The University of Tehran Press

Pollution

Print ISSN: 2383-451X
Online ISSN: 2383-4501

https://jpoll.ut.ac.ir/

Predictive Modeling of Traffic-Related Air Pollution in Urban Areas: Insights from Road Mid-Block Sections

Rama Kanth Angatha[1] | Sai Sahitya Kurre[2]✉

1. Department of Civil Engineering, Marri Laxman Reddy Institute of Technology and Management, Hyderabad-500043, India
2.Department of Civil Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India

Article Info

Article type:
Research Article

Article history:
Received: 26 May 2025
Revised: 29 June 2025
Accepted: 8 September 2025

Keywords:
Pollutant Concentrations
Models
Traffic Flow
Road Sections
Modelling
Mid-Block Sections

ABSTRACT

Roads and highways are vital to a nation's economic and social development. However, the surge in transportation demand has led to heightened vehicular emissions, particularly at urban mid-block sections. This study presents predictive models for estimating pollutant concentrations of CO, HCHO, TVOCs, $PM_{2.5}$, and PM_{10} a Green House Gas (GHG) CO_2 based on traffic flow and vehicle type data collected across 18 urban mid-blocks in Warangal, Tirupati, and Vijayawada. Three models such as MLR, SVR, and ANN were developed, with ANN achieving the highest performance ($R^2 > 0.90$ for all pollutants). Peak concentrations of CO (1180 ppm) and $PM_{2.5}$ (over 100 µg/m³) were observed during evening hours (7–8 PM), coinciding with traffic volumes exceeding 4000 PCU/hr. A strong correlation ($R^2 > 0.7$) between traffic volume and pollutant levels was confirmed across all models. These findings provide actionable insights for urban transport planners to forecast and mitigate traffic-related air pollution at mid-block sections in similar urban environments.

Cite this article: Kanth Angatha, R., & Sahitya Kurre, S. (2025). Predictive Modeling of Traffic-Related Air Pollution in Urban Areas: Insights from Road Mid-Block Sections. *Pollution*. 11(4), 1473-1488.
https://doi.org/10.22059/poll.2025.396065.2951

DOI: https://doi.org/10.22059/poll.2025.396065.2951

INTRODUCTION

Transportation enriches civilization by facilitating connectivity and economic development; however, it significantly impacts environmental quality. The transportation sector is one of the primary contributors to increasing pollution levels. Motor vehicles release several atmospheric pollutants including carbon monoxide (CO), carbon dioxide (CO_2), oxides of nitrogen (NO_x) such as nitric oxide (NO) and nitrogen dioxide (NO_2), particulate matter ($PM_{2.5}$ & PM_{10}), and volatile organic compounds (VOCs), notably hydrocarbons (HCs) like benzene (Barth et al., 2000).

In Buenos Aires, vehicular traffic was identified as the main source of CO and NO_x emissions, with ozone formation linked to the mixing of clean air with high-NO emissions (Bogo et al., 1999). In China, the MOBILE 5 model was employed to estimate emission factors for CO, HC, and NO_x. Despite lower vehicle counts, metropolitan areas exceeded national air quality

*Corresponding Author Email: *saisahitya_k@vnrvjiet.in*

standards (Fu et al., 2001). In Los Angeles, it was found that CO and black carbon from heavy-duty diesel vehicles declined sharply with increased distance from highways (Zhu et al., 2002).

The CAL3QHC model was used to predict Carbon Monoxide (CO) concentrations at signalized intersections in a Thai city, utilizing site, traffic, meteorological, and emission parameters as input data (Tippichai et al., 2005). Delhi reported that approximately 72% of its air pollution was attributed to the transport sector (Goyal et al., 2005). In Hong Kong, traffic emissions of CO, NO_x, and PM_{10} were modeled based on traffic flow data (Xia & Shao, 2004). A study conducted in Calicut demonstrated that simple linear regression models outperformed others, such as CALINE4 and IITLS, when comparing CO levels to national standards (Anjaneyulu et al., 2008). Additionally, vehicles were confirmed as major VOC sources both inside and outside road tunnels (Ho et al., 2009). The influence of synchronized traffic signals on roadside pollution concentrations was analyzed in a Southern Italian urban area using the DRACULA microsimulation model, with neural networks modeling CO and C6H6 levels based on variable signal cycles and offset times (Zito, 2009).

The relationship between traffic flow, vehicle characteristics, and road features with vehicular exhaust emissions near intersections was examined using traffic and emission models (Pandian et al., 2009).

The COPERT III model applied in Ghana revealed that conventional passenger vehicles were significant contributors to emissions (Bonsu et al., 2010). Principal Component Analysis (PCA) was used to correlate fuel types of heavy-duty vehicles with emissions of formaldehyde and acetaldehyde (Rodrigues et al., 2011). Portable instruments facilitated the development of stable modal emission rates under varying driving conditions (Christopher et al., 2012). CO concentrations at a signalized intersection in a Malaysian city were predicted using the CAL3QHC dispersion model, showing no increase in CO levels from 2006 to 2014 despite a rise in vehicle numbers, and remained within national air quality standards (Wee and Ling, 2014). In Delhi, it was determined that private vehicles contributed 23–44% of the city's formaldehyde emissions (Nagpure et al., 2016; Masood et al., 2017).

CO emissions were spatially mapped in New Delhi using CALINE4 integrated with ArcGIS to identify high-emission zones (Masood et al., 2017). In Nigeria, emissions from diesel generators were predicted using multiple linear regression techniques (Akinyemi et al., 2018). Elevated road gradients and traffic congestion were found to significantly increase pollutant concentrations in urban settings (Abusalem et al., 2019). In Seoul, a spatiotemporal deep learning model was developed to predict citywide air pollution, incorporating traffic volume and average driving speed data, demonstrating improved forecasting accuracy (Le et al., 2019). VOC emissions from various types of vehicles were assessed using a Portable Emission Measurement System (PEMS), revealing that intersection road conditions significantly impact tailpipe emissions (Wang et al., 2020). A hybrid CNN-LSTM framework, Deep-AIR, was proposed for air quality modeling in metropolitan cities, accounting for urban dynamics such as road density and building height (Han et al., 2021).

In Wenling City, the life cycle assessment of public bicycle systems revealed that carbon balance is achieved within seven months and long-term use significantly reduces emissions (Guangnian et al., 2022). Emission reductions through cooperative urban delivery systems were achieved via optimized routing strategies (Du et al., 2022). A spatiotemporal stratified approach was employed to investigate the interrelation between urban form, traffic volume, and air quality, revealing that highly aggregated roads and industrial areas are more associated with traffic volume in polluted zones (Tian & Yao, 2022). Recent studies developed real-time machine learning models to monitor vehicular emissions in urban areas (Wang et al., 2023). An integrated simulation platform was introduced to quantify traffic-induced environmental and health impacts, combining traffic modeling, emissions modeling, dispersion modeling, and human exposure assessment (Zhao et al., 2023). The impact of electric vehicle adoption on reducing air pollutants was also evaluated (Chen et al., 2024). In Dublin, regression models utilizing Google Project Air View data and traffic data indicated that Gaussian Process Regression outperformed other models in predicting air quality, emphasizing the importance of considering spatial variability (Tafidis et al., 2024). Furthermore, big-data analytics have been used to optimize traffic flow and mitigate pollution in metropolitan areas (Kumar et al., 2025). The role of NO_2 as a reliable indicator of traffic-related air pollution was confirmed in a coastal city study, reinforcing its value for targeted mitigation policies (Hadine et al., 2025).

