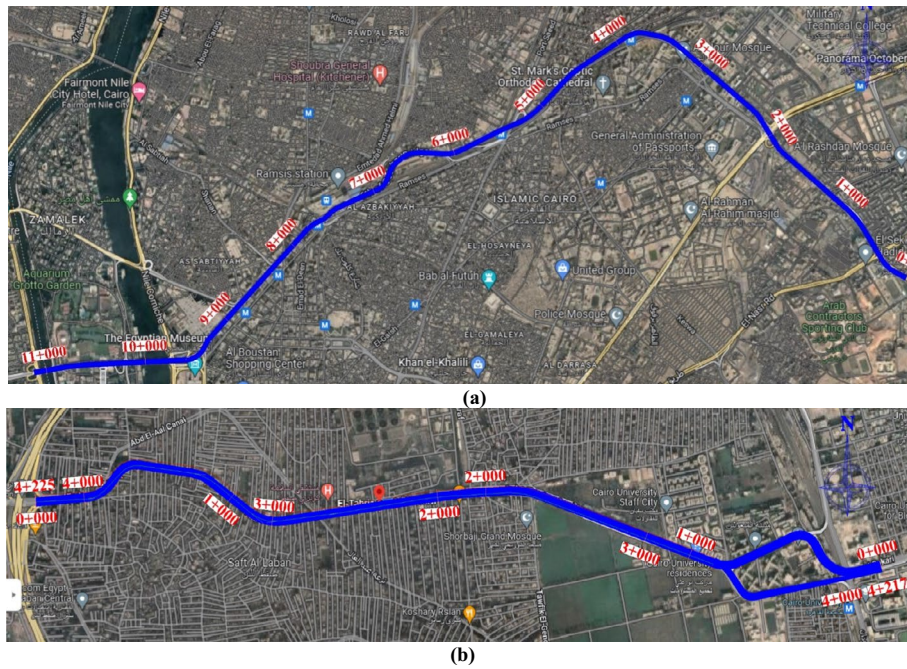
layers, number of neurons per layer, and associated functions; they all had one neuron in the output layer representing segment operating speed. Evaluation of investigated ANN models was based on prediction accuracy indicators. Finally, a design consistency evaluation exercise was undertaken to validate the performance of the developed ANN model.

### Case study data collection

Two elevated multilane urban arterials were selected for this study: 6th of October and Saft Al-Laban Corridors as shown in Fig. 2. Both corridors were chosen to ensure a wide range of sharp and smooth curves (with radii between 70 and 1400 m), and similar posted speeds of 60 km/h on most of the sections and 40 km/h on sharp alignment sections. Operating speeds data was collected during free-flow conditions to examine the effect of the geometric characteristics on vehicle speeds. Characteristics of the selected corridors are described in Table 2.

A mobile equipped with a GPS-based tracker was used in a test vehicle for collecting latitude and longitude coordinates of the two corridors. A horizontal alignment was developed for each direction of the two corridors using the vehicle’s position inside the lane. The recorded coordinates were used in AutoCAD Civil 3D® to develop horizontal alignments. The horizontal geometric features (tangent lengths, curve lengths, and curve radii) were determined. The alignments were calibrated using Google Earth® to ensure that the alignments were developed based on coordinates that can be used on Google Maps APIs. The alignments were divided (segmented) into tangents and curves to detect the impact of a certain geometric element on the following element. Geometric features were collected for each segment: radius (*R*), curve length (*CL*), and tangent length (*TL*).





**Fig. 2** Case study routes **a** 6th of October Corridor and **b** Saft Al-Laban Corridor

**Table 2** Summary of geometric features for the case study

	6th of October Corridor	Saft Al-Laban Corridor
Length in both directions	20.6 km	9.4 km
Number of lanes/direction	3–4	3
Number of curves	40	20
Posted speed (km/h)	60 (40)	60 (40)
Maximum curve radius ( $R_{max}$ )	1400 m	1300 m
Minimum curve radius ( $R_{min}$ )	110 m	70 m

GPS-based speeds were measured during free-flow conditions in between 3:00 AM and 4:00 AM, through one test run for each moving direction. This time slot was selected to ensure free-flow conditions, as our case-study corridors are heavily used during day-time. Nonetheless, the corridors are fully lighted which is believed to limit the impact of darkness on drivers’ behavior.

Since operating speed models are normally distributed and require a sample size of spot speeds of at least 30 vehicles, the conducted test run was not enough to secure reliable operating speed data. Alternatively, open-sourced travel time data provided by Google Cloud Services was used in this research to estimate operating speeds on different road segments. The data is publicly accessible through Google APIs<sup>®</sup>. While the test vehicle data represents only one run speed data, Google-based data relies on a larger pool of vehicles in addition to a historical data component. It is important to note that

the adequacy of using Google API data in capturing speed data has been investigated repeatedly in the literature [19]. Speeds were captured using a developed Python code which develops a distance matrix. The inputs for the developed code were the final segment coordinates, and the outputs were travelled distance (segment length) and travel time under free-flow condition. Since speeds were estimated under free-flow conditions, they were considered operating speeds. Google-based speed data estimation accuracy was evaluated by comparing it to the ground truth speed data of 6th of October Corridor segments. A reasonable difference was estimated; MAE and MAPE were found to be 4.2 km/h and 5% respectively which can be considered as the difference between the one run-based speed (collected from geo-tracker) and multiple vehicle speeds (collected by Google).

### Data descriptive analysis

Collected geometry and speed data were used to construct two datasets to be used in the development of two types of operating speed models. Dataset #1 was for developing operating speed models on curves only (Model #1), and Dataset #2 was for developing operating speed models on all segments (curves and tangents) (Model #2).

Dataset #1 consisted of 60 horizontal curves extracted from the developed alignments for both corridors. Table 3 shows the description of the Dataset #1 variables.

Spearman correlation analysis was adopted to analyze the correlation between the independent variables and the dependent variable in Dataset #1, using SPSS<sup>®</sup> analysis tool. The reason for using Spearman correlation is that it relaxes the normality assumption of Pearson correlation. Table 4 shows the results of the correlation analysis, where operating speed on curve ( $n$ ) ( $V_{85}C_{(n)}$ ) is highly correlated with  $NL$ ,  $PS$ ,  $R$ ,  $V_{85}T$ ,  $TL$ , and  $V_{85}C_{(n-1)}$ .

The analysis also shows that there is high correlation between the independent variables themselves which indicates a potential presence of multicollinearity between the predictors. To ensure that the multicollinearity will not affect the regression model, the variance inflation factor (VIF) was calculated for each independent variable. All predictors reported  $>1VIF > 10$  which illustrates that although there is multicollinearity between the predictors, it is not significant to affect the accuracy of the MLR model.

**Table 3** Description of dataset #1 variables

Variable Name	Symbol	Unit	Range of Measurement	Mean	Coefficient of Variation
<b>Independent variables</b>					
Number of lanes/direction	$NL$	-	3, 4	3.44	0.146
Posted speed	$PS$	km/h	40, 60	55.93	0.145
Curve length	$CL$	m	52 – 265	110.22	0.47
Curve radius	$R$	m	70 -1400	341.36	0.892
Speed on tangent	$V_{85}T$	km/h	48 – 87	67.68	0.135
Tangent length	$TL$	m	60 – 1488	353.83	1
Operating speed on previous Curve (n-1)	$V_{85}C_{(n-1)}$	km/h	50 – 90	68.83	0.139
<b>Dependent variables</b>					
Operating Speed on Curve (n)	$V_{85}C_{(n)}$	km/h	50 – 90	69.12	0.133



