Time-Space Bounded Traffic Forecasting Model for Smart Cities using IoT and Machine Learning

Hamdan Alshehri*

Department of Information Technology and Security, Jazan University, Saudi Arabia

**Corresponding author; E-mail:halshehri@jazanu.edu.sa*

Abstract:
Every country is now embracing the smart city (SC) concept with the goal of incorporating cutting-edge technologies into every aspect of urban life. Maintaining traffic in cities is one of the most difficult tasks and a variety of approaches are used to accomplish it. Considering this issue, an Internet of Things (IoT)-based Machine Learning (ML) framework is proposed to estimate vehicle traffic in a particular area of the city to effectively manage the traffic. The traffic data was collected using IoT edge sensors. The data obtained from the Cloud server has been processed using a data simulator along with other traffic-related data. To create the learning machines for the proposed forecasting system, the Support Vector Machine (SVM) regression algorithm was used. The trained model was then applied to test samples and evaluated using the Pearson r correlation (PrC) and mean absolute error (MAE). Time and location-specific forecasting were exercised to predict traffic delays in the city. On evaluation, the proposed model demonstrated better performance and is therefore recommended for use in SC traffic management.

Key words: *Internet of Things (IoT), Machine Learning (ML), Support Vector Regression (SVR), Smart City (SC), Forecasting, Traffic Prediction.*

1. INTRODUCTION:

A smart city is a framework, primarily made up of information and communication technologies (ICT), that is used to create, deploy, and promote sustainable development methods to handle the growing issues of urbanization (Camero and Alba, 2019). A smart network of connected items and equipment transferring data utilizing wireless technologies and the Cloud is a key component of this ICT system. Real-time data is received, analyzed, and managed by Cloud-based IoT apps to assist municipalities, businesses, and individuals in making better choices that improve people's lives. Citizens interact with green technology ecosystems in a variety of ways, including through their smart phones and mobile devices, as well as through connected cars and residences. When devices and data are linked to a city's tangible infrastructural facilities, the costs can be lowered while sustainability is increased. With the help of IoT, communities can enhance energy distribution, speed trash collection, improve traffic, and improve air quality.

Traffic management in smart cities is a technological solution that enables rapid cost-effective improvements in street safety and traffic flow. By installing these technologies to enhance the city's existing traffic infrastructure, the goal is to achieve efficient traffic management, cost and energy savings, and to contribute to the smart city initiative (Cvar *et al.*, 2020).

Sensors, cameras, cellular routers, and automation are used in these systems to monitor and automatically divert traffic, reducing congestion. The appropriate technological solution may be expanded and upgraded at any point in time. Concurrently, these technological solutions position Smart Cities for future technological evolutions like Internet of Vehicles (IoV) (Khoukhi *et al*., 2021). Figure 1 illustrates the overall IoT architecture of a smart city.

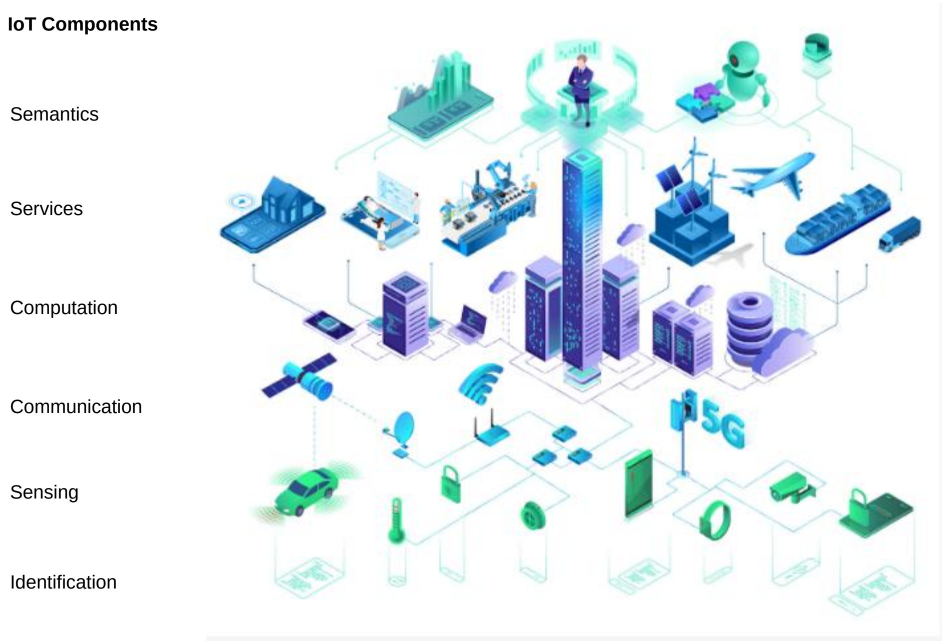


Figure 1. Smart City IoT Architecture (Image source: Nina Cvar et al. [2])

Smart Cities combine a wide range of ICT and advanced technologies that have the potential to revolutionize the country's social and economic factors such as education, energy, healthcare, sanitation, transportation, and trade. Many promising studies in this field have emerged with very novel approaches involving Big Data Analytics (BDA), Artificial Intelligence (AI), Machine Learning (ML), edge computing, and the Internet of Things (IoT).

Ang *et al.* (2022) published a comprehensive review article that covered a wide range of emerging concepts associated with the implementation of SCs. The review article includes detailed descriptions of the geospatial monitoring of SCs transportation, data analytics-driven transportation in SCs, machine learning (ML), deep learning (DL) supported classification, and forecasting models for SC transportation, as well as future SC deployment prospects. The usefulness of the aforementioned strategies was also evaluated by examining the outcomes of the various SC models proposed by various researchers in the recent past. Yin *et al*. (2022) conducted a review focused on DL-based traffic prediction as well as a few strategies that are important for SC transportation. IoT-based frameworks are not covered in this article.

