



# Short-term traffic prediction model for urban transportation using structure pattern and regression: an Indian context

Sathiyaraj Rajendran<sup>1</sup> · Bharathi Ayyasamy<sup>2</sup>Received: 22 November 2019 / Accepted: 22 May 2020 / Published online: 3 June 2020  
© Springer Nature Switzerland AG 2020

## Abstract

The paper presents a model for predicting short term traffic flow using structure pattern and regression. The relationship between traffic flow on the current section and upstream stations are used for predicting the short-term traffic flow and is exposed by the traffic flow structure pattern. The structure pattern of the traffic flow can be pinched from freeway toll data based on which, a new prediction model has been proposed. Locally weighted learning is used to predict the next instance's traffic flow of current section by considering the traffic flow of current section and the next station entrance flow. This learning model deploys both linear and non-linear models to fit the nearby points and then applies those values to forecast the values of the query points. Experimental results illustrates that the short term prediction method based on structure pattern and regression is an effective approach for traffic flow prediction, this is applicable especially during the abnormal traffic states. The proposed system outperforms the conventional systems in terms of accuracy and proves to facilitate in conserving energy. The proposed work benefits the travelers by rerouting and saves fuel consumption.

**Keywords** Short-term traffic flow · Prediction · Structure pattern · Locally weighted learning · Regression · Congestion avoidance

## 1 Introduction

India is the second largest country in terms of growth in population and economy. Most of the cities in India are facing road congestion problems. There are practical difficulties in maintaining Intelligent Transport Management Systems (ITMS) in developed countries and many difficulties are facing widely in improving the metropolitan cities. This is one of the most important problems in India to be addressed. This is due to slow growth in infrastructure compared to the rapid increase in the number of vehicles, space and cost constraints. The Traffic flow information is needed for travelers to help them to make better travel decisions on congestions and to improve traffic operation

efficiency. Predicting short-term traffic flow will be more helpful in managing freeway networks. This traffic flow prediction makes use of real-time data in predicting the traffic status in the subsequent 5–20 min. All countries in the world are trying to improve their traffic management system to be more efficient. Researchers have used different methods to predict freeway traffic in urban areas. The Indian Government committed Rs. 234,000 crores for the development of urban infrastructure sector. To promote the use of public transport most of the cities offers convenient bus rapid transit, metro and monorail services. Even with the development of transportation services, traffic congestion is still the most common problem in many of the developing cities. In metropolitan cities like Bangalore

✉ Sathiyaraj Rajendran, [rsr026@gmail.com](mailto:rsr026@gmail.com); Bharathi Ayyasamy, [bharathia@bitsathy.ac.in](mailto:bharathia@bitsathy.ac.in) | <sup>1</sup>Department of Information Technology, G.G.R. College of Engineering, Anna University Chennai, Vellore, Tamil Nadu, India. <sup>2</sup>Department of Information Technology, Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu, India.



and Delhi, with a sudden growth of the IT sector, population and number of vehicles also increased. Smart and intelligent traffic information system is required to manage traffic congestion and to intimate travelers to take different routes to avoid congestion and to save time. This is the need for Intelligent Traffic System (ITS) in urban areas in India.

Green computing supports the use of resources that are efficient in computing and infrastructure support. Green computing emphasizes to minimize the harmful environmental impact and helps to achieve enhanced system performance [2, 3, 6]. Most of the nations apply information and communication technologies to support sustainable urban development. In the transportation sector, the initiatives are intended to provide a safe and secure transportation system thereby ensuring sustainability of huge cities. ITS focuses primarily on road transportation services whereas it also has wide applications in rail, air and water transport services and incorporates navigational devices. Facilitating services to urban road transportation services is the scope of this paper.

Intelligent Traffic Management System is an interdisciplinary research area. Most number of researchers from various domains has shown their interest in developing an intelligent traffic management system. Auto Regressive Integrated Moving Average (ARIMA) was used by most of the researchers in providing a solution to ITS with regression models. Traffic information can be categorized as abnormal and normal components [7–10, 13] based on its statistical short time traffic characteristics using Wavelet Transform (WT) theory. The major tendency of traffic flow is the normal component, whereas the abnormal component is the unusual traffic flow that makes the concrete traffic flow very near the common tendency. Most of the academic and research studies have focused on predicting short-term traffic which focuses on regular traffic states. To predict the traffic flow on holidays, Online-Support Vector regression was used [20]. Traffic flow congestion can be predicted based on historical and real-time traffic data gathered from several sensors, inductive loops, cameras, radars, mobile GPS, social media, etc. [22]. Data posted by users in various online social media can also be used in predicting the traffic flow and collecting data from a highway toll is uncomplicated in freeway network. The information about the vehicles entry and exit of the freeway network can be found using the freeway toll data. The origin and destination point of the vehicles can be obtained easily. The objective of the paper is to predict the traffic flow in abnormal conditions with the information collected through toll data.

The major contributions of this manuscript are (1) Provides the traffic flow structure pattern in a freeway

network, (2) An efficient short-term traffic prediction method based on big data analytics is proposed, (3) Case study of our proposed system with simulation results. The left over part of the paper is structured as follows: Section 2 provides a brief overview of the related works evolved in this area. Section 3 provides the analysis of traffic flow; this summarizes the traffic data collection process and analysis. Section 4 provides a Traffic flow based prediction method. Results are presented in Sect. 5 and discussed in Sect. 6. Finally, Sect. 7 provides the concluding remarks of the proposed work.