The emissions from vehicles at the surface would have the biggest impact on people in general. Furthermore, automobiles contribute substantially to the overall level of air pollution in several urban areas. The research lacks in analysing the impact of traffic volume and vehicle composition on air quality. The models which provide the relationship between the pollutants and traffic volume are necessary to observe the change in different pollutants with respect to increase in traffic volume. In this context, the present study attempts to measure and assess the concentration of various pollutants with respect to traffic volume observed at sections of multi-lane divided road mid-blocks. The study also focuses to develop models for predicting concentration of major air pollutants on multi-lane divided roadway mid-blocks under mixed traffic conditions.

MATERIALS AND METHODS

A detailed methodology is proposed to study the impact of traffic flow on air quality in urban areas in various phases. The first phase of the study involves a critical literature review to identify the objectives of the study. The second phase of the study deals with the selection of study areas such as urban road mid-block sections and signalized intersections in selected cities. The third phase of the study involves collection of data on sections of multilane divided road ways, traffic volume and vehicle composition data at urban road mid-blocks. The fourth phase of the study involves the detailed pollution data analysis, volume data analysis. In fifth phase of the study, development of models to predict the concentration of major pollutants on multi-lane divided urban road sections. A procedure followed to carry out the present study is presented in Fig 1.

The road mid-block sections with flat terrain and straight alignment are considered for the

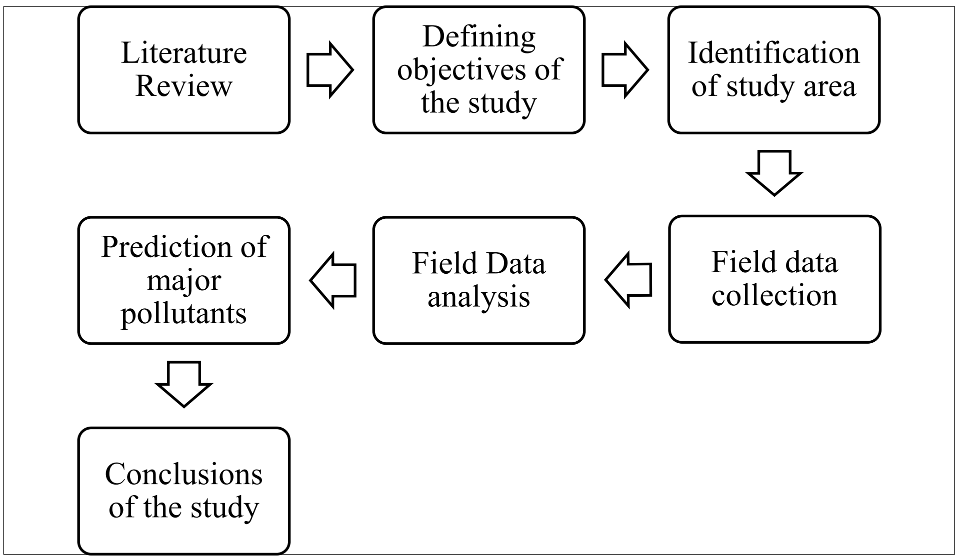

Fig. 1. Methodology flow chart of the study

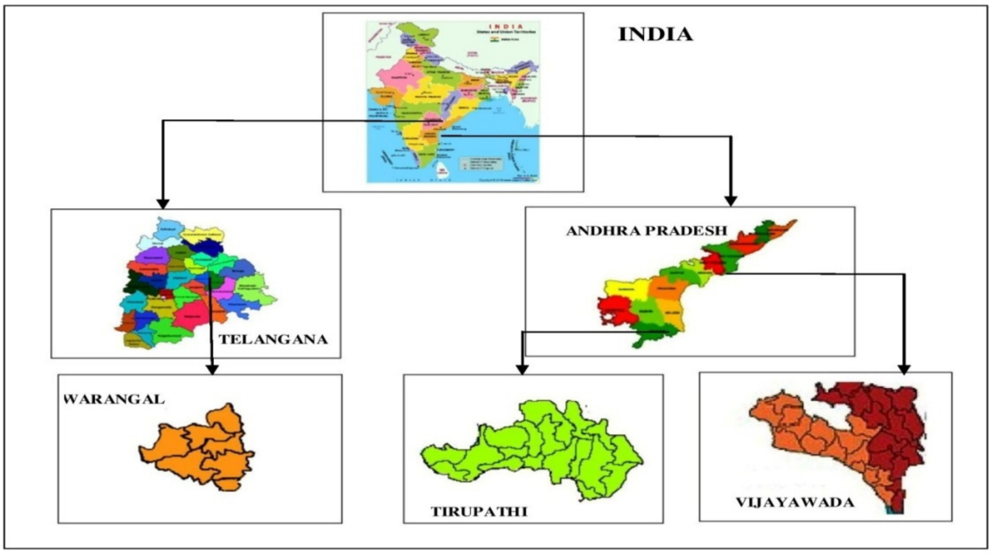

Fig. 2. Region of the cities chosen for the study

current study. Field data is collected in three cities in India namely Warangal, Vijayawada and Tirupathi. Six different mid-block sections have been identified for data collection from each city chosen. The region or map of the cities chosen for the study is shown in Fig 2.

The present study required to collect different types of data such as traffic volume and composition data at road mid-block sections and pollutant data at mid-blocks. The traffic volume (V) data was collected for eight hours in each mid-block section using video cameras. The classified traffic volume data was extracted from the video by playing the videos on wide screen display. Six major pollutants namely CO, CO_2, HCHO, TVOC, $PM_{2.5}$ and PM_{10} are observed at road mid-block sections.The concentration of pollutants was measured using handheld equipment. The data obtained for all the pollutant concentrations at every 5 min and aggregated for every 1 hour to measure average and total concentration. The present study used different portable equipment such as CO meter, laser particle multi-functional detector and air quality detector.

Traffic data were collected via videography at mid-block sections and converted to PCUs using Indo-HCM (2017) values. Section XIV recorded the highest traffic volume (5586 PCU/hr) between 7–8 PM, while Section IX had the lowest (121 PCU/hr) between 7–8 AM. Pollution data for six major pollutants were gathered every 5 minutes using handheld devices, with hourly aggregation. The highest pollution level (1180 ppm) aligned with peak traffic in Section XIV, whereas the lowest (563 ppm) occurred in Section XIII during low traffic. A strong correlation between traffic volume and pollution levels was observed.