Recently, Razali *et al.* (2021) published a review article with the objective of finding the limitations of the traffic prediction frameworks using ML and DL since 2016. The review work was conducted based on highly impacted research articles from top-rated databases. They came up with important findings as a result of their analysis. According to their findings, convolutional neural networks (CNN) and long-short term memory (LSTM) are the most commonly used machine learning algorithms for traffic flow prediction. Another important finding is that most of the research did not reveal the sources of the data and only a few used the existing databases for prediction. This could be overcome by adopting IoT technology to get real-time data for prediction. Yuan *et al*. (2021) presented a review study with noteworthy discoveries related to next-generation smart transportation. It focuses more on ML approaches and future potential to improve ML frameworks in relation to smart city transportation.

All of the review articles came to the same conclusion: ML-based frameworks for smart city traffic prediction and management must be combined with IoT in order to support real-time data collection and analysis. IoT is by definition ubiquitous, so being available anywhere is one of its primary goals. ML will play a key role in this by mining the data generated by thousands and millions of connected devices. To enable the IoT to play a significant role in achieving its goals, embedded intelligence (EI) will be at the core. EI is the combination of products and intelligence to improve automation, efficiency, productivity, and connectivity (Guo, B et al., 2013) Intelligence is acquired through learning, whether in the physical or virtual world. By bolstering this reasoning, this research proposes an ML-based prediction system based on IoT mechanisms. The major objectives of this proposed framework are:

- To review the related literatures on smart city traffic prediction and transportation managements.
- To look for the limitations of the related works.
- To design and evolve an ML-based traffic forecasting model for use in a smart city in a time-space bound manner.

Following the introduction section, this research paper is formatted as follows: section 2 provides a detailed analysis of the outcomes of the related literature. Section 3 details the proposed methodology for the development of the IoT enabled forecasting model using ML. The simulated results of the proposed ML model are presented in Section 4 with interpretations. The paper concludes with Section 5 by relating the objective of the proposed research to the results.

2. RELATED WORKS

This section presents a comprehensive assessment of similar research in smart city traffic prediction and forecasting with a focus on ML in conjunction with an IoT-based data collection method. The articles for the review were chosen from a variety of databases of the top journals.

Neelakandan *et al.* (2021) reported a benchmarking research outcome of an IoT-based traffic prediction and signal management system for a smart city utilizing an Intel 286 microprocessor. The proposed system was created in five phases: data gathering via IoT, feature engineering, data separation, traffic data optimization, and traffic management. To classify the traffic data in crowded areas, the Elman neural network method was used. The proposed system was claimed to outperform state-of-the-art approaches.

Lilhore *et al.* (2022) designed an adaptive traffic management system for SCs using ML and IoT. The work incorporated multiple scenarios to address all potential concerns within the city's transportation infrastructure. To detect any anomaly, the suggested system additionally employed a ML-based clustering algorithm. The traffic light schedules were regularly updated based on traffic movement and expected flow using the neighboring signal junctions. The simulation results show that the suggested approach performs better than the existing transportation strategies and shall be an added advantage in smart-city-based transportation systems, according to the authors. It is also claimed that the proposed method reduces traffic jamming, vehicle idle times, and disasters.

Elsagheer and AlShalfan (2021) made an analysis of all the services that an intelligent traffic management system (TMS) can provide. They proposed a smart TMS, which uses the existing IoV and VANET infrastructure to create an efficient and intelligent TMS without the need for any additional components, hardware, or deployment. They designed a traffic signal architecture and operation that is suitable for SCs. They used an optimization algorithm to provide efficient and near-optimal traffic management for local roads with any number of phases and completely customizable traffic management. The effectiveness of the suggested method was investigated and compared to the fixed-time approach using a simulator under various traffic situations. It is shown that their model surpasses standard traffic management systems in terms of the average waiting time and number of vehicles serviced.

Culita *et al.* (2020) suggested a neuro-inspired traffic control framework as part of a complex system of SC traffic. It was designed to demonstrate a holistic approach to the global problem of smart city traffic management, and it incorporates previously built urban traffic control architecture with the goal of proactively insuring it through traffic flow prediction. An analysis of the needs and prediction methods was carried out in order to establish the optimum strategy to satisfy the architecture's perception function in relation to the traffic control problem specification. On the basis of real-world data, both parametric and AI-based methods for predicting traffic flow are explored.

Lee and Chiu (2020) devised and implemented a smart traffic signal control mechanism to support numerous smart city transportation applications. It was engineered to be compatible with older traffic signal controllers, allowing for rapid and cost-effective deployment. According to the authors, the novel traffic signal scheme has been developed specifically for the emergency vehicle signal preemption concept. It can alert all drivers around the intersection to which direction the emergency vehicle is approaching, improving traffic flow and safety. The associated control algorithms were implemented and field tests were undertaken to demonstrate the system's performance.

In comparison to baseline approaches such as the historical average, the current time-based, and the Double Exponential Smoothing predictors, Toan *et al.* (2020) presented an efficient method for short-term traffic flow prediction utilizing an SVM. They used one-month time-series traffic flow data on a stretch of an expressway in Singapore for training and testing the SVM-based predictive model. According to the authors, for most of the prediction intervals and under varied traffic conditions, the SVM technique significantly outperforms the baseline methods with a rolling horizon of 30 minutes. The k-Nearest Neighbor (kNN) approach is also used with actual and simulated data replacing SVM training. The kNN approach is said to allow for a significant reduction in SVM training size, allowing for faster training without diminishing the predictive performance.

