Research Journal of Engineering Sciences ISSN 2278 – 9472
Vol. **15(3),** 10-14, September **(2026)** *Res. J. Engineering Sci.*

A Data-Agnostic Hybrid SARIMA–LSTM framework for Urban Traffic Monitoring in Smart cities

Shruti Vohra[1]*, Smita Verma[2] and Mohan Rawat[1]
[1]Department of Engineering Sciences and Humanity, Chameli Devi Group of Institutions, Indore 452020, MP, India
[2]Department of Applied Mathematics and Computational Science, SGSITS, Indore 452003, MP, India
shrutiavohra@gmail.com

Available online at: www.isca.in, www.isca.me
Received 9th June 2026, revised 26th July 2026, accepted 17th August 2026

Abstract

The monitoring of urban traffic is critical for effective management in smart city and transport systems. Quite often, existing forecasting approaches tend to be heavily dependent on bulky location-specific data sets, which results in their applicability being questionable under conditions of limited data. In this paper, we suggest developing a novel SARIMA-LSTM approach that combines the principles of statistical time-series analysis with deep learning-based correction of residuals. In the first place, the SARIMA algorithm is introduced in order to guarantee interpretability and statistically proven characteristics of our model due to the identification of linear and seasonal trends. At the same time, the LSTM network is applied to imitate the dynamic non-linearities in residuals in order to improve short-term forecasts. The theoretical development of the approach, its design and experimental verification are provided in the paper. Moreover, the suggested model structure offers an environmentally friendly opportunity for monitoring of urban traffic and facilitates the integration with future spatial-temporal and real-time intelligent transport systems.

Keywords: Data-agnostic modeling, Deep learning, Hybrid SARIMA-LSTM, Traffic forecasting.

Introduction

Rapid growth of urbanization and increased movement of people make traffic management systems in RBANs increasingly complex. Intelligent cities require continuous monitoring of traffic to mitigate traffic congestion, prevent collisions, and promote sustainable transport systems[1,2]. ARIMA and SARIMA models have traditionally been used in predicting traffic and have proven efficient in capturing linear trends and seasonality in traffic data[3,4]. Urban traffic data exhibits non-linear and complex temporal dynamics influenced by several factors, including road infrastructure, events, and weather conditions[5]. Deep learning approaches have proved superior in modeling non-linear and complex temporal dependencies. LSTM architectures have been widely applied for traffic prediction due to their gate structure, which helps in maintaining long-term dependencies[6–8]. Studies have found that deep learning LSTM-based models, coupled with external variables like meteorology and accidents, provide improved accuracy in traffic prediction tasks. Recent surveys suggest an increasing trend of applying hybrid and ensemble deep learning models to tackle non-linear and multi-scale urban traffic.

Despite significant advancements in deep learning, there are two primary challenges that it is facing—first, its effective implementation requires a substantial amount of data, and second, it is extremely challenging to interpret the model. Both problems make life difficult when dealing with city-level data, as it can be challenging to gather sufficient data. To overcome the limitations faced in implementing traditional deep learning models, scientists came up with hybrid models that would incorporate the benefits of statistical models and deep learning. Inspired by this, along with real-world issues, this research paper aims to propose a novel data-agnostic hybrid SARIMA-LSTM framework for urban traffic surveillance.

The objectives of this study include: The proposed methodology utilizes an innovative hybrid model which incorporates the statistical precision of Seasonal Autoregressive Integrated Moving Average (SARIMA) along with the deep learning capabilities of Long Short Term Memory (LSTM) to forecast urban traffic. The model can be applied to urban traffic data sets collected from various cities or even from varying platforms, including limited data sets, increasing its applicability to diverse urban settings. The hybrid model identifies linear relationships using SARIMA, which is useful for identifying seasonal and steady-state trends in traffic data, along with detecting nonlinear relationships using LSTM. Such a hybrid solution can increase efficiency and robustness compared to pure statistical or deep learning methods by itself. Performance evaluation using measures such as RMSE, MAE, and MAPE of actual or artificial urban traffic data sets proves the efficacy of the presented hybrid model. Finding suggests that the model suits well to modern smart city scenarios, where efficient traffic control and management become possible, resulting in reduced congestion. In addition, the method is relatively easy to implement and can facilitate real-time analysis and management decisions.