RESULTS AND DISCUSSIONS

The relationship between different variables can be analyzed using various modelling techniques. The present study implemented three different modelling techniques namely Multiple Linear Regression (MLR), Support Vector Regression (SVR) and Artificial Neural Networks (ANN) to predict the concentration of different pollutants on road mid-block sections. The concentration of pollutants is taken as dependent variable and proportional share of different type of vehicles such as P_{2W}, P_{Car}, P_{3W}, P_{LCV} and P_{HV}, traffic volume (V) and temperature (T) are considered as independent variables for all the models.

Models were developed using MLR analysis on data from 12 road mid-block sections across Warangal, Tirupathi, and Vijayawada to predict concentrations of six pollutants (CO, CO_2, HCHO, TVOC, $PM_{2.5}$, and PM_{10}). The models used pollutant concentrations as dependent variables and factors like vehicle proportions (P2W, PCar, P3W, PLCV, PHV), traffic volume, and temperature as independent variables, achieving an R^2 of about 0.6, with the percentage of 3W, percentage of LCV, and traffic volume emerging as key influencers. The regression outputs obtained from MLR model are presented in Table 1.

$$CO = 0.49 + 137.31 \times P_{LCV} + 0.002 \times V \tag{6.1}$$

$$CO_2 = 418.45 + 1437.03 \times P_{3W} - 4.80 \times T + 0.04 \times V \tag{6.2}$$

$$HCHO = 0.01 + 0.36 \times P_{3W} - 0.02 \times T + 0.0007 \times V \tag{6.3}$$

$$TVOC = 0.03 + 0.55 \times P_{3W} - 0.005 \times T + 1.71 \times P_{LCV} \tag{6.4}$$

$$PM_{2.5} = 7.38 + 166.32 \times P_{HV} + 0.01 \times V \tag{6.5}$$

$$PM_{10} = 28.54 + 183.17 \times P_{3W} - 1.12 \times T + 0.007 \times V \tag{6.6}$$

This regression output table presents the statistical relationship between pollutant concentrations and various traffic-related factors at urban mid-block sections. For carbon monoxide (CO), the model achieved an R^2 of 0.68, indicating a good fit, with the percentage of light commercial vehicles (PLCV) and overall traffic volume emerging as highly significant contributors to CO levels ($p < 0.001$). Carbon dioxide (CO_2) showed a slightly better fit ($R^2 = 0.70$), with strong positive associations with the percentage of three-wheelers (P3W) and traffic volume, while temperature had a slight negative effect. Formaldehyde (HCHO) concentrations were moderately explained ($R^2 = 0.61$), where P3W and traffic volume had positive impacts, and temperature showed a significant negative influence. The model for total volatile organic compounds (TVOC) had the highest explanatory power ($R^2 = 0.80$), driven significantly by P3W, PLCV, and temperature (which had a negative effect). For particulate matter, $PM_{2.5}$ ($R^2 = 0.74$) was influenced mainly by the percentage of heavy vehicles (PHV) and traffic volume, both showing strong positive effects, while PM_{10} ($R^2 = 0.66$) was significantly affected by P3W, temperature (negative), and traffic volume. Across all pollutants, traffic volume consistently emerged as a key predictor, underlining its critical role in urban air quality deterioration.

The pollutant concentration models were validated using data from 6 mid-block sections, with observed and predicted values compared against a 45° reference line. Chi-square tests yielded p-values above 0.05, indicating no significant difference between observed and modeled values, as shown in the validation plots in Fig 3.

The validation plots presented in subplots (a) through (f) compare the observed and predicted pollutant concentrations for CO, CO_2, HCHO, TVOC, $PM_{2.5}$, and PM_{10}, respectively, using Multiple Linear Regression (MLR) models developed in this study. These plots demonstrate how accurately the models replicate real-world pollution levels observed at urban mid-block sections. A strong alignment of points along the 45° reference line in subplots (a) and (b) indicates that the MLR model performs reliably in estimating CO and CO_2, which is further supported by R^2 values of 0.68 and 0.70, respectively (Table 2). Subplot (c), showing HCHO, and subplot (d), showing TVOC, also reflect acceptable model performance, with data clustering near the line and respective R^2 values of 0.61 and 0.80. These results are in line with studies such as Pandian et al. (2009) and Xia & Shao (2004), who also demonstrated strong pollutant-traffic volume correlations using statistical models.

Table 1. Regression outputs of the MLR models developed at road mid-block sections

Dependent variable	Independent variables	p-value	t-stat value	R^2	Adjusted R^2
CO (ppm)	Intercept	0.004	2.92	0.68	0.67
	P_{LCV}	0.000	11.68		
	Traffic volume	0.000	9.41		
CO_2 (ppm)	Intercept	0.000	8.10	0.70	0.71
	P_{3W}	0.000	9.52		
	Temperature	0.020	-2.35		
	Traffic volume	0.000	6.34		
HCHO (mg/m^3)	Intercept	0.001	2.15	0.61	0.61
	P_{3W}	0.000	6.00		
	Temperature	0.003	-3.04		
	Traffic volume	0.006	2.76		
TVOC (mg/m^3)	Intercept	0.041	1.98	0.80	0.79
	P_{3W}	0.000	3.58		
	Temperature	0.000	-4.07		
	P_{LCV}	0.022	2.32		
$PM_{2.5}$ (µg/m^3)	Intercept	0.000	6.11	0.74	0.73
	P_{HV}	0.000	4.81		
	Traffic volume	0.000	16.13		
PM_{10} (µg/m^3)	Intercept	0.033	2.16	0.66	0.65
	P_{3W}	0.000	7.21		
	Temperature	0.005	-2.85		
	Traffic volume	0.000	7.03		

Subplots (e) and (f), which validate $PM_{2.5}$ and PM_{10} predictions, show slightly more dispersion but still maintain a clear upward trend, with R^2 values of 0.74 and 0.66, respectively, suggesting that the model effectively captures the particulate pollution dynamics influenced by vehicle types and volumes. The p-values from the Chi-square tests are above 0.05 (e.g., 0.91 for CO_2), indicating no statistically significant difference between observed and predicted values, validating the models' robustness. Similar modeling approaches and validation techniques were employed in previous works such as Goyal et al. (2005) and Tippichai et al. (2005), which also found traffic flow to be a critical determinant of urban air pollution levels. Overall, the