**Table 4** Spearman correlation analysis for dataset # 1

Variables		$V_{85}C_{(n-1)}$	$TL$	$V_{85}T$	$R$	$CL$	$PS$	$NL$	$V_{85}C_{(n)}$
$V_{85}C_{(n-1)}$	Correlation coefficient	1.000	0.323 <sup>a</sup>	0.453 <sup>b</sup>	0.326 <sup>a</sup>	0.147	0.262 <sup>a</sup>	0.363 <sup>b</sup>	0.413 <sup>b</sup>
	Sig.		0.013	0.000	0.012	0.268	0.045	0.005	0.001
$TL$	Correlation coefficient	0.323 <sup>a</sup>	1.000	0.425 <sup>b</sup>	0.359 <sup>b</sup>	0.240	0.368 <sup>b</sup>	0.126	0.412 <sup>b</sup>
	Sig.	0.013		0.001	0.005	0.067	0.004	0.340	0.001
$V_{85}T$	Correlation Coefficient	0.453 <sup>b</sup>	0.425 <sup>b</sup>	1.000	0.370 <sup>b</sup>	0.186	0.331 <sup>a</sup>	0.300 <sup>a</sup>	0.476 <sup>b</sup>
	Sig.	0.000	0.001		0.004	0.159	0.011	0.021	0.000
$R$	Correlation Coefficient	0.326 <sup>a</sup>	0.359 <sup>b</sup>	0.370 <sup>b</sup>	1.000	0.250	0.649 <sup>b</sup>	0.264 <sup>a</sup>	0.646 <sup>b</sup>
	Sig.	0.012	0.005	0.004		0.056	0.000	0.044	0.000
$CL$	Correlation Coefficient	0.147	0.240	0.186	0.250	1.000	.268 <sup>a</sup>	-0.143	0.071
	Sig.	0.268	0.067	0.159	0.056		0.040	0.279	0.593
$PS$	Correlation Coefficient	0.262 <sup>a</sup>	0.368 <sup>b</sup>	0.331 <sup>a</sup>	0.649 <sup>b</sup>	0.268 <sup>a</sup>	1.000	0.109	0.366 <sup>b</sup>
	Sig.	0.045	0.004	0.011	0.000	0.040		0.410	0.004
$NL$	Correlation Coefficient	0.363 <sup>b</sup>	0.126	0.300 <sup>a</sup>	0.264 <sup>a</sup>	-0.143	0.109	1.000	0.380 <sup>b</sup>
	Sig.	0.005	0.340	0.021	0.044	0.279	0.410		0.003
$V_{85}C_{(n)}$	Correlation Coefficient	0.413 <sup>b</sup>	0.412 <sup>b</sup>	0.476 <sup>b</sup>	0.646 <sup>b</sup>	0.071	0.366 <sup>b</sup>	0.380 <sup>b</sup>	1.000
	Sig.	0.001	0.001	0.000	0.000	0.593	0.004	0.003	

<sup>a</sup> Correlation is significant at the 0.05 level (2-tailed)

<sup>b</sup> Correlation is significant at the 0.01 level (2-tailed)

On the other hand, Dataset #2 was constructed using data on all segments (tangents and curves), from both corridors. It consisted of 122 segments (60 curves and 62 tangents). A categorical approach was adopted to differentiate between segments with respect to curvature characteristics; segments were classified into 10 categories which are described as below:

- Category #1 ( $C_1$ ): Curves with radii < 100 m
- Category #2 to Category #9 ( $C_2$  to  $C_9$ ): Curves with radii of 100 to 550 m with an increment of 50 m
- Category #10 ( $C_{10}$ ): Tangents and curves with radii > 550 m

Moreover, to investigate the effect of upstream segments' geometric and operational features on the accuracy of the prediction model, the dataset included the characteristics of four previous segments. Table 5 below shows the shape of Dataset #2, the symbol of each variable, and its range of measurements to be used in the second model that predicts the speed of all segments.

Results of the correlation analysis, presented in Table 6, illustrates that the speed on segment ( $n$ ) ( $V_{85(n)}$ ) is significantly correlated with some features of segments  $n$ ,  $n-1$ ,  $n-2$ ,  $n-3$ , and  $n-4$  ( $PS_n, L_{(n-1)}, V_{85(n-1)}, C_{(n-1)}, V_{85(n-2)}, V_{85(n-3)}, V_{85(n-4)}$ ). It is also noticed from the correlation analysis that there is multicollinearity among the independent variables.

To evaluate the effect of multicollinearity on regression Model #2, the variance inflation factor (VIF) for Dataset #2 predictors were calculated. The VIF for all independent variables was between 1 and 10 ( $1 < VIF < 10$ ) which indicated that there is multicollinearity among the independent variables, but it does not significantly affect the regression models.

**Table 5** Variables for dataset #2

Segment position	Variable name	Symbol	Unit	Range of measurement	Mean	Coefficient of variation
<b>Dependent variable</b>						
Segment (n)	Speed	$V_{85(n)}$	km/h	48—90	69.19	0.14
<b>Independent variables</b>						
Segment (n)	Length	$L_n$	m	52—1488	235.98	1.19
	Category	$C_n$	-	1—10	7.57	0.42
	Posted Speed	$PS$	km/h	40, 60	58.03	0.1
Segment (n-1)	Length	$L_{(n-1)}$	m	52—1488	232.04	1.22
	Speed	$V_{85(n-1)}$	km/h	48—90	69.38	0.14
	Category	$C_{(n-1)}$	-	1—10	7.49	0.43
Segment (n-2)	Length	$L_{(n-2)}$	M	52—1488	236.16	1.21
	Speed	$V_{85(n-2)}$	km/h	48—90	69.43	0.14
	Category	$C_{(n-2)}$	-	1—10	7.55	0.41
Segment (n-3)	Length	$L_{(n-3)}$	m	52—1488	226.66	1.21
	Speed	$V_{85(n-3)}$	km/h	48—90	69.54	0.14
	Category	$C_{(n-3)}$	-	1—10	7.46	0.42
Segment (n-4)	Length	$L_{(n-4)}$	m	52—1488	230.93	1.2
	Speed	$V_{85(n-4)}$	km/h	48—90	69.48	0.14
	Category	$C_{(n-4)}$	-	1—10	7.63	0.4

### Model development

This research investigated two modeling techniques: multilinear regression (MLR) and feedforward artificial neural networks (ANN). Each of the mentioned modeling techniques was used to develop two types of models. Model #1 predicts speed on curves only using predictors from Dataset #1. Model #2 predicts speed on any segment (tangent or curve) using predictors from Dataset #2. MLR model was developed using SPSS<sup>®</sup>, while ANN model was developed using a tailored Python script. The dataset used in developing the ANN models was divided into training and testing data (with a ratio of 70:30). The predicting power of each modeling technique was assessed by comparing the prediction accuracy of each model.

#### MLR-Model #1 speed on curves

A stepwise procedure was used to model operating speed on curves using Dataset #1 as shown in Eq. 30 with  $MAE=7.87$  km/h,  $MAPE=11.66\%$ , and  $R^2=0.37$ . The positive sign of the coefficient for the radius ( $R$ ) is logical as the speed on curve ( $V_{85C}$ ) increases with the increase in curve radius. Moreover, the positive sign for the coefficient ( $V_{85T}$ ) also indicates that the higher the speed on the tangent before the curve, the higher the expected speed on the curve. Table 7 presents the significance of each predictor.