A ML and DL used framework was proposed by Navarro-Espinoza *et al.* (2022) for traffic flow prediction, which is area specific. It is stated that the DL-based model outperformed the ML-based model in terms of accuracy and training time. All of the ML and DL algorithms had high performance metrics, indicating that they might be used in smart traffic light controllers. Likewise, Ata *et al.* (2019) proposed an ANN-based traffic congestion control model with the goal of providing the greatest road service possible to citizens. The neural network was trained using the backpropagation technique which was used to forecast congestion. The proposed system intends to provide a solution that will boost the travelers' comfort levels, allowing them to make more informed and better transportation decisions. The authors suggest that the neural network is a feasible method for detecting traffic problems. Later, Ata *et al.* (2019) enhanced the framework by adopting the IoT mechanism for the real-time monitoring of the road traffic.

Hong *et al.* (2018) proposed a hybrid forecasting model based on the time-space bounded multifractal characteristics of traffic flow for short-term traffic. To uncover the traffic features behind the data, the proposed model decomposed the traffic flow series into four various components: a periodic part, a trend part, a stationary part, and a volatility part. To better understand the underlying traffic patterns and increase in forecasting accuracy, four elements were analyzed and modelled individually using various methods such as spectral analysis, time series, and statistical volatility analysis. The suggested hybrid model's performance was evaluated using real-time data from two cities. The experimental results showed that the suggested model outperforms the existing forecasting methods in terms of capturing nonlinear volatility and enhancing forecasting accuracy, particularly for multi-step forward forecasting, according to the authors. Likewise, several statistical (Ramesh *et al.*, 2021), machine learning (Meena *et al.*, 2020), and deep learning-based (Du *et al.*, 2018) traffic prediction frameworks put forward in the recent past as part of transforming the conventional traffic management enhance the traffic flow and SC implementation.

Certain facts were derived from the above papers after a careful review of the outcomes. Table 1 contains excerpts from the recent literature that are self-explanatory. One of the important facts found in the existing literature is that there is a lack of accurate ML analysis which relates to guaranteeing prediction accuracy. The proposed model was developed with this in mind. The methodology of the suggested framework has been detailed in the next section.

Table 1. Summary of the related works on SC traffic Prediction

Ref.	Proposed Work	Methods adopted	Outcome /Limitation
Neelakandan *et al.*, 2021	Traffic Prediction and Signal Management	IoT, Elman NN	Effective in prediction
Lilhore *et al.*, 2022	Adaptive Traffic Management System	IoT, ML	Reduced traffic jamming, vehicle idle times, and accidents
Elsagheer *et al.*, 2021	Intelligent Traffic Management System	IoV, VANET, Optimization and control algorithms	Reduced average waiting time and Increased number of vehicles serviced.
Culita *et al.*, 2020	Neuro-inspired traffic control framework/SC traffic flow prediction	Parametric and AI	Found to be effective in prediction
Lee *et al.*, 2020	Smart traffic signal control mechanism	Optimization and control algorithms	Found effective in a field test
Toan *et al.*, 2020	Short-term traffic flow prediction	ML: SVM, kNN	kNN found to be effective.
Navarro-Espinoza *et al.*, 2022	Traffic flow prediction	ML and DL	Found to be effective
Ata *et al.*, 2019	Traffic congestion control model	ANN	ANN is found to be feasible for more informed traffic predictions
Hong *et al.*, 2018	Hybrid forecasting model based on time-space	Spectral analysis, Time series, and Statistical volatility analysis	Found to be effective in prediction accuracy

3. MATERIALS AND METHODS

This section describes the strategies and algorithms that have been incorporated into the proposed framework. The suggested IoT architecture for traffic data gathering in the SC is detailed in the first part. The subsequent section outlines the SVM-based regression method. The dataset's specifics are then presented, followed by the evaluation metrics.

3.1 IoT Architecture for Smart City Traffic

An IoT-based data collection model has been used to generate the dataset for ML training. The AnyLogic simulator (Abdellah *et al.*, 2020) was used to model the IoT system for the proposed framework. Figure 2 shows the architecture of an IoT system for collecting and modelling traffic data. The model includes an IoT traffic system that simulates the process of an IoT device or set of IoT devices, a traffic source for traditional communication facilities, and traffic management. The purpose of incorporating IoT architecture in the proposed prediction problem is to obtain live traffic data and feed to the machine learning based regression model.

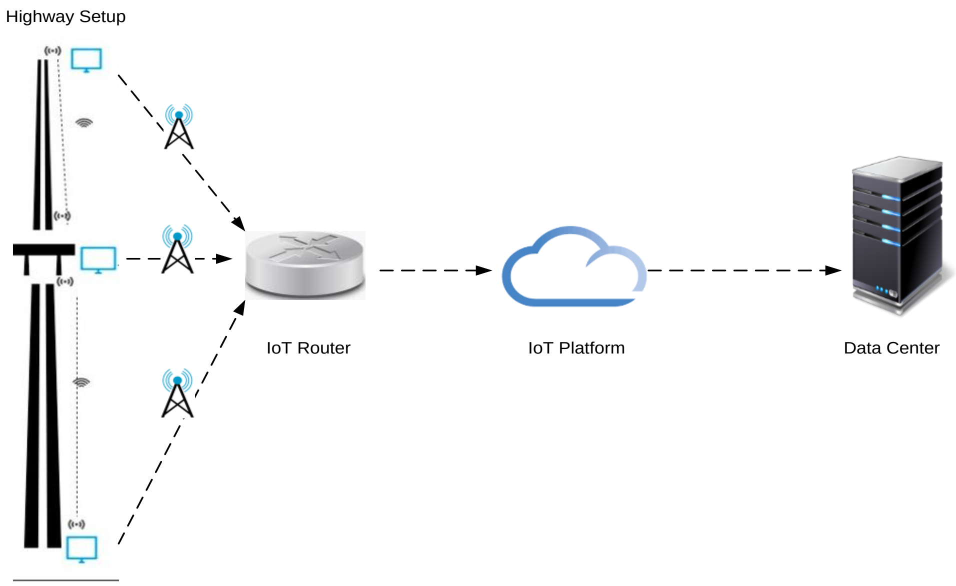


Figure 2. IoT Architecture for SC Traffic Data Collection

3.2 Time-Space Bounded Traffic Prediction

Foreseeing full city traffic is important for a smooth traffic flow as time and region-specific predictions are critical for adding value to the congestion-free movement of vehicles. The framework is designed to forecast traffic conditions in a certain area at a specific time. Because the traffic state change process is a real-time, nonlinear, high-dimensional, non-stationary random process, the randomness and uncertainty of traffic state changes increase when the statistical period is reduced. Short-term variations in traffic status are affected not only by the traffic status of the previous few periods of the road but also by the upstream and downstream traffic conditions. The difficulty when attempting short-term traffic state prediction is determining the regularity and establishing the forecast from traffic state changes with randomness and uncertainty based on the traffic state parameters received by the vehicle detector and other influencing factors (Tian *et al*., 2019).