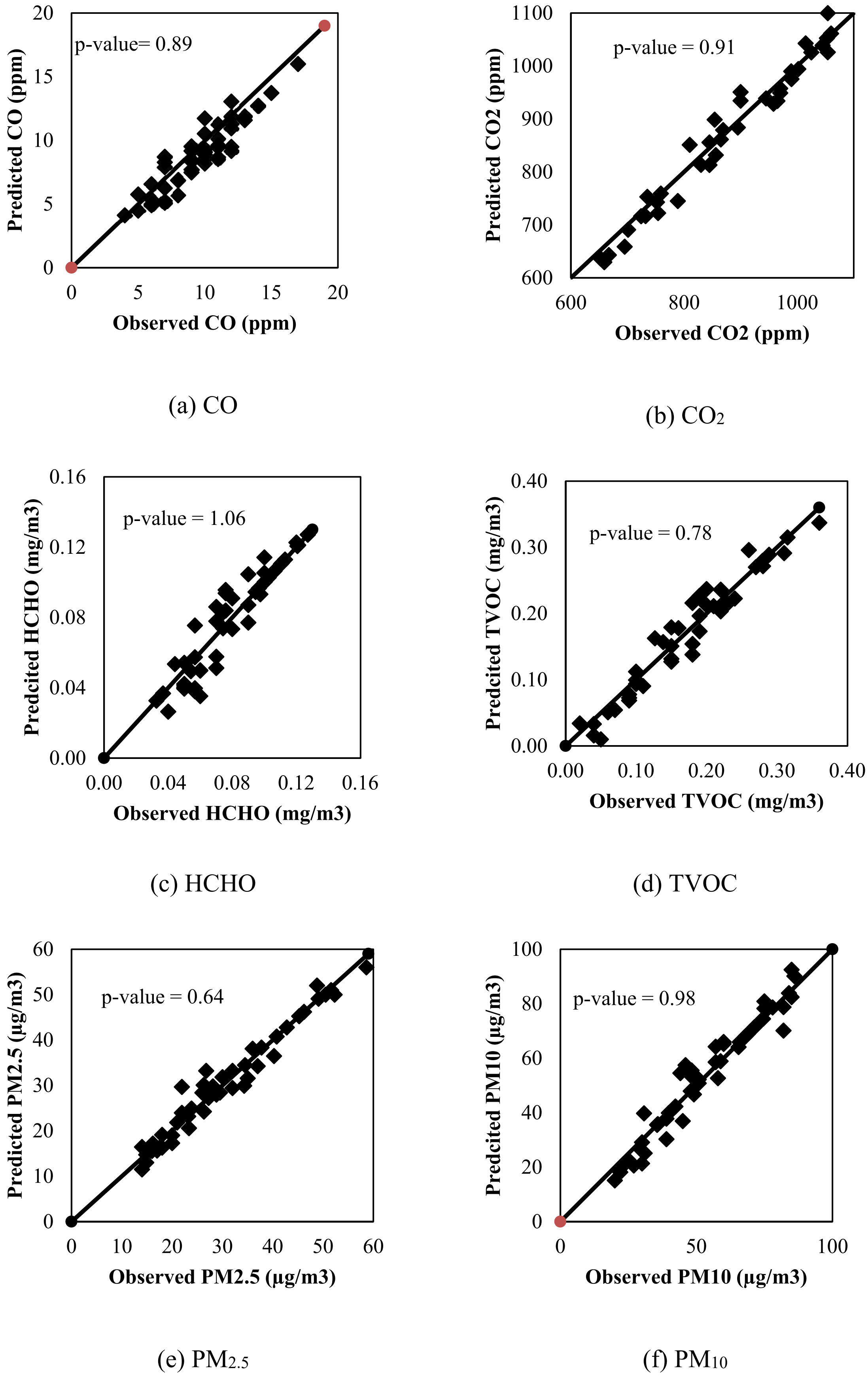

Fig. 3. Validation plots for different pollutant concentrations

Table 2. Regression outputs of the SVR models developed at road mid-block sections

Dependent variable	Independent variables	Co-efficient	p-value	t-stat value	R^2	Adjusted R^2
CO (ppm)	Intercept	1.75	0.002	3.12		
	P_{LCV}	143.9	0.000	11.33	0.72	0.71
	Traffic volume	0.001	0.000	9.49		
CO_2 (ppm)	Intercept	425.45	0.000	7.21		
	P_{3W}	762.34	0.000	3.96		
	Temperature	-3.81	0.015	-2.22	0.75	0.75
	Traffic volume	0.06	0.000	8.19		
HCHO (mg/m^3)	Intercept	0.004	0.001	2.24		
	P_{3W}	0.028	0.000	4.42		
	Temperature	-0.002	0.004	-2.89	0.65	0.66
	Traffic volume	0.0006	0.003	2.92		
TVOC (mg/m^3)	Intercept	0.001	0.041	2.01		
	P_{3W}	0.42	0.021	2.92		
	Temperature	-0.003	0.042	-2.00	0.85	0.86
	P_{LCV}	1.40	0.022	2.12		
$PM_{2.5}$ (µg/m^3)	Intercept	5.42	0.000	5.37		
	P_{HV}	92.69	0.000	2.49	0.81	0.80
	Traffic volume	0.02	0.000	14.29		
PM_{10} (µg/m^3)	Intercept	24.97	0.033	2.60		
	P_{3W}	135.49	0.000	4.45		
	Temperature	-0.58	0.005	-2.15	0.70	0.71
	Traffic volume	0.007	0.000	5.99		

validation plots confirm that the MLR models developed in this study are reliable for predicting traffic-induced air pollutant concentrations at urban mid-block sections.

The Support Vector Regression (SVR) modeling approach provides a robust framework for regression tasks. SVM technique is efficient in handling high-dimensional data and non-linear relationships. The present study implemented SVR in R-software to predict the concentration of different pollutants on road mid-block sections. SVR modelling considers 70% of the data for training and 30% of the data for testing. The SVM type taken for the regression is eps-regression. The SVM Kernel function used for the regression is radial. The confidence interval used for the regression is 95%. The libraries, functions and data assumptions used to develop