$$V_{85C} = 36.6 + .015R + 0.341V_{85T} \quad (30)$$

**Table 6** Spearman correlation analysis for dataset # 2

Variables	$V_{85}(n)$	$L_n$	$C_n$	$PS_n$	$L_{(n-1)}$	$V_{85}(n-1)$	$C_{(n-1)}$	$L_{(n-2)}$	$V_{85}(n-2)$	$C_{(n-2)}$	$L_{(n-3)}$	$V_{85}(n-3)$	$C_{(n-3)}$	$L_{(n-4)}$	$V_{85}(n-4)$	$C_{(n-4)}$
$V_{85}(n)$	Corr	1	0.094	0.065	.184 <sup>a</sup>	.331 <sup>b</sup>	.404 <sup>b</sup>	.293 <sup>b</sup>	.443 <sup>b</sup>	-0.003	0.12	.280 <sup>b</sup>	0.126	0.119	.316 <sup>b</sup>	-0.038
	Sig		0.169	0.253	0.03	0	0.001	0.377	0	0.486	0.111	0.002	0.1	0.113	0	0.349
$L_n$	Corr	0.094	1	.493 <sup>b</sup>	.286 <sup>b</sup>	-0.04	.249 <sup>b</sup>	.226 <sup>b</sup>	0.057	.477 <sup>b</sup>	-0.086	0.057	-4.25 <sup>b</sup>	.232 <sup>b</sup>	0.001	.444 <sup>b</sup>
	Sig	0.169		0	0.001	0.342	0.002	0.01	0.28	0	0.191	0.282	0	0.008	0.495	0
$C_n$	Corr	0.065	.493 <sup>b</sup>	1	.515 <sup>b</sup>	.253 <sup>b</sup>	-.661 <sup>b</sup>	.444 <sup>b</sup>	-0.048	.791 <sup>b</sup>	-.374 <sup>b</sup>	.198 <sup>a</sup>	-6.72 <sup>b</sup>	.369 <sup>b</sup>	-0.111	.640 <sup>b</sup>
	Sig	0.253	0	0	0	0.003	0.004	0	0.313	0	0	0.021	0	0.128	0	
$PS_n$	Corr	.184 <sup>a</sup>	.286 <sup>b</sup>	.515 <sup>b</sup>	1	.226 <sup>b</sup>	-.254 <sup>b</sup>	.237 <sup>b</sup>	-0.001	.420 <sup>b</sup>	-0.061	.171 <sup>a</sup>	-2.58 <sup>b</sup>	.183 <sup>a</sup>	-0.05	.214 <sup>a</sup>
	Sig	0.03	0.001	0	0.23	0.01	0.004	0.007	0.496	0	0.266	0.04	0.004	0.03	0.304	0.014
$L_{(n-1)}$	Corr	.331 <sup>b</sup>	-0.04	-.264 <sup>b</sup>	0.072	0.137	.455 <sup>b</sup>	-0.005	.271 <sup>b</sup>	-.264 <sup>b</sup>	.236 <sup>b</sup>	0.083	.443 <sup>b</sup>	-0.061	0.049	-.373 <sup>b</sup>
	Sig	0	0.342	0.003	0.23	0.081	0	0.478	0.002	0.003	0.007	0.197	0	0.268	0.309	0
$V_{85}(n-1)$	Corr	.404 <sup>b</sup>	.249 <sup>b</sup>	.253 <sup>b</sup>	.226 <sup>b</sup>	1	0.139	.347 <sup>b</sup>	.423 <sup>b</sup>	.275 <sup>b</sup>	0.043	.451 <sup>b</sup>	0.052	.167 <sup>a</sup>	.305 <sup>b</sup>	.188 <sup>a</sup>
	Sig	0	0.005	0.004	0.01	0.081	0.078	0	0	0.002	0.332	0	0.297	0.044	0.001	0.027
$C_{(n-1)}$	Corr	.293 <sup>b</sup>	-.278 <sup>b</sup>	-.661 <sup>b</sup>	-.254 <sup>b</sup>	.455 <sup>b</sup>	1	-.216 <sup>a</sup>	.286 <sup>**</sup>	-.672 <sup>b</sup>	.439 <sup>b</sup>	0	.766 <sup>b</sup>	-.352 <sup>b</sup>	.250 <sup>b</sup>	-.661 <sup>b</sup>
	Sig	0.001	0.002	0	0.004	0	0.078	0.013	0.001	0	0	0.499	0	0	0.005	0
$L_{(n-2)}$	Corr	0.031	0.226 <sup>b</sup>	.444 <sup>b</sup>	.237 <sup>b</sup>	.347 <sup>b</sup>	-.216 <sup>a</sup>	1	0.151	.428 <sup>b</sup>	-0.024	.290 <sup>b</sup>	-2.05 <sup>a</sup>	.242 <sup>b</sup>	0.067	.431 <sup>b</sup>
	Sig	0.377	0.01	0	0.007	0.478	0.013	0	0.062	0	0.404	0.001	0.017	0.006	0.246	0
$V_{85}(n-2)$	Corr	.443 <sup>b</sup>	0.057	-0.048	-0.001	.423 <sup>b</sup>	.286 <sup>b</sup>	0.151	1	0.081	.325 <sup>b</sup>	.433 <sup>b</sup>	.321 <sup>b</sup>	0.022	.453 <sup>b</sup>	0.015
	Sig	0	0.28	0.001	0.496	0.002	0	0.062	0	0.204	0	0	0	0.413	0	0.438
$C_{(n-2)}$	Corr	-0.003	.477 <sup>b</sup>	.791 <sup>b</sup>	.420 <sup>b</sup>	.275 <sup>b</sup>	-.672 <sup>b</sup>	.428 <sup>b</sup>	0.081	1	-.294 <sup>b</sup>	.228 <sup>b</sup>	-.683 <sup>b</sup>	.425 <sup>b</sup>	-0.078	.791 <sup>b</sup>
	Sig	0.486	0	0	0	0.003	0.002	0	0.204	0	0.001	0.009	0	0.214	0	
$L_{(n-3)}$	Corr	0.12	-0.086	-.374 <sup>b</sup>	-0.061	.325 <sup>b</sup>	.439 <sup>b</sup>	-0.024	.325 <sup>b</sup>	-2.94 <sup>b</sup>	1	0.123	.433 <sup>b</sup>	-0.034	.250 <sup>b</sup>	-.223 <sup>a</sup>
	Sig	0.111	0.191	0	0.266	0.007	0.332	0.404	0	0.001	0.123	0.104	0	0.364	0.005	0.011
$V_{85}(n-3)$	Corr	.280 <sup>b</sup>	0.057	.198 <sup>a</sup>	.171 <sup>a</sup>	.451 <sup>b</sup>	0	.290 <sup>b</sup>	.433 <sup>b</sup>	.228 <sup>b</sup>	0.123	1	0.152	.331 <sup>b</sup>	.418 <sup>b</sup>	.284 <sup>b</sup>
	Sig	0.002	0.282	0.021	0.04	0.197	0	0.001	0	0.009	0.104	0	0.059	0	0	0.002
$C_{(n-3)}$	Corr	0.126	-.425 <sup>b</sup>	-.672 <sup>b</sup>	-.258 <sup>b</sup>	.443 <sup>b</sup>	.766 <sup>b</sup>	-.205 <sup>a</sup>	.321 <sup>b</sup>	-.683 <sup>b</sup>	.433 <sup>b</sup>	0.152	1	-.252 <sup>b</sup>	.260 <sup>b</sup>	-.672 <sup>b</sup>
	Sig	0.1	0	0	0.004	0	0.297	0	0	0	0	0.059	0	0.005	0.004	0

**Table 6** (continued)

Variables	$V_{85(n)}$	$L_n$	$C_n$	$PS_n$	$L_{(n-1)}$	$V_{85(n-1)}$	$C_{(n-1)}$	$L_{(n-2)}$	$V_{85(n-2)}$	$C_{(n-2)}$	$L_{(n-3)}$	$V_{85(n-3)}$	$C_{(n-3)}$	$L_{(n-4)}$	$V_{85(n-4)}$	$C_{(n-4)}$
$L_{(n-4)}$	Corr	0.119	.232 <sup>b</sup>	.183 <sup>a</sup>	-0.061	.167 <sup>a</sup>	-.352 <sup>b</sup>	.242 <sup>b</sup>	0.022	.425 <sup>b</sup>	-0.034	.331 <sup>b</sup>	-.252 <sup>b</sup>	1	0.114	.434 <sup>b</sup>
	Sig	0.113	0.008	0.03	0.268	0.044	0	0.006	0.413	0	0.364	0	0.005		0.123	0
$V_{85(n-4)}$	Corr	.316 <sup>b</sup>	0.001	-0.05	0.049	.305 <sup>b</sup>	.250 <sup>b</sup>	0.067	.453 <sup>b</sup>	-0.078	.250 <sup>b</sup>	.418 <sup>b</sup>	.260 <sup>b</sup>	0.114	1	0.109
	Sig	0	0.495	0.304	0.309	0.001	0.005	0.246	0	0.214	0.005	0	0.004	0.123		0.133
$C_{(n-4)}$	Corr	-0.038	.444 <sup>b</sup>	.214 <sup>a</sup>	-.373 <sup>b</sup>	.188 <sup>a</sup>	-.661 <sup>b</sup>	.431 <sup>b</sup>	0.015	.791 <sup>b</sup>	-.223 <sup>a</sup>	.284 <sup>b</sup>	-.672 <sup>b</sup>	.434 <sup>b</sup>	0.109	1
	Sig	0.349	0	0.014	0	0.027	0	0	0.438	0	0.011	0.002	0	0	0.133	