## 2 Related work

With the improvement in information and communication technologies, many traffic prediction models were developed to support traffic management and control. Predicting short-term traffic flow is one of the challenging tasks in abnormal traffic situations and this is the key in utilizing the collected traffic data efficiently thereby predicting the traffic flow in the subsequent phase of time (regularly 5–20 min). In 1970s, ARIMA (autoregressive integrated moving average) model was used to predict the short-term traffic flow in freeway networks [4, 23, 25]. Time-series is the major factor in almost any of the traffic flow prediction models. A linear regression model was used in forecasting short-term traffic [26]. In the same traffic flow forecasting, a Bayesian network method was employed [27] to predict traffic flow and a support vector based online learning weighted support vector regression approach were proposed [16]. Many researchers have focused on providing an efficient solution for traffic management and control [15, 17, 18, 21]. The focus was turned on machine learning techniques [30–32] which provided a path for academics and industries to focus their attention towards Smart Intelligent Transport Management System (SITMS).

A prediction methodology based on deep-learning is proposed [22] for analyzing the system. A stacked auto encoder model is employed in a greedy layer wise fashion. An approach for forecasting 2 min travel rate using fuzzy neural networks is presented [28]. The proposed mechanism was applied to the traffic data gathered from remote microwave sensors in Beijing city. The proposed approach results proved that this was an efficient mechanism and better than the traditional models. A vibrant traffic simulator was designed [1] to create flows in all relations with obtainable traffic data, required, and historical traffic information obtainable from relations set with sensors was proposed. The proposed scheme applies the simulator to alter the origin-to-destination matrices. In predicting the traffic flow on each available link up to 30 min ahead, real-time

and estimated traffic data was used. The proposed prediction algorithm is also based on an autoregressive model.

The traffic flow prediction algorithms have its own benefits and drawbacks. Few of the existing approaches can be applied on undisciplined traffic. Several prediction models of traffic flow were developed recently which can be employed in abnormal conditions. Online-Support Vector Regression [29] used in predicting traffic especially during holidays. Few researchers have used neural networks in predicting traffic during abnormal weather conditions. The drawback of the existing methods is that there is no analysis of the traffic structure between the current section and upstream section. Predicting the short-term traffic flow is the base for managing the traffic in freeway networks. This assists in dropping the mean travelling time and avoids from traffic congestion. Short-term traffic prediction is more useful in developing an advanced transportation system. This plays a major role in managing abnormal traffic conditions and especially on holidays and weekends. Predicted in short duration is more important because the traffic can rise severely from time to time in next 5–10 min and can be highly congested in the next few minutes. These sorts of prediction systems are especially applicable for Indian Traffic System.

In this paper, we propose a new short-term traffic prediction methodology. This is analyzed with a case study—the experiment is performed in one of the major metropolitan city of India. This case study uses toll data in a freeway network. The toll data contains information about the entry and exit of vehicles in the freeway network. Using the collected traffic data, source and destination can be obtained. The traffic flow pattern organization can be extracted with the assistance of few video cameras installed at a toll. With the collected information, we can reveal the relationship among the traffic stream on the present road segment and the upstream stations. This information will support in predicting and avoiding traffic congestion, especially in abnormal traffic conditions.

### 3 Analysis of traffic flow

ITS provides several solutions with big data analytics to improve the traffic scenario issues. Several ITS applications are launched to promote Indian traffic system in various aspects like Vehicle classification, Intersection control, Incident detection, historical traffic data, and monitoring. Traffic congestion is a most common problem in many of the metropolitan cities in India. Most of the cities need a smart traffic management system to predict congestions and to help avoid accidents before they occur. Since Indian traffic system is a heterogeneous traffic system, congestion is a major issue for travelers [11, 12]. Most of the cities

are experiencing a major growth in population and face many challenges as they are expanding and increasing at a larger rate. Traffic is one of the most challenging issues in all developing cities. Table 1 (Ref: World Economic Forum) provides the measures taken by the some of the admired cities in controlling the traffic.

Modern Traffic Management system is widely needed for Indian traffic system in managing traffic conditions across various developing cities. Mumbai introduced a modern traffic management system. Real-time tuning traffic flow system was adopted; traffic intensity was detected by the installed video cameras and the data was submitted to the computer. This assisted in real time adjustments of traffic signals and help travelers feel better and to generate an efficient traffic flow system across the city. It was also possible to track and assist the spots from the control center. Damaged vehicles or accidents or huge traffic jam can also be handled very quickly and monitored from the control center.

ITS and Big data analytics jointly provides solutions for traffic conditions under abnormal conditions. In this paper, we offer a prediction method which can be adopted by most of our Indian cities. Our prediction method uses the information collected from one of the toll roads connecting Bangalore. Travelers are charged for passage in a toll road. This is implemented to support in earning the expense of road building and safeguarding, it is a form of taxation. We consider one of the important IT development cities in India for prediction of traffic flow.

