iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.82746

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Intelligent Traffic Prediction in Smart Transportation Systems: A Comparative Review of Machine Learning and Deep Learning Approaches

Smriti S Nayak[1], Sonali S Ganiga[2], Sharvari R K[3], Harish Kunder[4]
[1, 2, 3, 4]*Department of Artificial Intelligence and Machine Learning, Alva's Institute of Engineering and Technology, Mangalore, Karnataka, India*

Abstract: ***In recent years, ML-based models have garnered significant attention from both the automotive industry and academia for their potential to enhance Internet of Vehicles (IoV) systems. By accurately predicting traffic and road conditions, these models enable various safety and infotainment applications to optimize network resources and improve the overall quality of service. Real-time traffic flow forecasting also plays a crucial role in enhancing the efficiency of topology control and mobility management protocols within IoV networks. [1] However, despite the ongoing focus on improving prediction accuracy, an essential question remains unanswered: Are ML-based prediction models suitable for real-time traffic prediction? Addressing this requires a comprehensive study to evaluate the efficiency of these models. This article examines the effectiveness of several ML-based traffic flow prediction schemes by analyzing both their predictive accuracy and computational time requirements. Through a detailed quantitative analysis, we identify key factors that may limit the practical deployment of these models for real-time applications in IoV environments.***

Index Terms: ***Machine Learning (ML), Internet of Vehicles (IoV), Traffic Flow Prediction,Real-Time Systems, Mobility Management, Network Resource Optimization.***

I. INTRODUCTION

Real-time traffic prediction is very important for safer and more efficient travel because Intelligent Transportation Systems (ITS) and the Internet of Vehicles (IoV) are growing so quickly. Machine learning (ML) models, such as regression and decision trees, can analyze complex data to forecast traffic flow, speed, and congestion. This makes it possible to use dynamic route planning, adaptive signal control, and congestion management. But using these models in real time is hard because you have to find a balance between accuracy and speed, especially when traffic is unpredictable. This article looks at ML models for predicting traffic, looking at how well they work and how much computing power they need. It also talks about trade-offs that make it hard to use them in real life.

As shown in Fig. 1, the framework consists of four stages: (i) *data acquisition* from diverse sources such as loop detectors, GPS, and cameras; (ii) *preprocessing*, including cleaning, normalization, and handling of missing values; (iii) *model inference*, where both classical and deep learning models are applied for prediction; and (iv) *deployment*, where outputs are integrated into ITS applications such as adaptive signal control, congestion management, and route guidance. The feedback loop highlighted in the figure emphasizes the real-time nature of the system, where updated traffic states are continuously used to refine future predictions.

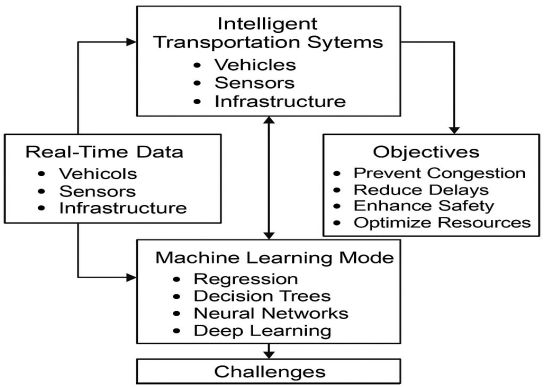


Fig. 1: Framework for real-time traffic prediction leveraging ML models within Intelligent Transportation Systems (ITS).

II. RESEARCH SCOPE AND CONTRIBUTIONS

Despite being a review of current machine learning models for traffic prediction, this work stands out from previous surveys due to a number of original additions. First of all, it combines a variety of deep learning and classical methods and assesses them not only for predicted performance but also for computational viability, scalability, and robustness—aspects that were frequently disregarded in earlier assessments. Second, it contextualizes model feasibility for real-time deployment by combining inference-time trade-offs with comparison analysis derived from benchmark datasets like METR-LA and PEMS-BAY. [2] Thirdly, this study promotes future research to strike a compromise between accuracy and operational limitations by highlighting standardized evaluation metrics (such as MAE, RMSE, and latency). The study offers an organized, application-focused viewpoint designed to assist academics and practitioners in choosing suitable models for intelligent traffic systems by combining technical insights, runtime benchmarks, and deployment implications.