<sup>a</sup> Correlation is significant at the 0.05 level

<sup>b</sup> Correlation is significant at the 0.01 level

**Table 7** MLR Models statistical significance

Model	Predictor	Coefficient	Sig.
MLR-Model #1	Intercept	36.597	0.001
	<i>R</i>	0.015	0.000
	$V_{85T}$	0.341	0.000
MLR-Model #2	Intercept	14.313	0.115
	$V_{85(n-2)}$	0.255	0.004
	$V_{85(n-1)}$	0.201	0.022
	$C_{(n-1)}$	0.814	0.002
	<i>PS</i>	0.303	0.026

**Table 8** ANN-Model #1 developed trials

	ANN M 1-1	ANN M 1-2	ANN M 1-3	ANN M 1-4
Inputs	<i>PS</i>	<i>PS</i>	<i>PS</i>	<i>PS</i>
	<i>NL</i>	<i>NL</i>	<i>NL</i>	<i>NL</i>
	<i>R</i>	<i>R</i>	<i>R</i>	<i>R</i>
	<i>CL</i>	<i>CL</i>	<i>CL</i>	<i>CL</i>
	--	<b><i>TL</i></b>	<b><i>TL</i></b>	<b><i>TL</i></b>
	--	--	$V_{85T}$	$V_{85T}$
			$V_{85}C_{(n-1)}$	

**MLR-Model #2 speed on all segments**

For MLR-Model #2, the significantly correlated variables from Dataset #2 were used to develop the best regression model using stepwise regression approach. The best model is shown in Eq. 31 and Table 7. The model is significant at a 95% confidence level with coefficient of determination ( $R^2$ ) equal to 0.31,  $MAE=6.17$  km/h, and  $MAPE=9.03\%$ .

$$V_{85(n)} = 14.313 + 0.201V_{85(n-1)} + 0.255V_{85(n-2)} + 0.814C_{(n-1)} + 0.303PS \quad (31)$$

The developed model reflects the impact of the residual speed from previous segments. The operating speed of the previous two segments ( $V_{85(n-1)} V_{85(n-2)}$ ), the radius category of the previous segment ( $C_{(n-1)}$ ), and the posted speed on the road ( $PS$ ) were all of significant impact.

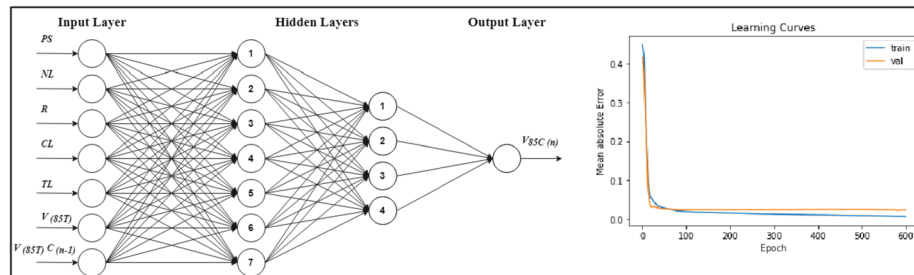
**ANN-Model #1 speed on curves**

ANN is commonly used to capture nondeterministic relationships between predictors and dependent variables. In the case of Model #1, the correlation analysis showed that there is high correlation between  $V_{85}C_{(n)}$  as the dependent variable and  $NL, PS, R, V_{85T}, TL,$  and  $V_{85}C_{(n-1)}$  as independent variables. However, the MLR model considered  $R$  and  $V_{85T}$  only as significant predictors. Thus, a feedforward ANN was used to model the operating speed considering all predictors of Dataset #1.

Since ANN-Model #1 was developed to predict speed on curves  $V_{85}C_{(n)}$ , therefore, tangent length ( $TL$ ), approaching speed (speed on tangent)  $V_{85T}$ , and speed on the previous

**Table 9** Model #1 prediction errors

	ANN models		MLR-Model #1	
	MAE (km/h)	MAPE (%)	MAE (km/h)	MAPE (%)
<b>M 1-1</b>	9.244	13.4%	7.87	11.66
<b>M 1-2</b>	7	10.1%		
<b>M 1-3</b>	6.5	9.4%		
<b>M 1-4</b>	5.5	8.0%		



**Fig. 3** ANN M 1–4 architecture and learning curve

curve  $V_{85}C_{(n-1)}$  are considered the geometric and operational properties of upstream segments. To determine the effect of adding the upstream characteristics on the accuracy of the operating speed model, a stepwise criterion was adopted in developing ANN-Model #1. The effect of adding each predictor was measured by developing four separate ANN models shown in Table 8. Models’ architectures were developed through a comprehensive trial error approach. The dataset was divided into training and testing subsets with a ratio of 70:30.

The effect of adding each predictor was measured by calculating MAE and MAPE for each model as shown in Table 9. The performance of ANN-Model #1–4 is considered the best performance among the developed models (MAPE of 8%) which illustrates that adding the upstream segments’ features has increased the accuracy of operating speed prediction model. Extensive efforts were exerted to fine tune the ANN model architecture to achieve the best possible performance. ANN M 1–4 had two hidden layers, with 7 and 4 neurons, respectively, and tanh activation function. Figure 3 presents M 1–4 model architecture and learning curve.

**ANN-Model#2 speed on all segments**

The developed MLR model considered four predictors only which were posted speed  $PS$ , category of the previous segment  $C_{(n-1)}$ , approaching speed  $V_{85(n-1)}$ , and speed on segment  $n-2$   $V_{85(n-2)}$ , although there were other predictors which were significantly correlated to the operating speed on segment  $n$   $V_{85(n)}$ . ANN-Model #2 was developed to consider all predictors of Dataset #2. To understand the contribution of data from previous segments, the model development process was conducted by developing five ANN models, as shown in Table 10.

Table 11 presents MAE and MAPE for each of the developed models. The results show that although models ANN M 2–4 and ANN M2-5 have more predictors



**Table 10** ANN-Model #2 developed trials

Model name	Inputs
ANN M 2-1	$L_n, C_n, PS_n$
ANN M 2-2	$L_n, C_n, PS_n$ $L^{(n-1)}, V_{85}^{(n-1)}, C^{(n-1)}$
ANN M 2-3	$L_n, C_n, PS_n$ $L^{(n-1)}, V_{85}^{(n-1)}, C^{(n-1)}$ $L^{(n-2)}, V_{85}^{(n-2)}, C^{(n-2)}$
ANN M 2-4	$L_n, C_n, PS_n$ $L^{(n-1)}, V_{85}^{(n-1)}, C^{(n-1)}$ $L^{(n-2)}, V_{85}^{(n-2)}, C^{(n-2)}$ $L^{(n-3)}, V_{85}^{(n-3)}, C^{(n-3)}$
ANN M 2-5	$L_n, C_n, PS_n$ $L^{(n-1)}, V_{85}^{(n-1)}, C^{(n-1)}$ $L^{(n-2)}, V_{85}^{(n-2)}, C^{(n-2)}$ $L^{(n-3)}, V_{85}^{(n-3)}, C^{(n-3)}$ $L^{(n-4)}, V_{85}^{(n-4)}, C^{(n-4)}$