3.3 ML Algorithm for Traffic Prediction

Because of its capacity to leverage the data potential that has become generally used for SC traffic by officials and academics, machine learning presents an excellent opportunity for the same. As well as the large volume of data created edge devices such as sensors, smart cards, cameras, etc. that cannot be reviewed individually, a system that can learn and optimize on its own is required. This is the significance of using ML in addition to dynamic and continuous learning techniques for SC traffic management. As a result, investigating the potential of machine learning in the development of personalized services for traffic management is crucial for greater efficiency in the SC establishment. Machine learning algorithms are continually evolving as a result of improved algorithms, data collection methods, and communication networks. The basic purpose of machine learning, similar to real-time data, is to effectively interpret incoming data and generate predictions well beyond the training set. Machine-learning-based approaches can provide descriptive, predictive, or prescriptive evaluations in general. The generalized framework of the SC traffic forecast is depicted in Figure 3.

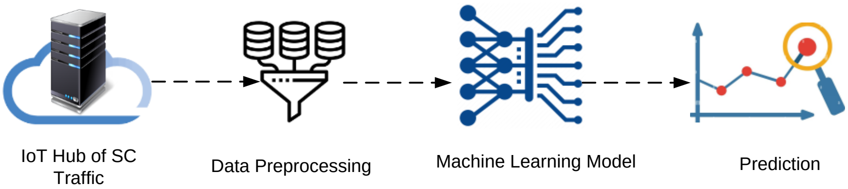


Figure 3. ML-based SC Traffic Prediction

3.3.1 Support Vector Regression (SVR)

A significant number of machine-learning-based research has appeared in the literature in recent decades, notably within a diverse usage of numerous machines learning approaches to examine a wide range of challenges in SC frameworks. Support vector regression (SVR), the machine learning algorithm used in our proposed SC traffic prediction system, is detailed here.

The Support Vector Machine (SVM) is a well-known supervised learning model for classification. Support Vector Regression (SVR) (Ho *et al.*, 2012) is a regression problem-solving technique that is analogous to SVM. SVR is an effective technique for real-value dataset predictions, despite its lack of popularity compared to SVM. The goal of the SVM is to find a hyperplane in an N-dimensional space where N is the number of features that classify the data points. The SVM is an extended form of the maximal margin classifier that allows for the improved classification of data in high-dimensional space (Franc *et al.*, 2003).

The implementation of the kernel function method can create a multiplier effect in traffic flow prediction based on large, high-dimensional, random, and nonlinear datasets. As a result, selecting the proper kernel function is a large part of employing the SVM method. Because of its low structural risk and the ability to avoid dimensionality issues and over-fitting, SVR is commonly employed in the transportation industry. SVR is a sparse-solution nonlinear kernel learning algorithm that translates linear inseparable problems into high-throughput kernel functions. The purpose of the linear problem in dimensional space is to minimize the risk of model structuring and to strike a balance between the data fitting precision and the approximation of function complexity to avoid overfitting. It is frequently used to anticipate traffic flow based on the difference in loss function. The SVR model consists primarily of three components: *e*SVR, *v*SVR, and least square SVR, with *e*SVR being the most frequently utilized. A predictive model of SVM regression can be created using the *e*SVR algorithm which is the basic idea of short-term traffic status prediction.

When building an SVM-based optimizer, it is important to choose the parameter and kernel function. The weighted minimizing factor and the underlying problem can be derived using the least square regression principle:

$$min\|w\|^2 + \sum_{i}^{l} \xi_i^2 \tag{1}$$

such that $y_i - < w.x_i > = \xi_i (i = 1,2,3 \ldots . l)$

The Lagrange function was obtained using the Lagrange theory as follows:

$$\min L(w, \xi, a) = \|w\|^2 + \sum_{i}^{l} \xi_i^2 + \sum_{i}^{l} a_i (y_i - < w.x_i > \xi_i) \tag{2}$$

The derivative was calculated and set it to 0. $w = \frac{a_i}{2}$

This was made into a dual problem:

$$\max_{a^* \in R} W(a^{(*)}) = \sum_{i}^{l} y_i (a_i^* - a_i) - \frac{1}{2} \sum_{i}^{l} (a_i^* - a_i) \sum_{j}^{l} (a_j^* - a_j) K(x_i, x_j) \tag{3}$$

Such that $\begin{cases} \sum_i^l (a_i^* - a_i) = 0 \\ a_i^* \in \left[\frac{C}{l}\right] i = 1,2, .. l \\ \sum_i^l (a_i^* + a_i) \leq C \end{cases}$

Where

i) The original objective function $W(a^{(*)})$ represents the minimum function of the difference between the actual and predicted values.

i) $K(x_i, x_j)$ is the kernel function that characterizes the feature space and calculates the target's output directly from nonlinear data.

ii) C was chosen based on the predictions of several training iterations.

The optimal solution was obtained from the dual problem $\overline{a^{(*)}} = \overline{a_1^*}, \overline{a_1}, \overline{a_2^*}, \overline{a_2} \ldots \overline{a_l^*}, \overline{a_l}$ which is similar to choosing the support vector from the training set as the weight coefficient in the optimization function for each output-input vector in the optimization set.