Bengaluru which is also known as Bangalore, is the capital of the South Indian state of Karnataka. This is the third most populated city with more than 8 million people. This city is also called as Garden city and is well-known for its pleasant climate throughout the year and attracts many travelers to the city. Bengaluru is the fastest developing major metropolitan city in India. With the sudden growth of IT Sectors, this city has become very populated, congested and faces various issues with traffic management system. It is also home to many educational and research institutions in India. Many people also often visit the city for various purposes. Managing traffic and

**Table 1** Traffic measures

S. no	Name of the city	Measure
1.	Stockholm	Electronic road pricing
2.	Barcelona	Urban lab—dynamic traffic forecasting
3.	London	Electronic journey planner
4.	Hong Kong	Public light bus
5.	Copenhagen	Integrative public transport model
6.	Hangzhou	Public cycling system
7.	United Kingdom	Active traffic management approach

avoiding congestion during peak hours is a major issue to be resolved. Especially during weekends and holidays, this city receives a huge number of vehicles and managing traffic on those days is complicated. ITS applications provide smart solutions in resolving traffic issues with data analytics and machine learning algorithms. Predicting short-term traffic flow will assist in managing the traffic system and improving traffic issues. Our proposed short-term traffic flow prediction based on structure and pattern will improve the traffic system, which will also enhance green priority, rerouting, fuel consumption and be time-saving. The various highway roads connecting Bangalore is shown in Fig. 1. This express the highway routes connecting Bangalore city and it is used to indicate the amount of vehicles entering into city is more in numbers, and traffic prediction system is required to manage the traffic control system. One of the most heavily occupied route paths is shown in Fig. 2 and it is the toll route path from where the toll data information was gathered. The vehicles are always found to be higher and this route leads to congestion most often.

Traffic congestion can be avoided by predicting the number of vehicles entering the city and finding the traffic structure pattern between the current station and upstream station. As six wheelers and four wheelers are much more in Bangalore, it will be better if we predict congestion that happens because of these vehicles. To reduce congestion, four wheelers and two wheelers can be re-routed to save the time and cost. With this, fuel efficiency can also be achieved. The information from the toll was collected by manual observation of parameters from two toll plaza. The First Toll plaza is Chennai–Bangalore National Highways, and Second Toll plaza connecting Tirupathi–Bangalore. The traffic between the current section and the upstream station (entering into the city) can be predicted to manage traffic efficiently.

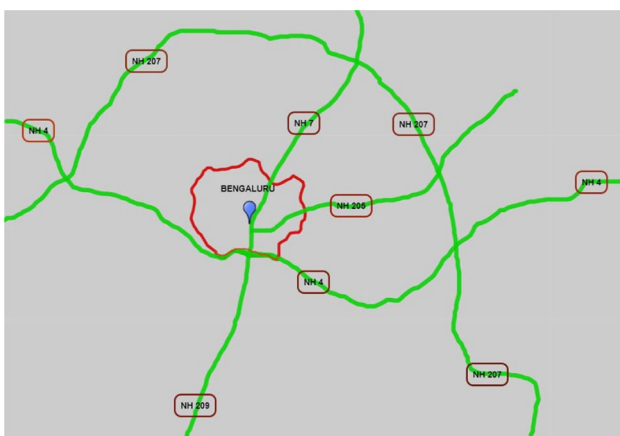


Fig. 1 Highways connecting Bengaluru

### 3.1 Data collection

To compute the traffic flow in a period, we analyze the toll data of two different plaza’s of Bangalore city for the period of last two weeks in the month of January 2018 in a freeway network. The collected information can be analyzed to predict the traffic structure and in avoiding the congestion during peak hours and especially on holidays. The properties of the Freeway toll data are given in Table 2.

### 3.2 Traffic flow analysis

The traffic flow analysis in a week of time is depicted in Fig. 3. Making use of the massive toll data and data’s from sensors and cameras installed across the city in a particular point of interest, we can identify the pattern of flow of the vehicles. The time of entry and exit of the vehicles from a particular region to another region and the total amount of time taken for transit between them which included overheads like traffic and other domestic problems is to be calculated. Buses and other heavy vehicles are the main reason for traffic congestion. With these vehicular data, traffic congestion can be predicted and light vehicles can be informed to divert their route. The entry/exit of vehicles

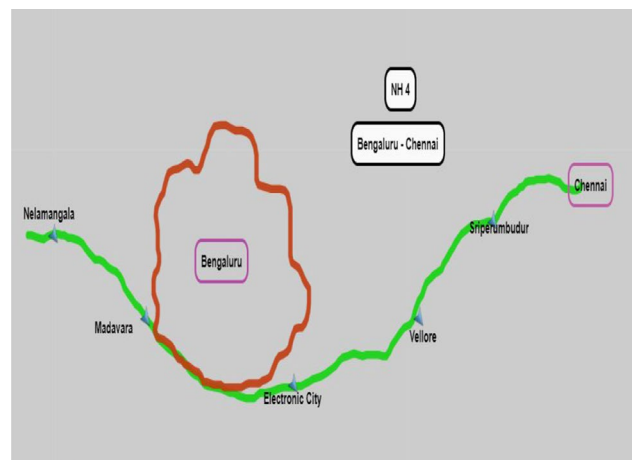


Fig. 2 NH-44 highway route (sample)

Table 2 Primary properties of the collected freeway toll data

S. no	Name	Description
1.	$V_{ID}$	Vehicle identifier
2.	$S_N$	Source name
3.	$D_N$	Destination name
4.	$T_{In}$	T-In time in time of the vehicle entry
5.	$T_{Out}$	T-out time out of the vehicle
6.	$V_h$	Number of vehicles crossed per hour