**Table 11** Model #2 prediction errors

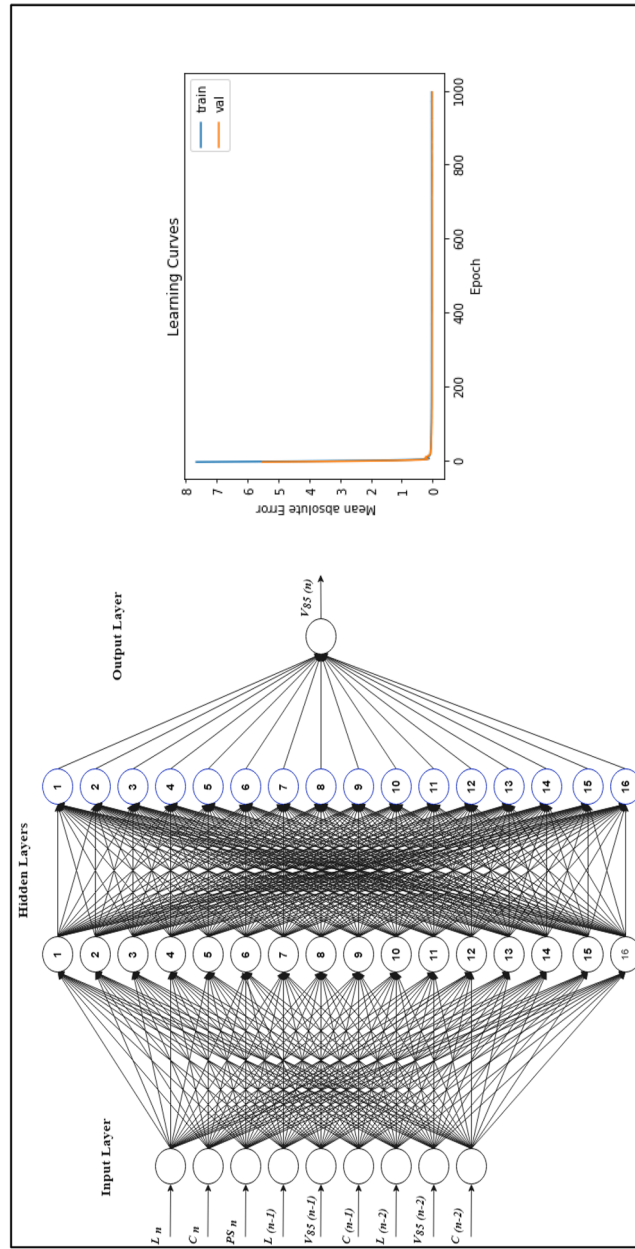
	ANN models		MLR Model #2	
	MAE (km/h)	MAPE (%)	MAE (km/h)	MAPE (%)
<b>M 2-1</b>	7.1	10.3%	6.17	9.03
<b>M 2-2</b>	5.7	8.3%		
<b>M 2-3</b>	4.6	6.7%		
<b>M 2-4</b>	5.1	7.4%		
<b>M 2-5</b>	5.2	7.5%		

including the operating speed and category of segments ( $n-3$ ) and ( $n-4$ ), the MAE and MAPE values were higher than those for ANN M 2-3. The best-performing model was ANN M 2-3, with  $MAE=4.65$  km/h,  $MAPE=6.7\%$ , and  $R^2=0.56$ . The model uses input data (predictors) from segment ( $n$ ) until segment ( $n-2$ ). The model has two hidden layers, 16 neurons each, and tanh activation function. Figure 4 presents ANN M 2-3 architecture and learning curve. On the other hand, using data from further segments ( $n-3$ ) and ( $n-4$ ) negatively affected the accuracy of the model as seen in the MAE and MAPE values of the 4th and 5th models. This implies that driver’s choice of speed was impacted by their previous experience (represented by their speed) on the previous two segments only ( $n-1$ ) and ( $n-2$ ), and that their speed on the segments before that ( $n-3$ ) and ( $n-4$ ) were not a contributing factor.

**Results and discussion**

**Model #1: Speed on curves**

ANN performance surpassed the performance of MLR as  $R^2$  and MAPE for ANN M1-4 were 0.5 and 8% respectively compared to 0.37 and 11.66% for MLR-Model #1. To further investigate the robustness of ANN M1-4, an error distribution analysis was conducted on the testing dataset. As shown in Fig. 5, the distribution of errors takes a normal shape with 40% of the error values in the range of 0 to 5 km/h.



**Fig. 4** ANN M2-3 architecture and learning curve

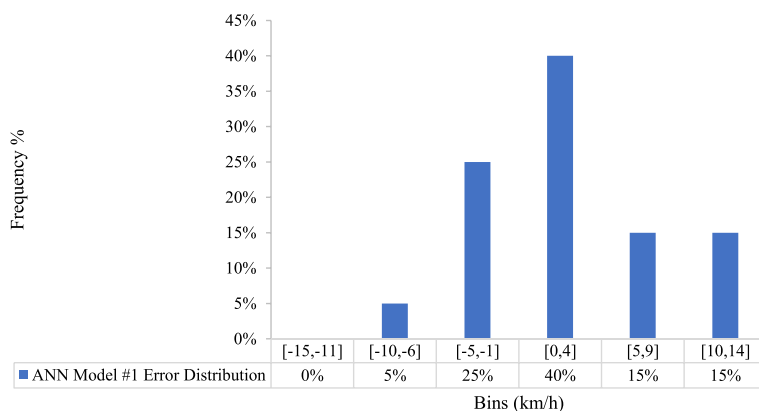


Fig. 5 Error distribution for ANN M 1–4

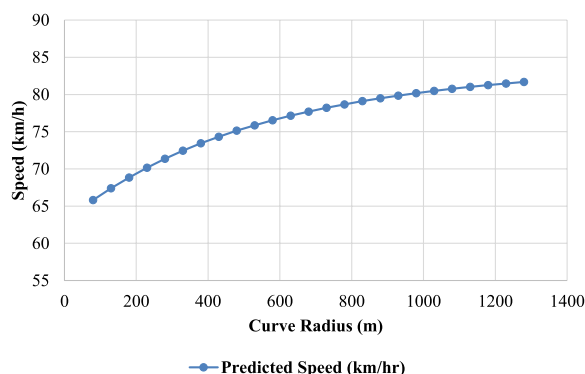


Fig. 6 Sensitivity of the predicted operating speed to change in curve radius

The ability of the developed ANN M1-4 to capture the sensitivity of operating speeds to changes in radius was examined on a curved segment from 6th of October corridor. The segment has the following characteristics ( $NL = 3$  lanes,  $PS = 60$  km/h,  $CL = 168$  m,  $R = 180$  m,  $V_{85T} = 68$  km/h,  $TL = 113$  m, and  $V_{85C_{(n-1)}} = 63$  km/h). All input parameters were fixed, and changes in curve radius have been performed incrementally to observe the change in predicted operating speed. It can be noticed from Fig. 6 that the rate of increase in the predicted operating speed decreases with the increase in curve radius. More specifically, the predicted operating speed is highly sensitive to the change in curve radius from  $R = 80$  m to  $R = 480$  m (10 km/h increase). Then, a reduced sensitivity could be depicted for curves with  $480 \text{ m} < R < 1280$  m.

**Model #2: Speed on all segments**

The performance of ANN M2-3 reported the best performance with  $R^2$  and MAPE of 0.56 and 6.7% respectively compared to 0.31 and 9.03% for MLR-Model #2. The testing data for ANN M2-3 was examined to determine the distribution of errors. The results shown in Fig. 7 depict a normal distribution of errors with a peak in the error range 0 to 5 km/h.