The prediction function was designed using the related regression equation $f(x) = < w.x > +b$ as follows:

$$f(x) = \sum_{i}^{l} (a_i^* - a_i) \; K(x, x_j) + b^* \tag{4}$$

x_i is the training sample in the training set S, which is obtained one-by-one by the i^{th} loop. x is the current time of the road segment *i* to be predicted back to the *p* time period (the t-p+1 time period to the first vector combination of various traffic state influence parameters at time *t*) in the kernel function of the prediction function.

Now, b^* was calculated as

$$b^* = \frac{1}{2} \left[y_i + y_k - \left(\sum_{i}^{l} (\overline{a_l^*} + \overline{a}_l) \, K(x_i, x_j) + \sum_{i}^{l} (\overline{a_l^*} + \overline{a}_l) \, K(x_i, x_k) \right) \right] \tag{5}$$

Where $\overline{a_j^*}$, $\overline{a_k}$ are lie between $\left(0,\frac{C}{l}\right)$. The projected result was represented by the value of the objective function *f(x)*.

A short-term traffic prediction was modelled using this SVR. Finally, the average traffic congestion of the future period can be forecasted using the prediction function *f(x)* and a combination of the traffic parameter x by applying Radial Bias Function (RBF) kernel. The Proposed SC traffic prediction framework using SVR is depicted in Figure 4.

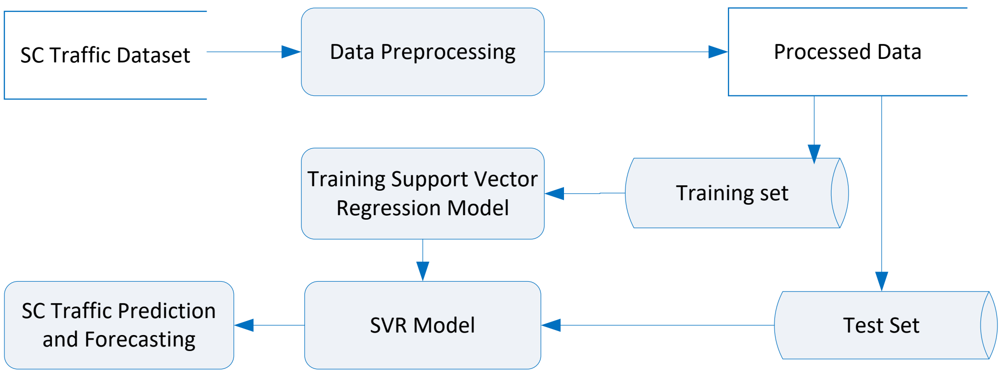


Figure 4. SC traffic prediction framework using SVR

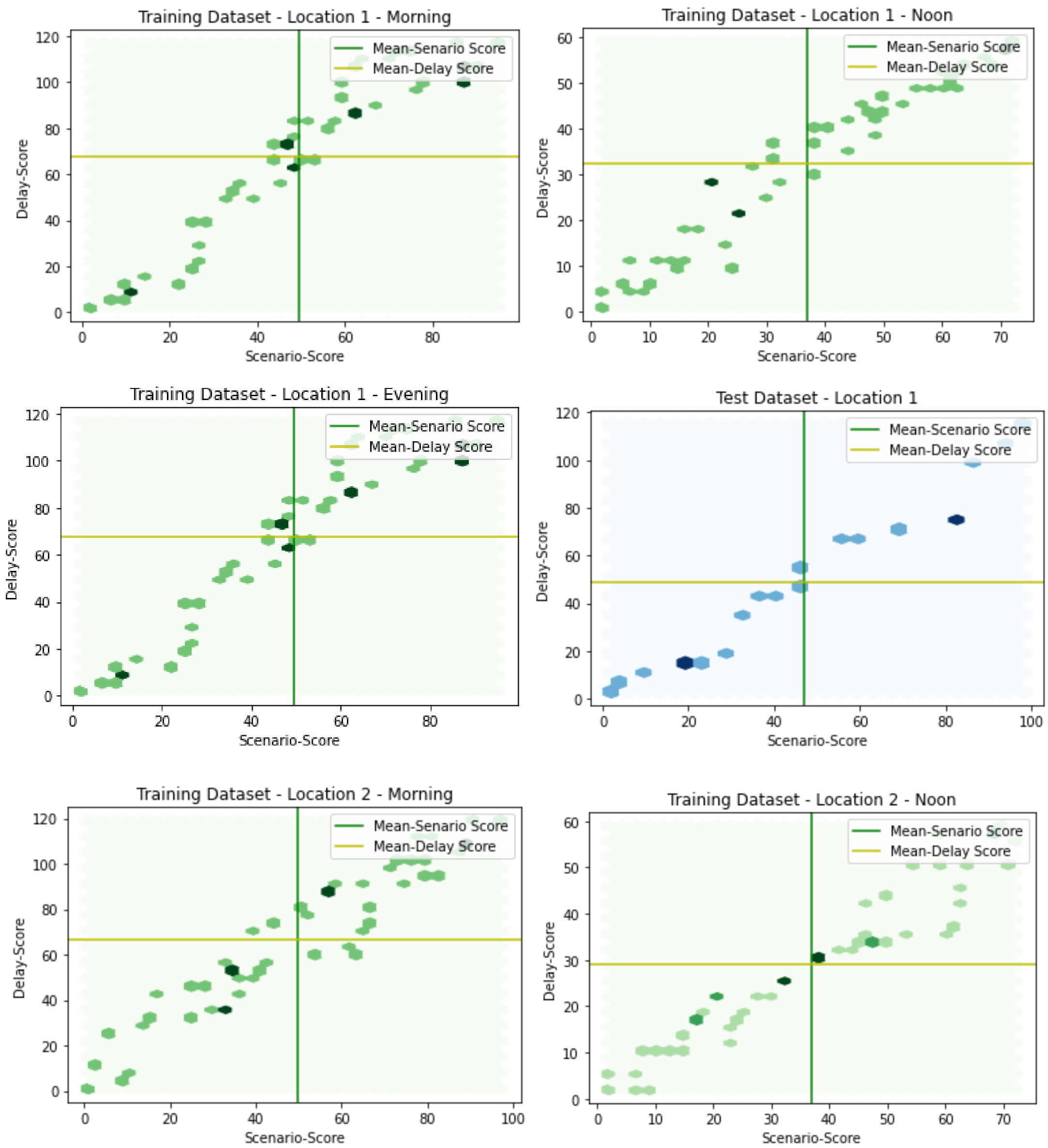
3.4 Training with the Dataset for SC Traffic Prediction