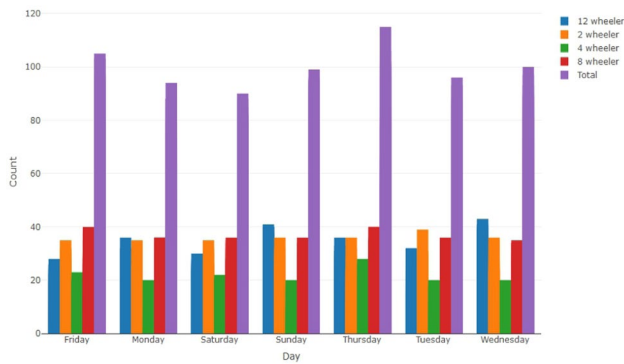


Fig. 3 Traffic analysis

in a certain area and the traffic flow at a definite time interval can be found. The time period is assigned to be 20 min. The flow path of a vehicle can be characterized as follows:

$$V_{Traj} = \{(T_{V-in}, T_{V-out})\} \quad (i \geq 1) \tag{1}$$

where  $V_{Traj}$  means that a trajectory of a vehicle is through a uninterrupted road section in a city.  $T_{V-in}$  means the entry time of the vehicle into the road segment 'i' entering into city and  $T_{V-out}$  means the way out time. We can cluster the set of vehicles those have passed. The traffic flow structure at current time period  $SV_R$  can be articulated as a vector

$$SV_{Ri} = \langle V_{Gi_1}, V_{Gi_2}, \dots, V_{Gi_N} \rangle \tag{2}$$

where  $SV_{Ri}$  represents the structure of a section of different vehicle types;  $V_{Gi_1}$  is the quantity of vehicles on the current road segment at the current time period. The sum of  $V_{Gi_1}$  is the traffic flow of this current section. To compute the traffic flow on a definite road section, we use an undersized set of components  $V_{Gi_1}$ .

If the vehicles are grouped majorly under three categories like Light Motor Vehicles (LMV), Heavy Motor Vehicles (HMV) and Heavy Passenger Motor Vehicle (HPMV), then the above traffic flow equation can be written as:

$$SV_{R1} = \langle V_{G1_1}, V_{G1_2}, \dots, V_{G1_N} \rangle \tag{3}$$

where i represents the flow of LMV

$$SV_{R2} = \langle V_{G2_1}, V_{G2_2}, \dots, V_{G2_N} \rangle \tag{4}$$

where i represents the flow of HMV

$$SV_{R3} = \langle V_{G3_1}, V_{G3_2}, \dots, V_{G3_N} \rangle \tag{5}$$

where i represents the flow of HPMV.

To compute the total number of vehicles in a particular road section at the current time interval,

$$V_{G1_1} + V_{G2_1} = V_{TC1}$$

$$V_{G1_2} + V_{G2_2} = V_{TC2}$$

Similarly,

$$V_{G1_N} + V_{G2_N} = V_{TCN} \tag{5}$$

where  $V_{TCk}$  ( $k \geq 1$ ) represents the total number of vehicles of different types in a particular road section at the current time interval.

Therefore

$$SV_R = \langle V_{TC1} + V_{TC2} + \dots + V_{TCN} \rangle \tag{6}$$

The traffic on a particular road at a time interval of 20 min can be known using  $V_{TC1}$  which represents the traffic flow in the given time interval and  $V_{TC2}$  is the traffic flow on that particular road section in the next time interval.

## 4 Traffic flow-based prediction method

To predict the traffic flow structure pattern between the upstream and downstream stations, we require the information about the vehicles origin and destination and facts in the field of traffic. The accuracy of the prediction method is most important in abnormal traffic states and especially during peak hours. The prediction of traffic flow structure pattern will enhance the accuracy of the prediction system. Prediction methodologies may not avoid the entire congestion rate, but this can be helpful in congestion avoidance. ITS supports in various aspects of improving traffic efficiency in many of the smart cities. From a holistic perspective, ITS exploitation is a key to maximize traffic efficiently and improving the congestion scenario. For example, ITS services have helped the Republic of Korea in increasing the average speed of traffic by 15–20%.

Providing real-time traffic information to travelers on the road is one of the efficient way to improve ITS. This will help to avoid traffic jams and also in managing their speed wisely. With the reduced traffic congestion and improved traffic efficiency, most of the public transportation services can be more reliable and punctual. This will also help in reducing traffic accidents and rash driving. Emergency vehicles like ambulances can be alerted before their arrival so that an alternate path can be taken to save the life and thus make the vehicles free from congestion. This further reduces congestion and pre-empts accidents.

### 4.1 Locally weighted learning

It is a lazy learning method in which generalization beyond the training data is delayed until a query is made to the system. This learning model deploys both linear and

non-linear models to fit the nearby points and then applies those values to forecast the values of the query points. Locally Weighted Learning (LWL) can also be used to include currently obtained data into the historical record to assist with learning. To get a high accuracy of prediction [5], this algorithm can be deployed. This algorithm is lazy behavior and is based on a plentiful training set. In some cases, unfortunately, this is too unfortunate to support the learning process suitable to this insufficient historical information and new queries. The regular traditional LWL will be unable to employ in abnormal traffic state and certain strategy can be introduced to assist regression [14].

LWL algorithm can be effectively used in cases where historical data is abundantly available and there are no new queries. In the traffic prediction system, the historical data is abundant making LWL algorithms most suitable for traffic prediction involving both upstream and downstream traffics. This method can also be employed in undisciplined traffic state scenarios. With this learning algorithm, the traffic structure pattern of the road section can be found and the spatiotemporal connection between the traffic stream on the current segment and the upstream section will be developed.