For further understanding of the behavior of the developed ANN M2-3, a sensitivity analysis was performed on a segment from 6th of October Corridor with the following

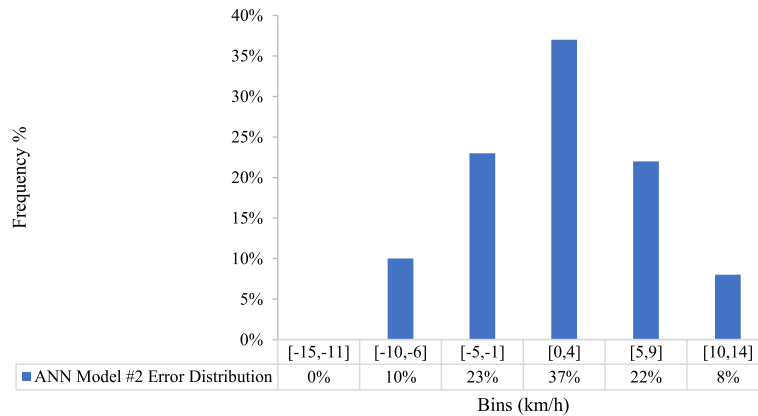


Fig. 7 Error distribution for ANN-Model #2  $V_{85}$  estimation

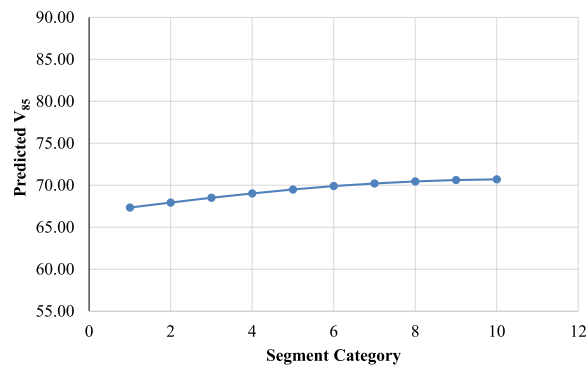


Fig. 8 ANN-Model #2 sensitivity to the change in segment (n) category

features ( $L_n=106$  m,  $C_n=4$ ,  $PS_n=60$  km/h,  $L_{(n-1)}=175$  m,  $V_{85(n-1)}=70$  km/h,  $C_{(n-1)}=10$ ,  $L_{(n-2)}=103$  m,  $V_{85(n-2)}=74$  km/h,  $C_{(n-2)}=5$ ). The effect of segment category  $C_n$  on the operating speed was examined by changing the category of the selected segment from 1 to 10, while the other variables were fixed. It can be noticed from Fig. 8 that the model sensitivity to changes in segment category is limited (less than 5 km/h) and more pronounced in categories 1 to 6.

The conducted sensitivity analysis depicts a variation in sensitivity to segments' curvature in between Model #1 and Model #2. Model #1 predicts operating speeds only on curves, and hence, the data set used for training is specific to curved segments. The narrow data range allowed for higher sensitivity to curve radii. On the other hand, Model #2 is a generalized version that predicts operating speeds on any segment (tangents or curves). This comes at the expense of a wider range of data resulting in diluted sensitivity to the impact of curve radii.

### Validation of the usability of ANN-Model #2

#### Assessment of model usability in design consistency evaluation

Design consistency is one of the major concerns in geometric design of roads as it is considered one of the main factors that control the quality of the trip. Lamm et al. [20]

developed three main criteria to evaluate the design consistency of highways; each evaluation criterion depends on evaluating the consistency of a specific measure through the successive road segments. The safety criterion (II) depends on evaluating the operating speed. In this research, the developed models predict operating speeds for arterial roads; thus, safety criterion (II) was used to evaluate the design consistency of 6th of October and Saft Al-Laban Corridors. The safety criteria developed by Lamm et al. [20] were for two-lane, two-way highways, where a good design had a change in operating speed between two successive elements of less than 10 km/h, a fair design had a change of operating speed between two successive elements between 10 km/h and 20 km/h, and a poor design had a change of operating speed of more than 20 km/h. In this study, the evaluation ranges of criterion II were modified to fit the arterial roads based on the design speed ratio between highways and arterial roads as shown in Table 12.

A comparative design consistency evaluation was conducted on each movement direction for the case study roads by using the proposed modified ranges of Safety Criterion II presented in Table 12. The first consistency evaluation was conducted for the case study road segments using the operating speed data obtained from Google Maps Distance Matrix API as a reference for the comparative evaluation. The second evaluation was conducted for the same segments using the predicted operating speed data ( $V_{85}$ ) based on the trained ANN-M2-3. The purpose of the application was to compare the design consistency rating using the ground truth data (raw data) operating speed values and the ratings from operating speeds inferred from the ANN model to test how different were the two results and whether the errors of the models resulted in different consistency ratings.

The results of the comparative design consistency evaluations are presented in Table 13 for 6th of October Corridor and Table 14 for Saft Al Laban Corridor and summarized in Table 15. By comparing the conducted evaluations for the two corridors, it was found that 71% of the segments had the same evaluation category, while 29% of the segments had different evaluation rating. Furthermore, the segments with different consistency evaluation were found to be 14 tangents and 15 curves with different range of characteristics which indicates that there is no pattern/bias in observed differences. The differences, though, could be justified by two main reasons. The first reason is the accuracy of the developed ANN model which reflects the prediction power of the model. It was noticed that the majority of evaluation differences (97%) are one category difference, and the remaining 3% is two categories difference. Referring to the MAE of ANN 2–3 which was 4.6 km/h and by comparing it with the speed range difference of each evaluation category which is 7 km/h per category, it can be concluded that the reason for most of the differences in the consistency evaluation (one category difference) is because of the MAE of the prediction model itself. The second reason which was identified through

**Table 12** Modified ranges for operating speed-based evaluation for elevated urban arterials

Evaluation	Operating speed consistency
Good	$\Delta V_{85} \leq 7$ km/h
Fair	$7 < \Delta V_{85} \leq 14$ km/h
Poor	$\Delta V_{85} > 14$ km/h

**Table 13** Comparative consistency evaluation results for 6th of October Corridor

6th of October Westbound							6th of October Eastbound						
#	APIs	ANN	Type	R (m)	L (m)	C	#	APIs	ANN	Type	R (m)	L (m)	C
1	F	G	T	-	134	10	1	G	F	T	-	366	10
2	G	G	C	300	161	5	2	F	G	C	1200	147	10
3	G	G	T	-	395	10	3	G	G	T	-	1025	10
4	G	G	C	650	202	10	4	P	F	C	200	215	4
5	G	F	T	-	889	10	5	F	F	T	-	59	10
6	G	G	C	400	123	7	6	G	G	C	200	57	4
7	G	G	T	-	286	10	7	G	G	T	-	1491	10
8	G	G	C	400	128	7	8	F	G	C	200	77	4
9	G	G	T	-	1010	10	9	G	G	T	-	63	10
10	F	G	C	400	199	7	10	G	G	C	300	105	5
11	G	G	T	-	156	10	11	F	G	T	-	50	10
12	G	G	C	250	199	5	12	G	G	C	200	75	4
13	G	G	T	-	555	10	13	G	G	T	-	461	10
14	G	G	C	1000	237	10	14	F	F	C	110	117	2
15	G	G	T	-	340	10	15	G	F	T	-	111	10
16	G	G	C	700	318	10	16	G	G	C	200	114	4
17	G	G	T	-	69	10	17	G	G	T	-	64	10
18	G	G	C	1000	98	10	18	F	G	C	300	210	5
19	G	G	T	-	442	10	19	G	G	T	-	233	10
20	F	G	C	300	159	5	20	G	G	C	300	159	5
21	G	G	T	-	233	10	21	G	G	T	-	442	10
22	G	G	C	300	210	5	22	G	G	C	1000	98	10
23	G	G	T	-	64	10	23	G	G	T	-	69	10
24	G	G	C	200	114	4	24	G	G	C	700	318	10
25	G	G	T	-	111	10	25	G	G	T	-	340	10
26	F	G	C	110	117	2	26	G	G	C	1000	237	10
27	F	F	T	-	461	10	27	F	F	T	-	555	10
28	G	G	C	200	75	4	28	G	G	C	250	199	5
29	G	G	T	-	35	10	29	F	G	T	-	156	10
30	F	G	C	300	105	5	30	G	F	C	400	199	7
31	G	G	T	-	63	10	31	G	G	T	-	1010	10
32	G	G	C	200	77	4	32	G	G	C	400	128	7
33	G	G	T	-	1491	10	33	G	G	T	-	286	10
34	G	G	C	200	57	4	34	G	F	C	400	123	7
35	G	G	T	-	59	10	35	F	G	T	-	889	10
36	G	G	C	200	215	4	36	G	G	C	650	202	10
37	F	F	T	-	1025	10	37	F	G	T	-	395	10
38	G	F	C	1200	147	10	38	G	G	C	300	161	5
39	G	G	T	-	366	10	39	G	G	T	-	134	10