The dataset for this prediction problem was simulated using SC-based traffic scenarios. Five SC locations were chosen and dynamic traffic conditions such as the speed of the vehicles, the number of vehicles passing through, the weather conditions, key events, seasonal activities, and festivals were monitored in each location. The parameters were observed at three different time intervals in a day. Table 2 shows the specification of the dataset, which is self-explanatory.

Table 2. Dataset Specification

Space	**Time**	**Parameters**	**Scenario Score (predictor Variable)**	**Delay Score (Outcome variable)**
Location 1	Time-1 Time-2 Time-3	Speed of the vehicle (IoT sensor data)		
Location 2	Time-1 Time-2 Time-3	No. of vehicles/minute (IoT sensor data)		
Location 3	Time-1 Time-2 Time-3	Weather condition (IoT Sensor data) Peak time status	100 point simulated score based on the parameters	Observed delay (in minutes) from the ideal traffic flow
Location 4	Time-1 Time-2 Time-3	Key events/ Festivals Seasonal activities		
Location 5	Time-1 Time-2 Time-	Observed for 50days		

Based on the parameter values, the predictor variable was named '*scenario score*' and the value ranged from 1 to 100. The '*delay score'* is the output variable measure which represents the observed delay in the midst of the various traffic parameters. The prediction framework is devised in two ways: i) location and time specific traffic predictions and ii) overall city traffic predictions. Accordingly, the datasets were prepared for training with the SVR model. Figure 5 depicts the training and test datasets for the SC traffic predictions. Three location (location 1, 2 and 5) datasets are presented here along with all of the city traffic test dataset.



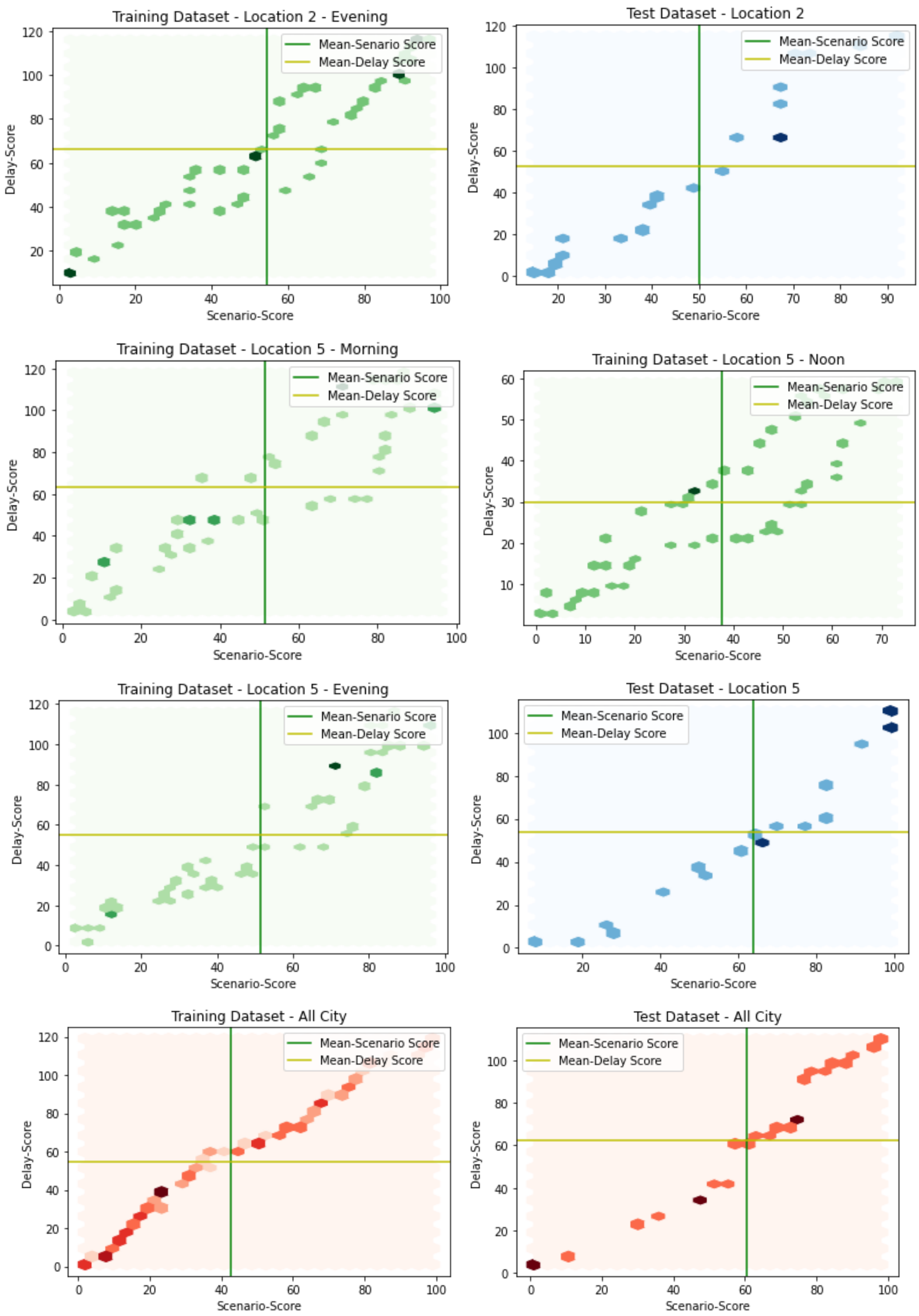


Figure 5. Training and Test Datasets

The preceding method incorporates all data from the study period and acquires all training sets from the research sample. Upon the successful training of the designed SVR model with each location dataset (50 samples each from three different time segments of each location), the model was then tested with test samples (20 random samples from each location). Similarly, to predict the overall city traffic, the SVR model was applied with separately prepared training samples and subsequently tested with the test samples.