### 4.2 Prediction technique based on traffic flow pattern

The primary idea is to improve LWL with the constancy structure outline of traffic flow is one part and the other is to make use of the upstream station's entry flow to accurate the prediction of the segment traffic flow. The initial concern is to direct LWL to select feature space and Inherent periodicity. This is since the segment flows have good innate periodicity, particularly those high flow zone their quantity at preceding hours can be selected as features. The structure pattern of the current section flow is collected of a smaller amount of four major components. To be clear, a small number of upstream locations have a control on the current section. The entry flow from these

**Table 3** Variables used in prediction methodology

Name	Description
$TF_{ts}$	Target section traffic flow
$TF_{up}$	Upstream position entry flow
$b_i$	Data point
$b_i$	Traffic flow of target segment corresponding to $X_i$
TD	Labeled training data
$a_q$	Query point, prediction
$d(a_i, \hat{b}_q = \hat{b}_q + S_r \beta_r)$	Relevance of training points
$C_s$	Components in flow structure
$C_r$	Input rate
$C(t)$	Correction function
$V(t)$	Prediction value
$K(d)$	Gaussian kernel

upstream stations can be preferred as the characteristic of LWL. The variables used in the prediction methodology are described in Table 3.

In general, a standard regression model like  $b = f(a) + \epsilon$  is assumed with a continuous function  $f(a)$  and noise  $\epsilon$ . The basic cost function of LWL is defined as

$$LW = \frac{1}{2} \sum_{i=1}^n w_i(a_q)(b_i - a_i \beta_b)^2 \tag{7}$$

With the components

- 1 Labeled training data  $TD = \{(a_i, b_i) - i = 1, 2, \dots, n\}$  where each data point  $a_i$  belongs to a corresponding output value  $b_i$ .
- 2 P.o.I (Point of Interest or Query Point)  $a_q$  which is the position where we want a prediction  $\hat{b}_q$ .
- 3 Weight ( $w_i$ ) describe the relevance of the corresponding training set for the current prediction.
- 4 Regression coefficient  $\beta_b$  needed for doing prediction

**Algorithm 1 - Prediction Method based on Traffic flow**

Step 1: Data Collection in the given time granularity.

Step 2: Compute  $SV_{R1}, SV_{R2}, SV_{R3}$

Step 3: Compute the total number of vehicles in the current point of interest

3.1 Calculate for each  $V_T$

Step 4: Sequence of the queries ( $a_q = a_q - S_r P_r^n$ )

Step 5: Trace the vehicle trajectory and find the flow path -  $V_{Traj}$

Step 6: Output – Prediction Vector ( $\hat{b}_q$ )

Step 7: Process

7.1 Initialization

- a) Select the structure pattern of the target section
- b) Decide upstream station based on the flow structure
- c) Compute the transfer time ( $\Delta t$ ) and input rate ( $C_r$ ) of each upstream station

Step 8: Regression

- a) For each Query ( $a_q$ )

Calculate  $d(a_i, a_q)$  for each history points ( $a_i, b_i$ )

- b) Choose nearby neighboring points and arrange them in an order according to their distance
- c) Obtain each points value based on Gaussian Kernel  $K(d)$
- d) Compute  $V(t)$  and  $C(t)$
- e) Prediction Vector ( $\hat{b}_q$ ), where  $\hat{b}_q = \alpha \times V(t) + \beta \times C(t)$ . Initially  $\alpha$  and  $\beta$  value will be equal to 1.
- f) Sum the newly obtained points into the historical database of the pattern.

**Algorithm 2: Locally Weighted Projection Regression (LWPR) Prediction Algorithm—nonlinear function approximation**

Step 1: Input each query point  $a_q$   
 Step 2: Initialize the values  $a_q = a_q - a_0$  and  $\hat{b} = \beta_0$   
 Step 3: Prediction  
 For  $i = 1:R$  do  
 $S_r = u_r^T a_q$ —latent variable  
 $\hat{b}_q = \hat{b}_q + S_r \beta_r$ —Update prediction  
 $a_q = a_q - S_r p_r^n$ —reduce input space.  
 End

**5 Results**

We analyzed the traffic flow structure pattern by considering the traffic flow in the last two weeks of January 2018. In the abnormal days like republic day and Sundays, the numbers of HPMV vehicles entering towards the city are more. Other than the frequent light motor vehicles and heavy motor vehicles were running on those days. This leads to traffic congestion on the road which connects the Electronic city, Majestic, Market and Shivaji Nagar. Predicting traffic on these days will help avoid the traffic congestion. Light vehicles can be informed to take alternate routes to save their time and improve fuel efficiency [24]. The minimum and maximum traffic during the duration period were shown in Fig. 4. In the considered two weeks of data, the first week data were used for the prediction, the later seven days are for prediction to test the model's performance. The first week data distribution is shown in Fig. 5, which is used for future traffic flow forecast.

Root Mean Square Error (RMSE) is used to compute the exactness of the prediction, defined as.

$$RMSE(b, b') = \sqrt{\frac{\sum_{n=1}^N ((b(n) - b'(n))^2)}{N}} \tag{8}$$

where  $b$  is the actual quantity and  $b'$  is the related prediction while  $N$  is the total amount of prediction values. This

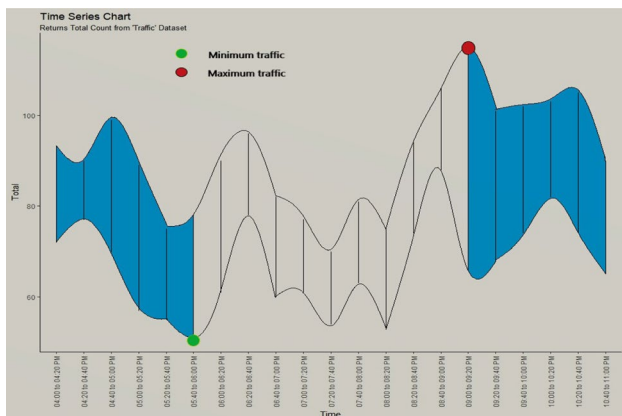


Fig. 4 Maximum and minimum traffic in the duration period

is a typical method to find out the error of a model in predicting quantitative facts. The model with a smaller RMSE has superior precision. This results in better accuracy. The experiment results are given in Table 4.