\*F Fair, G Good, P Poor

visual observations is the localized operational conditions that are not captured by the developed model and could have an impact on operating speeds such as radars, cameras, or deteriorated pavement condition in specific parts of some segments.

**Assessment of model transferability**

To further assess the developed ANN-Model #2 transferability, a validation exercise was undertaken to evaluate the potential application of ANN-Model #2 on other elevated urban arterials with similar characteristics. A test was conducted on both directions of a 1.7-km stretch of an elevated 6-lane divided urban arterial in Cairo, Egypt (Rod El Farag Corridor). Figure 9 and Table 16 display the geometric/operational characteristics of the considered portion of the corridor. Such data was used as model inputs to predict operating speeds on each segment of the corridor using ANN-Model #2.



**Table 14** Comparative consistency evaluation results for Saft Al Laban Corridor

Saft Al Laban Westbound							Saft Al Laban Eastbound						
#	APIs	ANN	Type	R (m)	L (m)	C	#	APIs	ANN	Type	R (m)	L (m)	C
1	G	G	T	-	147	10	1	G	G	T	-	65	10
2	G	G	C	120	160	2	2	G	G	C	100	76	2
3	G	G	T	-	55	10	3	G	G	T	-	92	10
4	G	G	C	90	132	1	4	G	G	C	100	94	2
5	P	G	T	-	277	10	5	G	G	T	-	273	10
6	G	G	C	150	102	3	6	G	G	C	150	76.1	3
7	G	F	T	-	837	10	7	G	G	T	-	196	10
8	G	G	C	330	123	5	8	G	G	C	150	110	3
9	G	G	T	-	139	10	9	G	F	T	-	122	10
10	G	G	C	1300	117	10	10	F	G	C	800	54	10
11	G	G	T	-	239	10	11	F	G	T	-	469	10
12	G	F	C	1300	73	10	12	G	G	C	980	87	10
13	G	G	T	-	439	10	13	F	G	T	-	317	10
14	G	G	C	200	159	4	14	F	G	C	300	111	5
15	G	G	T	-	187	10	15	F	G	T	-	852	10
16	G	G	C	150	76	3	16	G	F	C	110	78	2
17	G	G	T	-	274	10	17	G	G	T	-	57	10
18	G	G	C	100	93	2	18	G	G	C	70	78	1
19	G	G	T	-	93	10	19	G	G	T	-	333	10
20	G	G	C	100	72	2							
21	G	G	T	-	280	10							

\*F Fair, G Good, P Poor

**Table 15** Summary of consistency evaluation results

Consistency Level	Number of Segments							
	6 <sup>th</sup> of October Corridor				Saft Al-Laban Corridor			
	Eastbound		Westbound		Eastbound		Westbound	
	Actual API	ANN	Actual API	ANN	Actual API	ANN	Actual API	ANN
Good	28	31	32	35	14	17	19	18
Fair	10	8	7	4	5	2	0	2
Poor	1	0	0	0	0	0	1	0



**Fig. 9** Assessed route for validation

**Table 16** Descriptive statistics of assessed route

Variable	Range of measurement	Mean	Coefficient of variation
Length (m)	2681	-	-
No. of tangents	6	-	-
No. of curves	4	-	-
No. of lanes / direction	3	-	-
Posted speed (km/h)	60	-	-
Segment length (m)	40—856	221.43	1.305
Curve radius (m)	500—525	512.5	0.03
Segment category	9—10	9.5	0.055

**Table 17** Speed estimates and prediction errors of assessed route

Segment no.	Segment type	Segment length (m)	API speed (km/h)	Predicted speed (km/h)	Error (km/h)	% error
1	Tangent	235	60	68	8	13
2	Curve	40	72	73	1	1
3	Tangent	856	70	68	-2	3
4	Curve	162	73	74	2	3
5	Tangent	150	77	74	-3	4
6	Tangent	856	70	75	5	7
7	Curve	40	72	73	1	1
8	Tangent	235	60	64	4	7
9	Curve	56	61	63	2	3
10	Tangent	51	67	70	3	4

The performance of the developed ANN-Model #2 was tested by comparing the model predicted speeds to speeds obtained from Google Distance Matrix API (representing ground truth data). Table 17 depicts speed estimates and prediction errors. These results are for 10 out of the 14 original segments, since the first two segments from both directions were excluded from the prediction exercise. The operating speed on these excluded segments was assumed as the posted speed and used as input in predicting the operating speed of the following segment.

Estimated prediction errors ranged between 1 and 7% for all segments except for the first segment (where the prediction error increased to 13%). The estimated average prediction error was 4.6% (around 3 km/h). Reported results highlight the potential transferability of the developed model as it was able to reasonably predict operating speed patterns on a new urban arterial.

## Conclusions

This research focused on the development of operating speed prediction models in Egypt using two modeling approaches: multiple linear regression and artificial neural networks based on speed data extracted from Google Distance Matrix APIs. Two distinct models were developed in this research, Model #1 to predict operating speed on curves only ( $V_{85C}$ ) and Model #2 to predict speed on any segment (tangents and curves) ( $V_{85n}$ ). Notably, no separate model was developed for speeds on tangents

since speeds on the tangents largely depend on the length of the tangent, where short tangents will have speeds dependent on that of preceding and succeeding curves. The variability and driver's choice of speed mainly depend on the geometrics of the horizontal curves experienced by the driver. To our knowledge, this is the first research attempt to investigate the following: (1) the residual impact of upstream segments' operating speeds (segments  $n-1$ ,  $n-2$ ,  $n-3$ , and  $n-4$ ) on operating speed of a given segment ( $n$ ) and (2) the feasibility of developing one operating speed model that can predict operating speeds on both curves and tangents with acceptable level of accuracy.

A preliminary analysis was conducted on collected data to identify correlation patterns between different variables. Seven independent variables were selected to be investigated as potential inputs to Model#1, namely, number of lanes  $NL$ , posted speed  $PS$ , curve length  $CL$ , curve radius  $R$ , speed on tangent  $V_{85T}$ , tangent length  $TL$ , and operating speed on previous curve  $V_{85C_{(n-1)}}$ . A stepwise regression modeling exercise was undertaken. The developed model used only two independent variables ( $R$ ,  $V_{85T}$ ) and reported  $R^2$  of 0.37 and MAPE of 11.66%. On the other hand, the best performing ANN model used all seven variables and reported  $R^2$  of 0.5 and MAPE of 8%.

As for Model #2, fifteen independent variables, capturing geometric/operation features of the modeled segment  $n$  and upstream segments  $n-1$ ,  $n-2$ ,  $n-3$ , and  $n-4$ , have been investigated. The developed regression models used four independent variables: posted speed, operating speeds on segments  $n-1$  and  $n-2$ , and curvature category of segment  $n-1$ . The model reported  $R^2$  of 0.31 and MAPE of 9.03%. On the other hand, the best-performing ANN model used all inputs for segments  $n$ ,  $n-1$ , and  $n-2$ , reporting  $R^2$  of 0.56 and MAPE of 6.7%. The model performance was further evaluated in a design consistency evaluation exercise.