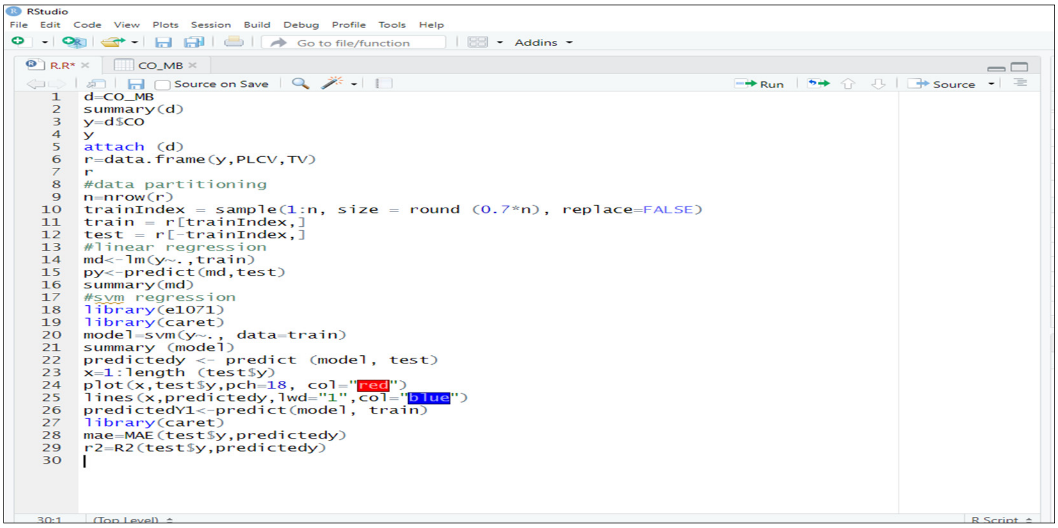

Fig. 4. Sample data input for SVR model developed for mid-block sections

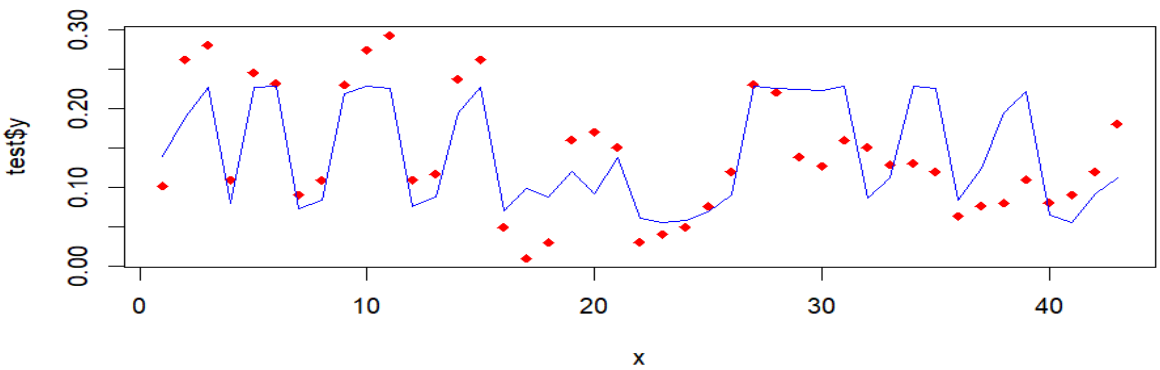

Fig. 5. SVR model predicted TVOC concentrations

the model is given as a sample in Fig 4.

This figure shows a screenshot of an RStudio script used for implementing Support Vector Regression (SVR) to predict pollutant concentrations (specifically CO in this case) based on traffic-related variables. The script first loads the dataset (d) and creates a data frame with three variables: the dependent variable y (CO concentration), and independent variables PLCV (percentage of light commercial vehicles) and TV (traffic volume). The dataset is then partitioned into training (70%) and testing (30%) sets using random sampling. A linear model is briefly created for comparison, but the main focus is on SVR, implemented using the e1071 and caret libraries. The SVR model (svm()) is trained on the training set (train) and used to predict values on the test set (test), which are visualized by plotting the observed and predicted values using a red line for actual and a blue line for predicted results. Performance metrics such as Mean Absolute Error (MAE) and Coefficient of Determination (R^2) are calculated using the caret package to assess model accuracy. This workflow is part of the model validation process described in the study to evaluate the effectiveness of SVR in forecasting traffic-induced pollution at mid-block road sections.

The proportional share of different type of vehicles such as P_{2W}, P_{Car}, P_{3W}, P_{LCV} and P_{HV}, traffic volume (V) and temperature (T) are considered explanatory variables and concentration of pollutants as dependent variables for all the models. The statistical outputs of support vector

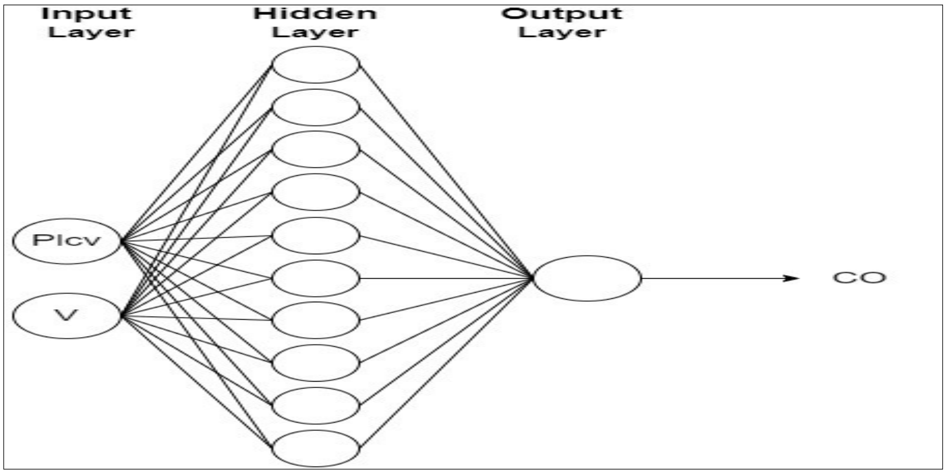

Fig. 5. Neural network architecture for CO model

regression models developed in the study are given in Table 2.

This table presents the results of the Support Vector Regression (SVR) models developed to predict the concentration of various air pollutants based on traffic and environmental factors at urban road mid-block sections. The models show strong predictive ability, with R^2 values ranging from 0.65 (HCHO) to 0.85 (TVOC), indicating a high degree of accuracy in capturing the variance in pollutant levels. For CO, the most influential predictors are the percentage of light commercial vehicles (PLCV) and traffic volume, both positively contributing to CO levels. In the case of CO_2, the model identifies a strong positive relationship with three-wheeler share (P3W) and traffic volume, while temperature shows a negative effect. Similarly, HCHO levels increase with P3W and traffic volume but decrease with rising temperatures. The TVOC model, which achieves the highest R^2 (0.85), shows significant influence from P3W, PLCV, and temperature, suggesting that both vehicle composition and environmental conditions play key roles. For $PM_{2.5}$, the concentration is driven mainly by the percentage of heavy vehicles (PHV) and traffic volume, with an exceptionally strong association indicated by a t-value of 14.29 for traffic volume. Finally, PM_{10} levels are significantly affected by P3W, temperature (negatively), and traffic volume. Across all models, traffic volume consistently emerges as a dominant predictor, reinforcing its critical impact on urban air quality. The SVR models thus offer a robust framework for accurately forecasting pollution levels using real-world traffic data.

The SVM model plots display predicted pollutant concentrations with test samples on the X-axis and predicted values on the Y-axis, where each point is generated by the SVR model. A trend line illustrates that predictions closely follow the expected pattern, as exemplified by the TVOC plot in Fig 5.