3.5 Performance Estimators

The Pearson *r* correlation (PrC) coefficient and mean absolute error (MAE) were used to evaluate the prediction model's performance. The PrC coefficient was a widely used scientific measure in predictive analytics to ensure the relationship between the predictor variable and outcome variable. This will help the researchers to validate the data even before training with predictive models. For each location's dataset, the MAE is a measurement of the difference between the actual output obtained while testing and the desired output estimated. The permissible value ranges from 0 to infinity.

4. RESULTS AND DISCUSSION

Phase 1: The time-space specific prediction framework was implemented in the first phase. The SVR model was trained using simulated datasets for each location at three different time periods. Test datasets were used to evaluate the trained model once it had been successfully trained. The results of the prediction performance utilizing PrC and MAE are shown in Table 3. The PrC for the datasets are close to 1 which means that the variables are strongly correlated. The gap between the measured and anticipated delays was assessed to be 2.18 minutes on average when evaluating the model using test datasets. This demonstrated that the SVR model effectively predicts the SC scenario, which is temporal and area specific.

Table 3. Time-space bounded traffic prediction estimators

Dataset		PrC (0 to 1)	MAE (in mts)
Location 1	Morning	0.982	1.87
	Noon	0.992	3.28
	Evening	0.967	1.65
Location 2	Morning	0.993	2.38
	Noon	0.990	1.76
	Evening	0.977	2.91
Location 3	Morning	0.987	2.43
	Noon	0.990	1.27
	Evening	0.985	2.58
Location 4	Morning	0.989	1.88
	Noon	0.992	3.63
	Evening	0.976	1.45
Location 5	Morning	0.969	2.54
	Noon	0.991	1.32
	Evening	0.993	1.81
Overall		**0.984**	**2.18**

Phase 2: The SVR model was trained with 100 samples from various locations in the city traffic dataset to predict overall city traffic. Test datasets (20 random samples) were used to evaluate the regression model after it had been trained. Table 4 shows the estimated results after testing the model. A Pearson r correlation coefficient (PrC) of 0.905 was achieved, indicating that the predictor variable (scenario score) and the outcome variable (delay score) are highly correlated. The MAE of the experiment was 2.58 minutes which represents the difference between measured and projected values. This low value implies that the SVR model is capable of projecting traffic delays over the entire city.

Table 4. Entire city traffic prediction estimators

Dataset		PrC (0 to 1)	MAE (in mts)
City	Morning	0.918	2.77
	Noon	0.892	3.04
	Evening	0.907	1.93
Overall		**0.905**	**2.58**

When we compared our work to the benchmark prediction frameworks in table 1, we discovered that the proposed work performed equally well. It should also be noted that the majority of the benchmark SC prediction works were implemented without the use of an IoT mechanism for data collection. As a result, our work is unique in SC traffic prediction.

5. CONCLUSION

The regularization of transportation in smart cities is a challenging task, especially when dealing with traffic congestion in peak hours. Predicting the city traffic in advance is important when seeking to maintain smooth traffic flow with minimal delay. Moreover, time and location-based prediction is important as the delay is usually caused by several other factors like local events, institutions, etc., which are dynamic in nature in addition to temporal factors. By keeping this in mind, the SC-based traffic prediction framework was carried out. Location specific real-time traffic data obtained through IoT sensors was subjected to preprocessing along with other location specific data. A support vector machine (SVM) regression algorithm was used for this prediction framework. On evaluation, the model showcased better performance when predicting the city's traffic delays. This is highly recommended for the management of SC transportation. Prediction accuracy and inappropriate evaluation methods were the major issues found in several studies. This has been overcome by adopting a highly scientific method in our framework for evaluating the prediction performance. The proposed framework can be extended by incorporating real-time IoT sensor data.

ABREIVATIONS

AI – Artificial Intelligence
BDA – Big Data Analytics
IoT – Internet of Things
IoV – Internet of Vehicles
MAE – Mean Absolute Error
ML – Machine Learning
PrC – Pearson r Correlation
SC – Smart City
SVM – Support Vector Machine
SVR – Support Vector Regression
TMS - Traffic Management System

REFERENCES:

Abdellah, A.R., Mahmood, O.A., and Koucheryavy, A. 2020. Delay prediction in IoT using Machine Learning Approach. In: *Proceedings of the 2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, Brno, Czech Republic. pp.275–279.

Ang, K., Seng, J., Ngharamike, E. and Ijemaru, G., 2022. Emerging Technologies for Smart Cities' Transportation: Geo-Information, Data Analytics and Machine Learning Approaches. *ISPRS International Journal of Geo-Information*, 11(2), p.85. https://doi.org/10.3390/ijgi11020085

Ata, A., Muhammad, Khan., Sagheer, Abbas., Gulzar, Ahmad., and Fatima, Areej. 2019. Modelling Smart Road Traffic Congestion Control System Using Machine Learning Techniques. *Neural Network World*. 99. https://doi.org/10.14311/NNW.2019.29.008

Ata, A., Khan, M. A., Abbas, S., Khan, M. S., and Ahmad, G. 2019. Adaptive IoT Empowered Smart Road Traffic Congestion Control System Using Supervised Machine Learning Algorithm. *The Computer Journal*, 64(11), pp.1672-1679.