Outputs of this investigation have highlighted the potential of using ANN in enhancing the prediction power of operating speed models. The ability of ANN to consolidate inputs from more than one segment, widening the spatial extent of influence, was proven to effectively boost the model prediction accuracy. Moreover, the capacity to include all types of segments (curved and tangents) in one model adds an advantage related to the ease of application. Furthermore, the study assessed the potential transferability of the developed model by applying it to a new elevated urban arterial. Promising results were reported, with an average prediction error of 4.5%. However, it is important to acknowledge that the developed models serve as proof-of-concept models as they are based on data consolidated from two elevated urban corridors. For model generalization, data from multiple elevated urban corridors should be considered.

The models can be valuable to road authorities during the design phase to check for any design inconsistencies that could be avoided. It is not common to test the operating speed profiles of drivers during the design phase, especially on high-speed urban arterials.

While this study attempted addressing specific research objectives related to developing operating speed models for urban arterial roads using ANN, it has triggered several research questions. Future research in this direction could contribute further to this point including the following:

1. Extending the used database to develop the operating speed model to incorporate more variability in road geometry (specifically curve radii, lane widths, speed limits).
2. Incorporating other variables that may impact operating speeds such as pavement condition.
3. Training and testing the dataset under other conditions such as lighting conditions and access points on the road segments.
4. Correlating the operating speed behavior of drivers with crash occurrences on the road segments.
5. Incorporating time-dependent variables such as traffic conditions for a wider scope of research study besides free-flow conditions.
6. Using the results of this study to provide recommendations to road authorities on suggested speed limits, suggested road signage requirements, and suggested traffic management procedures to guide driver's speed choice to be suitable to the alignment driven.

#### Abbreviations

ANN	Artificial neural networks
MLR	Multiple linear regression
API	Application programming interface
$R_{max}$	Maximum curve radius
$R_{min}$	Minimum curve radius
NL	Number of lanes
PS	Posted speed
CL	Curve length
R	Curve radius
$V_{85T}$	Speed on tangent
TL	Tangent length
$V_{85C(n-1)}$	Operating speed on previous curve (n-1)
$V_{85C(n)}$	Operating speed on curve (n)
$V_{85(n)}$	Operating speed on segment n
$L_n$	Length of segment n
$C_n$	Category of segment n
PS	Posted speed
$L_{(n-1)}$	Length of segment n-1
$V_{85(n-1)}$	Operating speed on segment n-1
$C_{(n-1)}$	Category of segment n-1
$L_{(n-2)}$	Length of segment n-2
$V_{85(n-2)}$	Operating speed on segment n-2
$C_{(n-2)}$	Category of segment n-2
$L_{(n-3)}$	Length of segment n-3
$V_{85(n-3)}$	Operating speed on segment n-3
$C_{(n-3)}$	Category of segment n-3
$L_{(n-4)}$	Length of segment n-4
$V_{85(n-4)}$	Operating speed on segment n-4
$C_{(n-4)}$	Category of segment n-4
ANN M 1-n	Artificial neural network model # 1 trial # n
ANN M 2-n	Artificial neural network model # 2 trial # n
G	Good
F	Fair
P	Poor

#### Acknowledgements

Not applicable.

#### Authors' contributions

All the authors confirm contribution to the paper as follows: study conception and design, DS and HT; data collection, FR; analysis and interpretation of results, FR, DS, and HT; draft manuscript preparation, FR, DS, and HT; and manuscript review, DS and HT. All authors reviewed the results and approved the final version of the manuscript.

#### Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Availability of data and materials

All the materials, including and not limited to the descriptive analysis, tables, figures, statistical models, and equations, are included in the manuscript. In addition to that, all the relevant raw data, Excel sheets, data collected, and SPSS files are freely available to any researchers who wish to use them for noncommercial purposes while preserving data collected confidentiality and anonymity from the corresponding author on reasonable request.

### Declarations

#### Competing interests

The authors declare that they have no competing interests.

Received: 19 July 2023 Accepted: 18 September 2023

Published online: 10 October 2023

### References

1. Adu-Gyamfi Y & Zhao M. (2018). Traffic speed prediction for urban arterial roads using deep neural networks. In American Society of Civil Engineers (pp. 85–96)
2. Lamm R, Choueiri EM and Mailaender T. (1990). Comparison of operating speeds on dry and wet pavements of two-lane rural highways
3. Krammes RA, Brackett RQ, Shafer MA, Ottesen JL, Anderson IB, Fink KL & Messer CJ. (1995). Horizontal alignment design consistency for rural two-lane highways (no. FHWA-RD-94-034). Federal Highway Administration, United States
4. Voigt A (1996). An evaluation of alternative horizontal curve design approaches for rural two-lane highways
5. Gong H, Stamatiadis N (2008) Operating speed prediction models for horizontal curves on rural four-lane highways. *Transport Res Record* 2075(1):1–7. <https://doi.org/10.3141/2075-01>
6. Zuriaga AM, García AG, Torregrosa FJ, D'Attoma P (2010) Modeling operating speed and deceleration on two-lane rural roads with Global Positioning System Data. *Transport Res Record* 2171(1):11–20. <https://doi.org/10.3141/2171-02>
7. Passetti KA, Fambro DB (1999) Operating speeds on curves with and without spiral transitions. *Transport Res Rec J Transport Res Board* 1658(1):9–16. <https://doi.org/10.3141/1658-02>
8. Abbas SK, Adnan MA, Endut IR (2011) Exploration of 85th percentile operating speed model on horizontal curve: a case study for two-lane rural highways. *Procedia Soc Behav Sci* 16:352–363. <https://doi.org/10.1016/j.sbspro.2011.04.456>
9. Mahmoud H, Said D & Radwan L. (2015). Three-dimensional modelling of operating speeds on horizontal curves for two-lane rural highways. In Proceedings of 94th Transportation Research Board Annual Meeting
10. Islam M, Seneviratne P (1994) Evaluation of design consistency of two-lane rural highways. *Institute Transport Eng J* 64(2):28–31
11. Bird RN, Hashim IH (2005) Operating speed and geometry relationships for rural single carriageways in the UK. In 3rd International Symposium on Highway Geometric Design. Transportation Research Board, Chicago
12. Hashim IH, Abdel-Wahed TA, Moustafa Y (2016) Toward an operating speed profile model for rural two-lane roads in Egypt. *J Traffic Transport Eng (English edition)* 3(1):82–88
13. Fitzpatrick K, Shamburger CB, Krammes RA, Fambro DB (1997) Operating speed on suburban arterial curves. *Transp Res Rec* 1579(1):89–96
14. Fitzpatrick K, Carlson P, Brewer M, Wooldridge MD (2003) Design speed, operating speed, and posted speed limit practices. In 82nd annual meeting of the transportation research board, Washington, DC
15. Wang J, Dixon KK, Li H, Hunter M (2006) Operating-speed model for low-speed urban tangent streets based on in-vehicle global positioning system data. *Transp Res Rec* 1961(1):24–33
16. McFadden J, Yang WT, Durrans S (2001) Application of artificial neural networks to predict speeds on two-lane rural highways. *Transp Res Rec* 1751(1):9–17
17. Semeida AM (2013) Impact of highway geometry and posted speed on operating speed at multi-lane highways in Egypt. *J Adv Res* 4(6):515–523
18. Semeida AM (2014) Application of artificial neural networks for operating speed prediction at horizontal curves: a case study in Egypt. *J Modern Transport* 22(1):20–29
19. Alsobky A, Mousa R (2020) Estimating free flow speed using Google Maps API: accuracy, limitations, and applications. *Advances in Transportation Studies* 50:49–64
20. Lamm R, Psarianos B & Mailaender T. (1999). Highway design and traffic safety engineering handbook

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.