Literature Review: As for traffic forecast modeling approaches, they include traditional time series algorithms such as ARIMA and SARIMA. These methods are widely utilized since they are easily interpretable and capable of modeling regularities like seasonality or trend in traffic. Nonetheless, let's face it; such methods will be rather incapable when it comes to abrupt fluctuations or non-linear relationships in traffic. This is where deep learning made its breakthrough. For instance, LSTMs are characterized by a better performance at extracting the dependencies between the variables for a very long period of time. Moreover, other models such as GRU and sequence-to-sequence are now used in traffic prediction successfully[5-6]. The use of extra traffic information, weather or time-based features, makes such predictions even more accurate, especially in smart cities. Hybrid models like CNN-LSTM combine spatial and temporal data and therefore are considered superior models. Graph Neural Networks (GNNs) are able to model the traffic in road networks since they use graph structures as inputs. Among the best GNNs models are the ones that combine DCRNN and Graph Wave Net which are highly accurate in the presence of multiple sensors[6-10]. DCRNN and Graph Wave Net models have achieved high precision; however, they require sufficient sensor data input into the network[11-13]. Additionally, there is an increasing tendency towards developing hybrid systems by combining statistical and deep learning techniques, particularly focusing on residual learning architectures. Such approaches have proven superior in accuracy, considering their complementary characteristics. They are applicable in urban environments where the data volume is limited, yet the forecast should be accurate[14-15].

Finally, deep learning algorithms have become popular in analyzing road accidents. LSTM-based models can detect accident pattern trends and provide practical assistance in optimizing traffic control measures and establishing safety policy guidelines[16]. The present study proposes a hybrid model of ARIMA and Conv LSTM with attention mechanisms. It resulted in higher short-term traffic forecast accuracy[17].

Problem formulation

Nonlinear interaction of vehicles, variations in the traffic demand pattern and various other external elements like weather, accidents, and road conditions contribute to making the dynamics of an urban traffic system complex. Such elements give rise to chaotic behavior in traffic flows and high degrees of time dependency, thus making traffic prediction very difficult. As a result, a good amount of modeling is required for traffic prediction.

Let X_t denote the observed traffic volume at time t. The traffic series can be decomposed into three fundamental components:

$X_t = T_t + S_t + N_t$
where: T_t represents long-term trend, S_t denotes seasonality, N_t captures nonlinear dynamics and residual fluctuations.

The aim is to predict the value of X_{t+h} through explicit modeling of both linear and nonlinear temporal structures employing a hybrid framework.

Proposed hybrid sarima-LSTM framework

Time-Series Diagnostics: Before modeling, the stationary of the series is checked using the Augmented Dickey-Fuller test. Seasonality decomposition, along with ACF, PACF analysis, will help determine the SARIMA order[3-13].

Statistical Component (SARIMA): The SARIMA equation is expressed as:

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^d Y_t = \theta_Q(B^s)\theta_q(B)\epsilon_t$$

Where: B: Backward shift operator, $BY_t = Y_{t-1}$, $\Phi_P(B^s)$: Seasonal AR polynomial. $\phi_p(B)$: Non-seasonal AR polynomial. $\theta_Q(B^s)$: Seasonal MA polynomial. $\theta_q(B)$: Non-seasonal MA polynomial. vi. ϵ_t: White noise error term.

Model orders (p, d, q) (P,D,Q,s) are selected using: i. ACF/PACF analysis, ii. Information criteria: AIC and BIC for optimal parameter selection[4,14], iii. Residual diagnostics to ensure white noise behavior. This component captures linear seasonal and trend effects.

Deep Learning (LSTM): The residuals obtained from the SARIMA model ($(R_t = X_t - \hat{X}_t^{SARIMA}$) can be modeled using an LSTM model that can deal with non-linear time-series dependency and long-range dependency[6,7].

For learning non-linear time-series dependency, the sequence of residuals is analyzed by applying a Long Short-Term Memory (LSTM) model. The LSTM model is specifically trained for learning long-term dependencies in time series. The LSTM cell is governed by the following equations:

Forget Gate: It filters out the no relevant information and define as $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$

Input Gate: It determines the impact of new information and written as $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$
$\widetilde{C_t} = tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$

Cell State Update: It stores long term memory
$C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C_t}$

Output Gate: Generates the final prediction
$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$

$h_t = o_t \cdot tanh(C_t)$
Where: W, b and σ denote weight matrices, biases & the sigmoid activation function respectively.

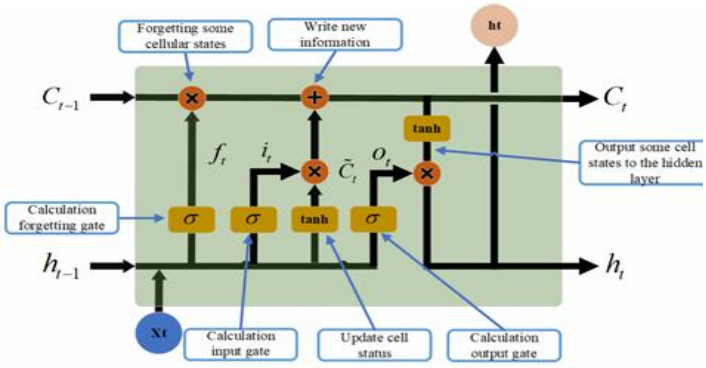


Figure-1: The Network Structure of LSTM at time t[8].