From the Table 4 and Figs. 4 and 5, it is proved that the proposed method is better than the traditional LWL method on the measured day. The RMSE value of the proposed work is smaller than the traditional LWL approach; this proves the results of the proposed work are more accurate than the traditional approach. The proposed method provides better results accurately and efficiently especially on holidays and abnormal traffic scenarios. This system is better to handle when there is a sudden increase in traffic flow and can support in finding the maximum traffic in the given duration.

**6 Discussions**

The proposed Short-Term Traffic Prediction Model (STTPM) is adaptable for Indian traffic system. This fits best for abnormal traffic cases and especially in holidays. The maximum and minimum traffic in the considered duration period are shown in Fig. 6. Traffic was found to be maximum during weekends and government holidays; this can be used to predict the traffic during abnormal states. To verify the stability of the structure pattern, we use the Coefficient of Variation (CV) as a stability index [17].

Figure 7 shows the mode of displaying the distribution of data based on five categories. Minimum, first quartile, median, third quartile and maximum. IQR is the Interquartile range. This helps in predicting and managing the traffic scenario particularly during weekends and holidays. The proposed approach is optimal for abnormal traffic scenarios.

Figure 8 shows the comparison of the proposed approach with regular LWL approach [14]. The line with red is the trained model prediction line, on which comparison is made. The yellow line is the regular approach and the green line is the proposed approach. The line which is closer to red line can be considered as a superior approach. In this, we can observe that at most of places or data points, the green line is smoother than the yellow line and closer to the red line. The proposed approach is superior to the existing approach. The performance of the prediction line of the trained model is wiser than the regular model approach. According to the data collected and used, STTPM is the finest prediction model. This show the proposed system results are overall 30% better than the regular approaches.

The performance evaluation of the proposed scheme with the various related schemes are presented in Fig. 9. Accuracy in prediction, cost, and energy are the various