This figure displays the Support Vector Regression (SVR) model validation plot for one of the pollutants—most likely TVOC (Total Volatile Organic Compounds)—based on test data. The red diamond points represent the actual observed values of the pollutant concentration (test$y), while the blue line shows the SVR-predicted values (predictedy) over the same data points indexed by x (the test sample index). Ideally, close alignment between the red points and the blue line indicates good predictive performance. In this plot, the SVR model captures the general trend of the pollutant variation, but with some fluctuations and deviations, especially at certain peak and trough points. This pattern suggests that while the model performs reasonably well in predicting pollutant concentrations, there are still some errors due to the complexity and variability of the underlying traffic-pollution dynamics. The model's effectiveness would

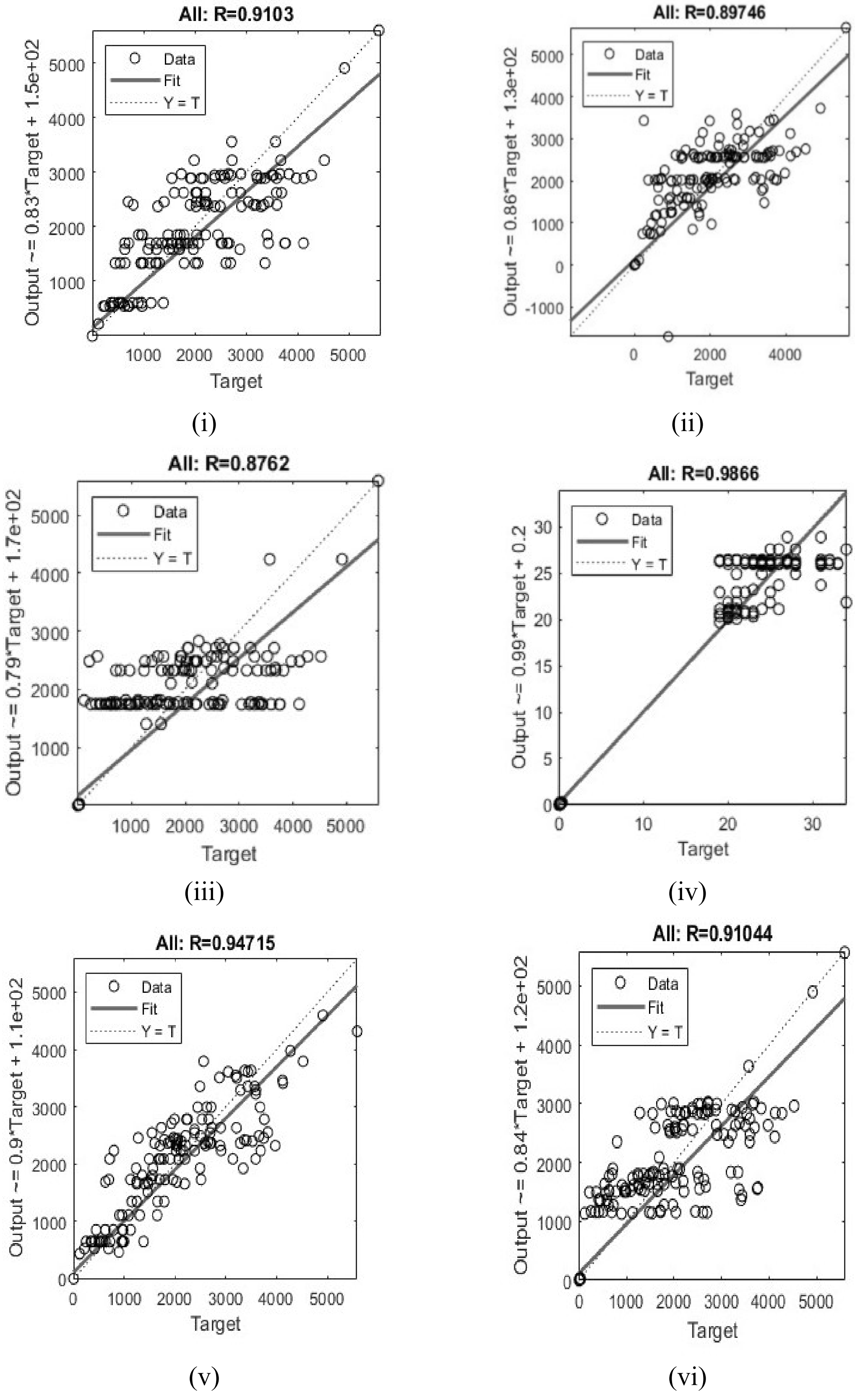

Fig. 6. Output of ANN models (i) CO, (ii) CO_2, (iii) HCHO, (iv) TVOC, (v)$PM_{2.5}$, (vi) PM_{10}

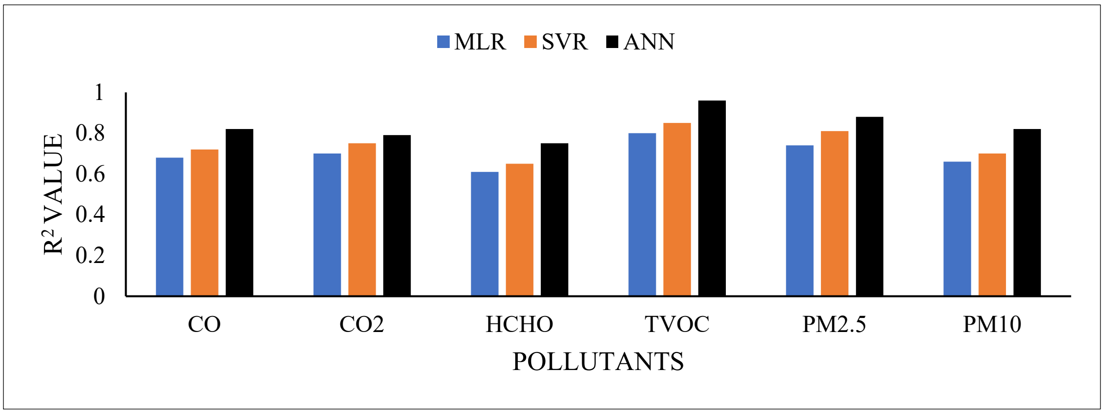

Fig. 6. Comparison of R^2 values for different models

typically be supported by statistical metrics such as R^2 and MAE, calculated in the associated R script, confirming its suitability for urban air quality modeling at mid-block road sections.

The present study also used the ANN technique to predict concentration of pollutants based on proportional share of different type of vehicles. Multi-layered feed-forward ANN is used by considering concentration of pollutants as the output layer in this study, and proportional share of different type of vehicles such as P_{2W}, P_{Car}, P_{3W}, P_{LCV} and P_{HV}, traffic volume (V) and temperature (T) as input layers.

Different iterations were carried out by changing the number of hidden layers and neurons to obtain the optimal network structure. The network structure with less error between the measured and ANN modelled values is the optimal structure. The neural network structure (one input layer, one hidden layer with ten neurons and one output layer) obtained for CO model is shown as a sample in Fig 5. Similar neural network structures are obtained for all the pollutants. The output of ANN models is given in Fig 6.