Hybrid Forecasting: Final traffic forecasting is obtained by combining the SARIMA and LSTM outputs:
$\hat{X}_t = \hat{X}_t^{SARIMA} + \hat{R}_t^{LSTM}$

This hybrid residual learning approach balances interpretability and nonlinear learning capability.

This hybrid strategy ensures that: i. SARIMA captures linear seasonal behavior, ii. LSTM captures nonlinear temporal dependencies.

Thus, the integrated model improves forecasting capability while maintaining interpretability.

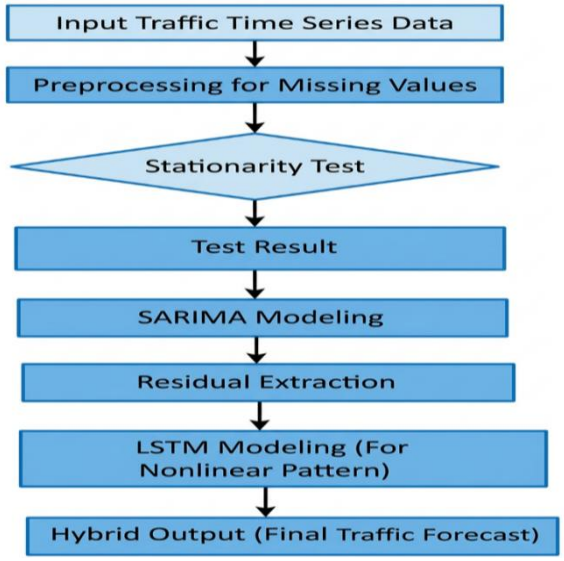


Figure-2: Proposed Hybrid SARIMA-LSTM Methodology Workflow.

Algorithmic framework: The procedure for this algorithm follows two stages. First, the SARIMA algorithm takes care of the linear components of the traffic data as well as the seasonality. This will give us a reliable baseline prediction. However, there are still some residual errors that we cannot predict using SARIMA. These are passed to the next phase, where they are fed into an LSTM model. The LSTM then focuses on the non-linear aspects that were overlooked by the previous model. We can just simply add the results from the two models.

Results and Discussion

Forecasting Performance Analysis: This is for the hybrid SARIMA-LSTM predictions of a traffic time series with both trend and seasonal components. It must be noted that the hybrid method is very accurate in capturing the seasonal component, while at the same time filtering out the volatile nature of the real traffic data. This comparison can be seen in Figure-2.

Quantitative Evaluation Metrics: The performance of the suggested hybrid Seasonal Autoregressive Integrated Moving Average-LSTM (SARIMA–LSTM) model may be measured by using the following time-series prediction evaluation metrics based on the discrepancy between actual values x_t and predicted values $\hat{x}_t$.

The Mean Absolute Error (MAE) measures the average magnitude of prediction errors:

$$MAE = \frac{1}{n}\sum_{t=1}^{n} | x_t - \hat{x}_t |$$

The Root Mean Square Error (RMSE) gives higher weight to large prediction errors:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_t - \hat{x}_t)^2}$$

The Mean Absolute Percentage Error (MAPE) expresses the error in percentage terms:

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{x_t - \hat{x}_t}{x_t}\right|$$

Lower values of these metrics indicate better forecasting accuracy and improved performance of the hybrid model.

Discussion: The main outputs obtained through the application of the hybrid framework (SARIMA+LSTM) demonstrate several important conclusions regarding the effectiveness of urban traffic prediction: i. The proposed approach adequately represents both linear and nonlinear characteristics of the time series concerning urban transportation. ii. The SARIMA model generates a predictable initial forecast by means of modeling seasonal and trend traffic flows. iii. The use of the LSTM network for finding nonlinear connections between residuals provides an opportunity to enhance the prediction accuracy. iv. The entire architecture proves to be highly adaptive to traffic variations, particularly when it comes to irregular traffic changes. v. The intelligent transportation systems may benefit

from the use of such an architecture, which combines both deep learning technologies and statistical models' interpretability. vi. The model may assist in traffic management, congestion avoidance, and intelligent transportation solutions development.

The above-listed outcomes of the research prove the advantages of using hybrid approaches in developing urban traffic forecasting models.