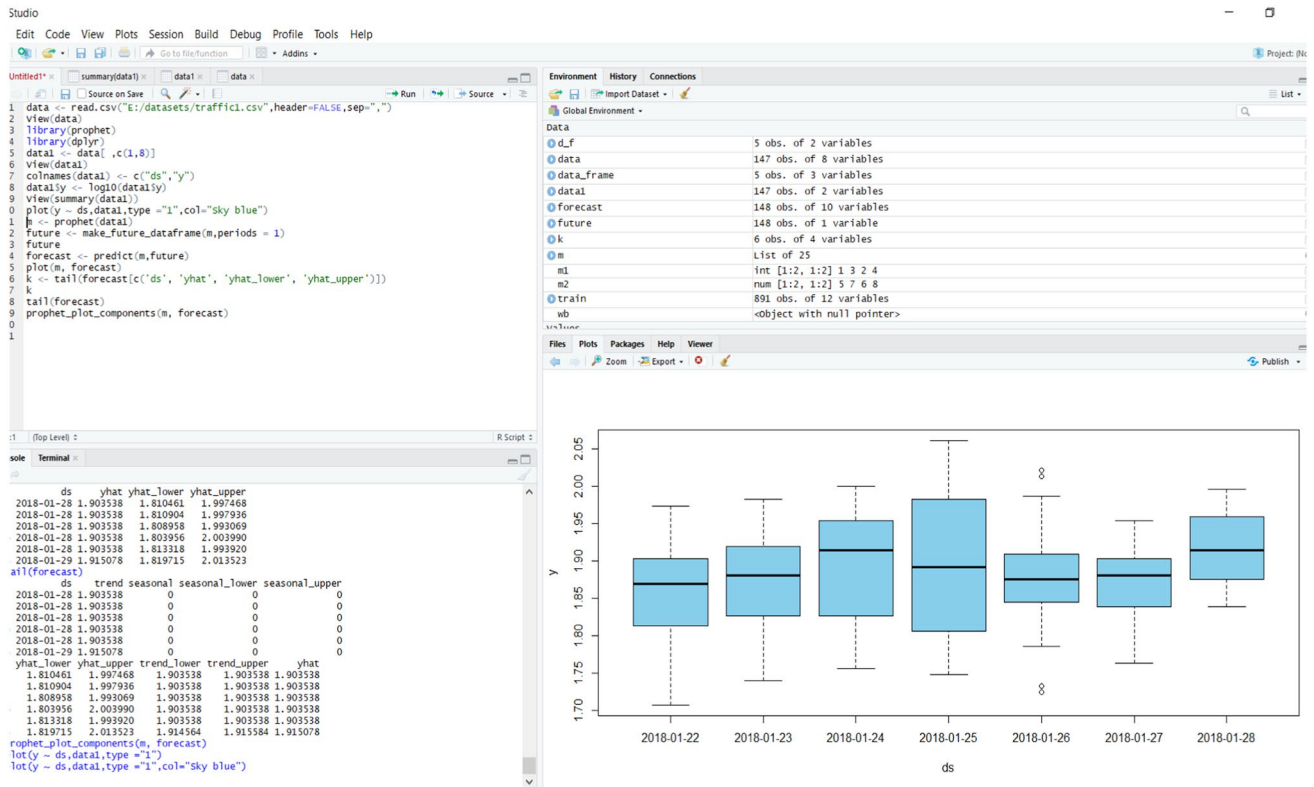


Fig. 5 Distribution of data in the duration period and future forecast

Table 4 RMSE of experiments

RMSE	Traditional LWL	STTPM
Day 4	35.60158	17.05670
Day 7	42.90654	20.04503

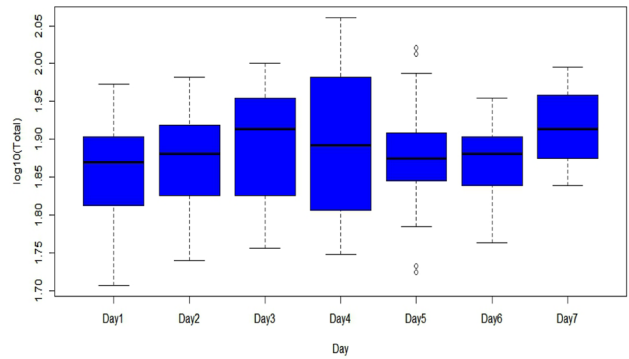


Fig. 7 Distribution of data especially on holidays

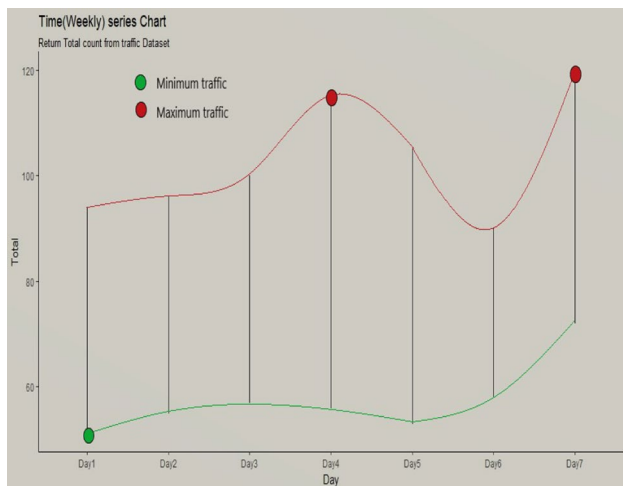


Fig. 6 Maximum and minimum traffic

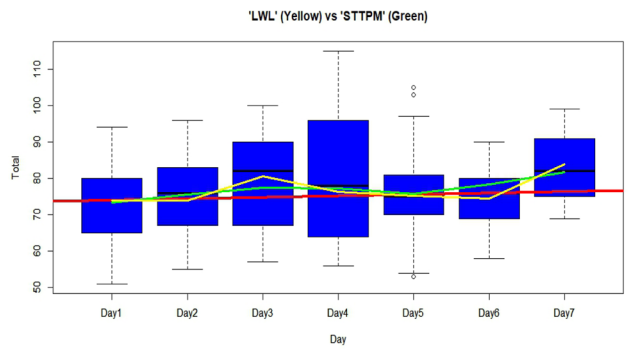
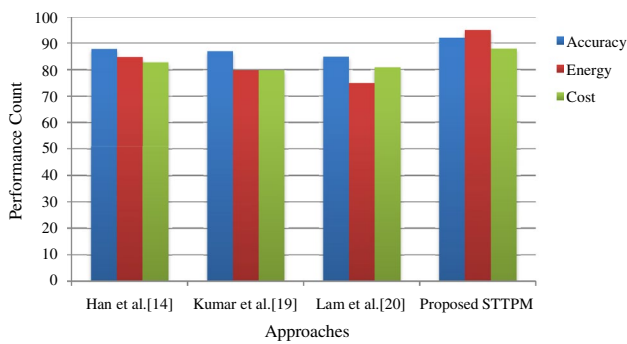


Fig. 8 Comparison of LWL versus STTPM



**Fig. 9** Comparison of various approaches with proposed

key factors considered for the comparison with various approaches. The accuracy of the results of the regular LWL approach [14] is proved to have higher RMSE value than the proposed STTPM approach. The proposed approach is suitable for short-term traffic flow prediction. The short-term traffic flow prediction using seasonal ARIMA model [19] is suitable for only limited set of data. This model is not suitable for abnormal traffic scenarios and data availability is also a major issue. Short-term travel time prediction [20] employs support vector regression for time-series travel time prediction. The proposed work integrates various factors like accuracy in short-term traffic flow prediction, consumes fuel and time, fits for abnormal traffic scenarios, especially on holidays and peak hours, moreover this fits best for Indian traffic system.

## 7 Conclusion and future work

We propose an efficient short-term traffic flow prediction methodology based on structure pattern traffic flow and regression methods. We analyzed the traffic flow congestion on the two most frequent traveling roads. The traffic flow at the next time point can be predicted by using entrance flow information. Initially, we collected the data to predict the traffic flow and then we compute the traffic occurring at the next stations. With the pattern model information and LWL, we predict the traffic flow in a week. The results show that the prediction based on structure pattern will be an efficient approach for predicting the traffic flow in metropolitan cities. This will be more useful in predicting the abnormal traffic state in India. This work can be extended by considering the other important external factors like climatic conditions, road diversions due to construction or road damage and traffic incidents. With the involvement of these factors, this will be a more effective system for smarter cities. The current proposed approach can be made capable to handle the data collection from