The ANN models developed in present study represents the graphs between the observed pollutant values, and ANN predicted pollutant values. The horizontal axis values display the field observed values. The vertical axis values are the forecasted values produced by the neural network model. The R-value represents the relation between the observed and ANN predicted values.

A comparison has been made between the developed models of MLR, SVM, and ANN to understand the accuracy of the prediction of concentration of pollutants corresponding to the traffic volume. The error between the measured and modelled values is quantified to examine different methods' performance in predicting concentration of pollutants. The statistical estimates such as R^2value and Mean Absolute Percentage Error (MAPE) are used for measuring accuracy and determining the error. The MAPE value is the error percentage between the observed and predicted values obtained for each model. The model with high R^2value and low MAPE values gives the best-predicted values. The R^2 values and MAPE values obtained for the models are shown in Fig 6 and 7 respectively.

The results of comparison indicate that high R^2value and low MAPE values are obtained when the ANN method is used to predict concentration of pollutants on road mid-block sections compared to the other methods. It represents that the ANN method could better predict the concentration of pollutants values regarding the rather than MLR and SVR methods.

CONCLUSİONS

The findings of this study reveal a clear statistical relationship between pollutant concentrations and traffic volume at urban road mid-block sections. Through the application of Multiple

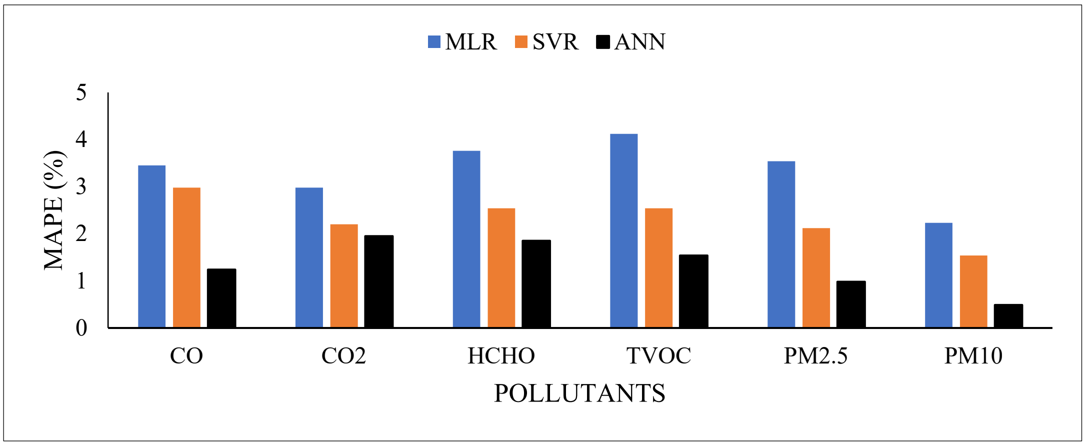

Fig. 7. Comparison of MAPE values for different models

Linear Regression (MLR) analysis, models were developed to estimate the concentrations of key pollutants, including CO, CO_2, HCHO, TVOCs, $PM_{2.5}$, and PM_{10}, with traffic volume and vehicle percentages as explanatory variables. Validation using Chi-square tests demonstrated that the predicted and observed values were highly consistent, confirming the models' reliability.

The analysis showed significant diurnal variations in pollutant concentrations, particularly between 6 PM and 9 PM, when traffic volumes ranged from 2500 to 4000 PCU/hr. During this period, pollutants such as CO, HCHO, TVOCs, and particulate matter (PM) exhibited substantial increases, whereas CO_2 concentrations remained largely stable. The total pollution concentration at mid-block road sections was found to increase by approximately 50% from morning to evening, correlating with a 73% rise in traffic volume. The highest pollutant concentration, recorded at 1180 ppm during the peak hour (7 PM to 8 PM), coincided with a traffic flow of 4000 PCU/h.

Furthermore, the study demonstrated that Artificial Neural Networks (ANN) outperformed both MLR and Support Vector Regression (SVR) models in predicting pollutant concentrations, as evidenced by superior R^2 values. These results underscore the critical role of increasing traffic volume in deteriorating air quality, with further traffic growth posing significant environmental risks.

This study provides valuable empirical insights into the intensity and dynamics of pollutant emissions from road traffic. It highlights the need for urban transportation planners and policymakers to address rising traffic volumes, as unchecked growth may lead to increasingly severe air quality degradation. The models developed offer a practical tool for forecasting pollution levels, enabling more informed decision-making to mitigate traffic-induced environmental impacts.

The results of this study, while derived from a specific urban context, have broad applicability to other cities experiencing similar traffic growth and air quality challenges. The modeling approaches using MLR, SVR, and ANN can be adapted to different urban environments with varying traffic compositions and infrastructure characteristics, offering a scalable method for estimating pollutant concentrations based on traffic patterns. The clear correlation between traffic volume and pollutant levels underscores a universal trend in urban areas, making the findings relevant for transportation planners and environmental regulators globally. These models can serve as predictive tools for assessing the environmental impact of proposed traffic management strategies, infrastructure developments, or policy interventions, thereby supporting data-driven decisions for sustainable urban mobility and improved public health outcomes.

Based on the findings of this study, several practical recommendations are proposed to mitigate traffic-related air pollution at urban road mid-block sections. Authorities should prioritize the deployment of real-time air quality monitoring systems integrated with Artificial Neural Network (ANN) models, which demonstrated superior predictive accuracy in this study. Traffic management strategies such as dynamic signal timing, vehicle restrictions, and staggered work hours should be implemented during evening peak periods (6 PM to 9 PM), when pollution levels were observed to rise by approximately 50%. Policymakers are encouraged to regulate the movement of high-emission vehicle types, particularly three-wheelers and light commercial vehicles, in sensitive zones. Promoting the use of public transport, non-motorized mobility, and electric vehicles can further reduce pollutant concentrations. Lastly, urban planners should incorporate emission modeling tools into infrastructure design and land-use policies, ensuring that mid-block sections are regularly assessed and improved for environmental resilience.

ABBREVATIONS

CO (Carbon Monoxide), CO2 (Carbon Dioxide), HCHO (Formaldehyde), TVOCS (Total Volatile Compounds), NO_x (Nitrogen Oxides), NO (nitrogen Oxide), NO2 (Nitorgen Dioxide), VOCs (volatile Organic Compounds), HCs (Hydro Carbons), HC (Hydro Carbon), CAL3QHC (Dıspersıon model of USA), CALINE4 (California Line Source 4 Model), COPERT III (EU standard vehicle emissions calculator), PCA (Principal Component Analysis), PEMS (Portable Emission Monitoring System), CNN-LSTM (Convalutional Neural Network – Long Short Term Memory Networks), V (Traffic Volume), PCUs (passenger Car Units), MLR (Multiple Linear Regression), SVR (Support Vector Regression), ANN (Artificial Neural Networks), P2W (Proportion of two wheeleres), Pcar (Proportion of cars), P3W (Proportion of three wheelers), PLCV (Proportion of Light Commercial vehicles), PHV (Proportion of Heavy vehicles), T (Temperature), LCV (Light Commerical Vehicles), PM (particluate Matter), MAPE (Mean Absolute Percentage error), R2 (Regression value), SVM (Support Vector Machine), PM2.5 (Particulate Matter 2.5) and PM10 (Particulate Matter 10) and IITLS (Indian Institute of Technology Line Source Model).