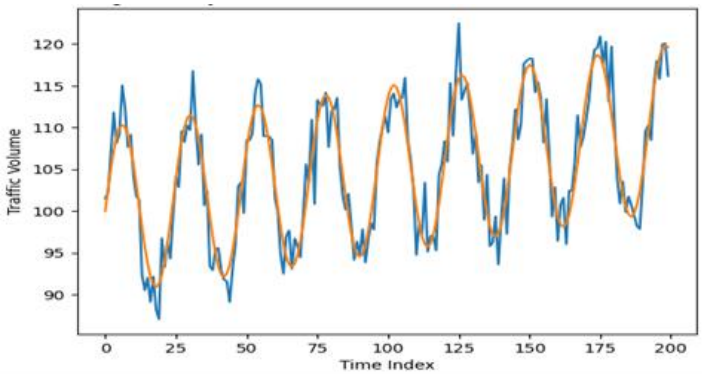


Figure-3: Presents a conceptual simulation to illustrate the hybrid forecasting behavior.

Limitations: The hybrid SARIMA-LSTM approach mainly emphasizes time series analysis without considering the spatial relationships of interconnected road networks. The hypothesis of partitioning traffic patterns into linear and nonlinear segments might not be appropriate when traffic experiences extreme volatility due to unexpected incidents or events. Being a purely theoretical model that can be applied to any type of dataset, it needs rigorous testing based on a large number of multi-city traffic datasets to establish its universal applicability. Furthermore, the optimization of LSTM parameters may lead to higher computational costs, and some exogenous variables like weather and policies are missing from the model.

Conclusion

This research paper describes the development of a new combined approach which uses both the SARIMA model and the LSTM network in a hybrid way for urban traffic forecasting. The advantage of this approach lies in the combination of the benefits from the usage of statistical and learning approaches by modeling the residuals in order to improve forecasting stability, flexibility, and efficiency. This technique is particularly relevant for smart cities and other urban areas with limited data. Future studies on the topic must consider multiple sources of data and spatial graph analysis.

References

1. Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43, 3-19.

2. Abdel-Aty, M., Zheng, O., Wu, Y., Abdelraouf, A., Rim, H., & Li, P. (2023). Real-time big data analytics and proactive traffic safety management visualization system. *Journal of transportation engineering, Part A: Systems*, *149*(8), 04023064.

3. Box, G. and Jenkins, G. M. (1976). Analysis: Forecasting and control. *San francisco*, *10*.

4. Williams, B. M., & Hoel, L. A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of transportation engineering*, *129*(6), 664-672.

5. Ahmed, S. A., & Cook, A. R. (1980). Time series models for freeway incident detection. *Transportation Engineering Journal of ASCE*, *106*(6), 731-745.

6. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.

7. Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation research part C: emerging technologies*, 54, 187-197.

8. Yao, E., Zhang, L., Li, X., & Yun, X. (2023). Traffic forecasting of back servers based on ARIMA-LSTM-CF hybrid model. *International Journal of Computational Intelligence Systems*, 16(1), 65.

9. Kong, F., Li, J., Jiang, B., & Song, H. (2019). Short-term traffic flow prediction in smart multimedia system for Internet of Vehicles based on deep belief network. Future Generation Computer Systems, 93, 460-472.

10. Tao, X., Cheng, L., Zhang, R., Chan, W. K., Chao, H., & Qin, J. (2023). Towards green innovation in smart cities: Leveraging traffic flow prediction with machine learning algorithms for sustainable transportation systems. *Sustainability*, 16(1), 251.

11. Mystakidis, A., Koukaras, P., & Tjortjis, C. (2025). Advances in traffic congestion prediction: An overview of emerging techniques and methods. *Smart Cities*, *8*(1), 25.

12. Leinonen, M., Al-Tachmeesschi, A., Turkmen, B., Atashi, N., & Ruotsalainen, L. (2025). Long short term memory based traffic prediction using multi-source data. *International Journal of Intelligent Transportation Systems Research*, 23(1), 354-371.

13. Hyndman, R. J., & Athanasopoulos, G. (2014). *Forecasting: principles and practice* (Vol. 2). Melbourne, Australia: OTexts.

14. Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716-723.

15. Mead, M. A. (2022). Hybrid CNN and LSTM model (HCLM) for short-term traffic volume prediction.

International Journal of Intelligent Computing and Information Sciences, 22(4), 51-61.

16. Angadi, V. S. and Halyal, S. (2024). Forecasting road accidents using deep learning approach: policies to improve road safety. *Journal of Soft Computing in Civil Engineering*, 8(4), 27-53.

17. Sattarzadeh, A. R., Kutadinata, R. J., Pathirana, P. N., & Huynh, V. T. (2025). A novel hybrid deep learning model with ARIMA Conv-LSTM networks and shuffle attention layer for short-term traffic flow prediction. *Transportmetrica A: Transport Science*, 21(1), 2236724.