Camero, A. and Alba, E., 2019. Smart City and information technology: A review. *Cities*, 93, pp.84-94. https://doi.org/10.1016/j.cities.2019.04.014

Culita, J., Caramihai, S., Dumitrache, I., Moisescu, M. and Sacala, I., 2020. An Hybrid Approach for Urban Traffic Prediction and Control in Smart Cities. *Sensors*, 20(24), p.7209. https://doi.org/10.3390/s20247209

Cvar, N., Trilar, J., Kos, A., Volk, M. and Stojmenova Duh, E., 2020. The Use of IoT Technology in Smart Cities and Smart Villages: Similarities, Differences, and Future Prospects. *Sensors*, 20(14), p.3897. https://doi.org/10.3390/s20143897

Du, Shengdong., Li, Tianrui., Gong, Xun., Yu, Zeng., Horng, Shi-Jinn. 2018. A Hybrid Method for Traffic Flow Forecasting Using Multimodal Deep Learning. *International Journal of Computational Intelligence Systems*. 13. https://doi.org/10.2991/ijcis.d.200120.001

Elsagheer Mohamed, S. and AlShalfan, K., 2021. Intelligent Traffic Management System Based on the Internet of Vehicles (IoV). *Journal of Advanced Transportation*, 2021, pp.1-23. https://doi.org/10.1155/2021/4037533

Franc, V., and Hlaváč, V. 2003. An Iterative Algorithm Learning the Maximal Margin Classifier. *Pattern Recognition*. 36(9), pp.1985-1996.

Guo, B., Zhang, D., Yu, Z., Liang, Y., Wang, Z., and Zhou, X. 2013. From the Internet of Things to Embedded Intelligence. *World Wide Web 2013*, 16, 399–420. https://doi.org/10.1007/s11280-012-0188-y

Ho, C. H., and Lin, C. J. 2012. Large-Scale Linear Support Vector Regression. *The Journal of Machine Learning Research*. 13(1), pp.3323-3348.

Hong, Zhang., Xiaoming, Wang., Jie, Cao., M, Tang., and Yirong, Guo. 2018. A hybrid short-term traffic flow forecasting model based on time series multifractal characteristics. *Applied Intelligence*. 48(8), pp.2429–2440. https://doi.org/10.1007/s10489-017-1095-9

Khoukhi, L., Xiong, H., Kumari, S. and Puech, N., 2021. The Internet of vehicles and smart cities. *Annals of Telecommunications*, 76(9-10), pp.545-546. https://doi.org/10.1007/s12243-021-00891-7

Lee, W. and Chiu, C., 2020. Design and Implementation of a Smart Traffic Signal Control System for Smart City Applications. *Sensors*, 20(2), p.508. https://doi.org/10.3390/s20020508

Lilhore, U., Imoize, A., Li, C., Simaiya, S., Pani, S., Goyal, N., Kumar, A. and Lee, C., 2022. Design and Implementation of an ML and IoT Based Adaptive Traffic-Management System for Smart Cities. *Sensors*, 22(8), p.2908. https://doi.org/10.3390/s22082908

Meena, G., Sharma, D., and Mahrishi, M. 2020. Traffic Prediction for Intelligent Transportation System using Machine Learning. In: *3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE).* pp.145-148. https://doi.org/10.1109/ICETCE48199.2020.9091758

Navarro-Espinoza, A., López-Bonilla, O., García-Guerrero, E., Tlelo-Cuautle, E., López-Mancilla, D., Hernández-Mejía, C. and Inzunza-González, E., 2022. Traffic Flow Prediction for Smart Traffic Lights Using Machine Learning Algorithms. *Technologies*, 10(1), p.5. https://doi.org/10.3390/technologies10010005

Neelakandan, S., Berlin, M., Tripathi, S., Devi, V., Bhardwaj, I. and Arulkumar, N., 2021. IoT-based traffic prediction and traffic signal control system for smart city. *Soft Computing*, 25(18), pp.12241-12248. https://doi.org/10.1007/s00500-021-05896-x

Ramesh, K., Lakshna, A., Renjith, P.N., and Sheema, D. 2021. Smart Traffic Prediction and Congestion Reduction in Smart Cities. *Turkish Journal of Computer and Mathematics Education*. 12(12).

Razali, N., Shamsaimon, N., Ishak, K., Ramli, S., Amran, M. and Sukardi, S., 2021. Gap, techniques and evaluation: traffic flow prediction using machine learning and deep learning. *Journal of Big Data*, 8(1). https://doi.org/10.1186/s40537-021-00542-7

Tian, X., Yu, D., Xing, X., Wang, S., and Wang, Z. 2019. Hybrid short-term traffic flow prediction model of intersections based on improved complete ensemble empirical mode decomposition with adaptive noise. *Advances in Mechanical Engineering*. https://doi.org/10.1177/1687814019841819

Toan, T. and Truong, V., 2020. Support Vector Machine for Short-Term Traffic Flow Prediction and Improvement of Its Model Training using Nearest Neighbor Approach. Transportation Research Record: *Journal of the Transportation Research Board*, 2675(4), pp.362-373. https://doi.org/10.1177/0361198120980432

Yin, X., Wu, G., Wei, J., Shen, Y., Qi, H. and Yin, B., 2022. Deep Learning on Traffic Prediction: Methods, Analysis, and Future Directions. *IEEE Transactions on Intelligent Transportation Systems*, 23(6), pp.4927-4943. https://doi.org/10.48550/arXiv.2004.08555

Yuan, T., Rocha Neto, W., Rothenberg, C., Obraczka, K., Barakat, C. and Turletti, T. 2021. Machine learning for next-generation intelligent transportation systems: A survey. *Transactions on Emerging Telecommunications Technologies*, 33(4). https://doi.org/10.1002/ett.4427