GRANT SUPPORT DETAILS

The present research did not receive any financial support.

CONFLICT OF INTEREST

The authors declare that there is not any conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/ or falsification, double publication and/or submission, and redundancy has been completely observed by the authors.

LIFE SCIENCE REPORTING

No life science threat was practiced in this research

ACKNOWLEDMENTS

The author(s) declare that there are no acknowledgements to report.

REFERENCES

Abusalem, S., Al-Ahmadi, F., & Al-Garni, A. (2019). Effects of road gradients and congestion on urban pollution. *Transportation Research Part D: Transport and Environment*, 67, 473-485.

Akinyemi, O., Oladipo, O., & Adebayo, O. (2018). Diesel generator emissions modeling in Nigeria. *Renewable and Sustainable Energy Reviews*, 82, 2224-2230.

Anjaneyulu, Y., Rao, D., & Reddy, A. (2008). Comparative analysis of CO measurement models in Calicut. *Transportation Research Part D: Transport and Environment*, 13(6), 404-411.

Barth, M., Seinfeld, J.H., & Hildemann, L.M. (2000). Motor vehicle emissions and air quality. *Atmospheric Environment*, 34(15), 2543-2554.

Bogo, H., Bermejo, V., & Gaggero, L. (1999). Vehicular sources of CO and NO_x in Buenos Aires. *Environmental Monitoring and Assessment*, 58(2), 123-134.

Bonsu, G., Kumah, D., & Adomako, D. (2010). Vehicle emission estimates using COPERT III in Ghana. *Environmental Science and Pollution Research*, 17(8), 1301-1310.

Chen, M., Huang, Y., & Zhao, J. (2024). Evaluating electric vehicle adoption impact on urban air quality. *Energy Policy*, 172, 113312.

Christopher, S.A., Anwar, S., & Wilson, J. (2012). Portable emission rates for different driving conditions. *Journal of the Air & Waste Management Association*, 62(8), 940-947.

Du, Y., Wang, Z., & Chen, L. (2022). Cooperative routing and emission reduction in urban delivery. *Transportation Research Part D: Transport and Environment*, 105, 103201.

Fu, L., Zhang, Y., & Wang, J. (2001). Estimation of vehicle emission factors in metropolitan China using MOBILE 5. *Journal of Environmental Sciences*, 13(4), 495-503.

Goyal, P., Goyal, R., & Khare, M. (2005). Air pollution in Delhi: Transport sector contribution. *Environmental Pollution*, 138(3), 521-529.

Guangnian, Z., Li, Y., & Xiong, W. (2022). Life cycle carbon assessment of public bike systems in Wenling City. *Journal of Cleaner Production*, 338, 130571.

Hadine, M., Mabrouki, J., Qisse, N., El-Moustaqim, K., El Yousfi, Y. (2025). 'Contribution of Predictive Statistics in the Evaluation of Correlations between Air Pollutants and Traffic Intensity in the City of Kénitra', *Pollution*, 11(1), 134-146. DOI: 10.22059/poll.2024.376596.2374

Han, D., Lee, J., & Kim, S. (2021). Deep-AIR: Hybrid CNN-LSTM model for urban air quality prediction. *Atmospheric Environment*, 244, 117962.

Ho, K.F., Lee, S.C., & Yu, J.Z. (2009). Vehicle emissions of VOCs inside and outside tunnels. *Atmospheric Environment*, 43(20), 3240-3246.

Kumar, A., Singh, R., & Verma, P. (2025). Big data analytics for optimizing traffic flow and pollution mitigation. *Transportation Research Part C: Emerging Technologies*, 145, 103882.

Le, T., Nguyen, H., & Park, S. (2019). Deep learning for spatiotemporal air pollution forecasting in Seoul. *Environmental Pollution*, 252, 338-348.

Masood, A., Saxena, M., & Singh, V. (2017). Spatial mapping of CO emissions using CALINE4 and ArcGIS in Delhi. *Environmental Monitoring and Assessment*, 189(8), 400.

Nagpure, A.S., Gurjar, B.R., & Kumar, S. (2016). Formaldehyde emissions from private vehicles in Delhi. *Atmospheric Environment*, 123, 144-151.

Pandian, S., Gokhale, S. and Ghoshal, A.K., (2009). Evaluating effects of traffic and vehicle characteristics on vehicular emissions near traffic intersections. *Transportation Research Part D: Transport and Environment*, 14(3), pp.180–196.

Rodrigues, S., Pereira, M., & Moreira, A. (2011). PCA of fuel types and aldehyde emissions from heavy-duty vehicles. *Environmental Science and Technology*, 45(15), 6448-6454.

Tafidis, P., O'Sullivan, D., & Brennan, R. (2024). Predicting urban air quality using Gaussian Process Regression and Google Air View data. *Atmospheric Environment*, 293, 119911.

Tian, Z., & Yao, L. (2022). Spatiotemporal analysis of urban form, traffic volume and air quality. *Science of the Total Environment*, 806, 150608.

Tippichai, A., Klungboonkrong, P., Aram, P. and Wongwises, P., (2005). Prediction of CO concentrations from road traffic at signalized intersections using CAL3QHC model: the Khon Kaen case study. *Songklanakarin Journal of Science and Technology*, 27(6), 1233–1242.

Wang, M., Lia, S., Zhu, R., Zhang, R., Zu, L., Wang, Z. and Bao, X., (2020). On-road tailpipe emission characteristics and ozone formation potentials of VOCs from gasoline, diesel and liquefied petroleum

gas fuelled vehicles. *Atmospheric Environment*, 223, 117294.
Wang, Q., Zhang, J., & Li, X. (2023). Real-time machine learning models for urban vehicular emission monitoring. *Environmental Modelling & Software*, 160, 105610.
Xia, L., & Shao, M. (2004). Modeling traffic-related emissions of CO, NO_x and PM_{10} in Hong Kong. *Science of the Total Environment*, 332(1-3), 135-144.
Zhao, H., Liu, J., & Sun, Y. (2023). Integrated simulation platform for traffic-induced environmental and health impact assessment. *Transportation Research Part D: Transport and Environment*, 115, 103690.
Zhu, Y., Hinds, W.C., Kim, S., & Sioutas, C. (2002). Concentration and size distribution of ultrafine particles near highways. *Environmental Science & Technology*, 36(6), 1044-1052.
Zito, P., (2009). Influence of coordinated traffic signals parameters on roadside pollutant concentrations. *Transportation Research Part D: Transport and Environment*, 14(9), 604–609.