sensors or from loop detectors and this could be achieved better by integrating with Internet of Things (IoT).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1. Abadi A, Rajabioun T, Ioannou PA (2014) Traffic flow prediction for road transportation networks with limited traffic data. *IEEE Trans Intell Transp Syst* 16(2):653–662
2. Agarwal S, Basu K, Nath A (2013) Green computing and green technology based teaching learning and administration in higher education institutions. *Int J Adv Comput Res* 3(3):295
3. Agarwal S (2014) Impact of green computing in it industry to make eco friendly environment. *J Glob Res Comput Sci* 5(4):05–10
4. Ahmed MS, Cook AR (1979) Analysis of freeway traffic time-series data by using Box–Jenkins techniques. No. 722
5. Atkeson CG, Moore AW, Schaal S (1997) Locally weighted learning for control. *Art Intell Rev* 11:75–113
6. Atrey A, Jain N, Iyengar N (2013) A study on green cloud computing. *Int J Grid Distrib Comput* 6(6):93–102
7. Balamurugan B, Venkatakrishna P, UmaMageshwari R, Gomathi P (2014) Prevention techniques for traffic congestion based on intelligent light control and deviation system. *Internetwork Indones J* 6(1B):7–13
8. Balsys K, Eidukas D, Marma A, Valinevičius A, Žilys M (2007) Systems of transport route development. *Electron Electric Eng* 75(3):17–22
9. Boltze M, Tuan VA (2016) Approaches to achieve sustainability in traffic management. *Procedia Eng* 142:205–212
10. Čepėnas M Research on sensors for transportation control systems. PhD diss., Kaunas University of Technology
11. Fleming CH, Leveson NG (2016) Early concept development and safety analysis of future transportation systems. *IEEE Trans Intell Transp Syst* 17(12):3512–3523
12. Garcia S, Luengo J, Sáez JA, Lopez V, Herrera F (2012) A survey of discretization techniques: taxonomy and empirical analysis in supervised learning. *IEEE Trans Knowl Data Eng* 25(4):734–750
13. Gu Y, Li Y, Xu J, Liu Y (2011) Novel model based on wavelet transform and GA-fuzzy neural network applied to short time traffic flow prediction. *Inf Technol J* 10(11):2105–2111
14. Han L, Wu J, Gu P, Xie K, Song G, Tang S, Yang D, Jiao B, Gao F (2010) Adaptive knowledge transfer based on locally weighted learning. In: 2010 International conference on technologies and applications of artificial intelligence. IEEE, pp 392–397
15. Huang W, Song G, Hong H, Xie K (2014) Deep architecture for traffic flow prediction: deep belief networks with multitask learning. *IEEE Trans Intell Transp Syst* 15(5):2191–2201
16. Jeong Y-S, Byon Y-J, Castro-Neto MM, Easa SM (2013) Supervised weighting-online learning algorithm for short-term traffic flow prediction. *IEEE Trans Intell Transp Syst* 14(4):1700–1707
17. Jin C, Wei DX, Low SH (2004) FAST TCP: motivation, architecture, algorithms, performance. In: IEEE INFOCOM 2004, vol 4. IEEE, pp 2490–2501
18. Khan Z, Anjum A, Kiani SL (2013) Cloud based big data analytics for smart future cities. In: 2013 IEEE/ACM 6th international conference on utility and cloud computing. IEEE, pp 381–386

19. Kumar SV, Vanajakshi L (2015) Short-term traffic flow prediction using seasonal ARIMA model with limited input data. *Eur Transp Res Rev* 7(3):21
20. Lam SHM, Toan TD (2008) Short-term travel time prediction using support vector regression. No. 08-0670
21. Lama V, Sharma SK, Goyal N, Singh M (2014) Going green: computing for a sustainable future for economy, environment and eco-friendly technology. *Int J Adv Res Comput Sci Softw Eng* 4(6):970–975
22. Lv Y, Duan Y, Kang W, Li Z, Wang F-Y (2014) Traffic flow prediction with big data: a deep learning approach. *IEEE Trans Intell Transp Syst* 16(2):865–873
23. Oh C, Park S (2011) Investigating the effects of daily travel time patterns on short-term prediction. *KSCE J Civ Eng* 15(7):1263
24. Osorio C, Nanduri K (2015) Energy-efficient urban traffic management: a microscopic simulation-based approach. *Transp Sci* 49(3):637–651
25. Park S, Kim S, Ha Y (2016) Highway traffic accident prediction using VDS big data analysis. *J Supercomput* 72(7):2815–2831
26. Sun H, Liu HX, Xiao H, He RR, Ran B (2003) Use of local linear regression model for short-term traffic forecasting. *Transp Res Rec* 1836(1):143–150
27. Sun S, Zhang C, Yu G (2006) A Bayesian network approach to traffic flow forecasting. *IEEE Trans Intell Transp Syst* 7(1):124–132
28. Tang J, Liu F, Zou Y, Zhang W, Wang Y (2017) An improved fuzzy neural network for traffic speed prediction considering periodic characteristic. *IEEE Trans Intell Transp Syst* 18(9):2340–2350
29. Wu C-H, Ho J-M, Lee D-T (2004) Travel-time prediction with support vector regression. *IEEE Trans Intell Transp Syst* 5(4):276–281
30. Xia X, Lo D, Wang X, Yang X, Li S, Sun J (2013) A comparative study of supervised learning algorithms for re-opened bug prediction. In: 2013 17th European conference on software maintenance and reengineering. IEEE, pp 331–334
31. Zhang P, Xie K, Song G (2012) A short-term freeway traffic flow prediction method based on road section traffic flow structure pattern. In: 2012 15th international IEEE conference on intelligent transportation systems. IEEE, pp 534–539
32. ZhaAtkeson, C.G., Moore, A.W. & Schaal, S. Locally Weighted Learning for Control. *Artificial Intelligence Review* 11, 75–113 (1997).ng Y, Hou Z (2018) Short term traffic flow prediction based on improved support vector machine. *J Appl Sci Eng* 21(1):25–32

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.