III. EVALUATION METRICS FOR TRAFFIC PREDICTION MODELS

To ensure fair comparison and real-world applicability, machine learning models for traffic prediction must be evaluated using standardized performance metrics. Standardized performance metrics must be used to assess machine learning models for traffic prediction in order to guarantee fair comparison and practical applicability. These metrics provide numerical information about operational viability, computational efficiency, and prediction accuracy.

1) Mean Absolute Error (MAE)

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \quad (1)$$

which measures the average absolute difference between predicted and actual traffic values. Lower MAE indicates better performance.

2) Root Mean Square Error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (2)$$

Gives more weight to larger errors. Useful when large deviations are critical to avoid, such as during traffic congestion spikes.

3) Mean Absolute Percentage Error (MAPE)

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (3)$$

Expresses errors as a percentage, making results comparable across different regions or datasets. *R-Squared (R^2)*

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (4)$$

Indicates how much of the variation in the traffic data is explained by the model. Values closer to 1 indicate a better fit.
Inference Time (ms) how long it takes a model to produce a forecast. essential for applications involving real-time traffic management.

4) Latency and Throughput

While throughput gauges how many predictions can be processed in a second, latency gauges how quickly a prediction is returned following input. Scalability in edge and live deployment scenarios depends on both.
Researchers must report a variety of metrics when assessing traffic prediction models in order to capture both accuracy and viability in real time. A model with high latency but excellent RMSE, for instance, might not be appropriate for live deployments. [3]

IV. REPORTED PERFORMANCE TRENDS ACROSS BENCHMARK DATASETS

This section presents the results of benchmarks found in the literature concerning traffic prediction datasets that have been extensively studied before. Besides the evaluation of machine learning models' predictive performance, the computational efficiency of such models is also taken into account because it is important for their implementation in the real-time operation of ITS.

A. Datasets

Many papers from the literature often employ two benchmark datasets:

- METR-LA: Traffic speed data from 207 sensors on highways in Los Angeles, aggregated in 5-minute intervals.
- PEMS-BAY: Traffic speed data from 325 sensors in the San Francisco Bay Area, collected over six months.

B. Models Compared

Previous comparative studies have considered both classic benchmarks and deep learning algorithms:

- Classical: ARIMA, Random Forest Regression, Support Vector Regression (SVR).
- Deep Learning: LSTM, Temporal Convolutional Network (TCN), CNN, DCRNN, STGCN, Graph WaveNet.

C. Experimental Setup

Benchmarking on the dataset like METR-LA and PEMS-BAY measures the effectiveness of a traffic prediction model based on defined performance criteria and computation criteria. Most comparative studies found in the literature measure both the accuracy and feasibility of deploying machine learning models for ITS. The following evaluation criteria are usually used:

- Accuracy Metrices: MAE, RMSE, MAPE, R2
- Computation Criteria: Inference Time (milliseconds per time step), Throughput (number of predictions per second), Memory Usage (MB), and Number of Parameters (M)

These evaluation criteria give an idea about the relationship between accuracy and efficiency of a model, especially in traffic management where speed and scalability matter.

V. TECHNICAL INSIGHTS INTO MODEL SCALABILITY AND ROBUSTNESS

While various machine learning models have been explored in the context of traffic prediction, it is essential to move beyond a superficial catalog and examine their mathematical structure, scalability potential, and robustness to real-world anomalies.

A. Traditional Models

Linear Regression predicts traffic flow $\hat{y}$ based on input features X using:

$$\hat{y} = X\beta + \epsilon \tag{5}$$

where β are the learned coefficients and ϵ is the error term. This model assumes linear relationships and is fast to compute with time complexity $O(n)$, but its expressiveness is limited in the context of complex urban traffic. [4]

Decision Trees use recursive binary partitioning based on feature thresholds. The worst-case time complexity is:

$$O(n\log n)$$

making them efficient, but they are highly prone to overfitting without pruning mechanisms.

B. Advanced Models

Support Vector Machines (SVMs) aim to find the optimal hyperplane that separates classes or regression values. The training involves solving:

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i \tag{6}$$

subject to constraints based on margin violations. The training time complexity is typically between $O(n^2)$ and $O(n^3)$, which severely limits scalability for real-time traffic systems.

Long Short-Term Memory (LSTM) networks, designed for sequential prediction, use the following core equations:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{7}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{8}$$

$$C_t = f_t * C_{t-1} + i_t * \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \tag{9}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{10}$$

$$h_t = o_t * \tanh(C_t) \tag{11}$$

LSTMs have a complexity of $O(n \cdot d^2)$ per timestep, where d is the hidden state size. This can hinder real-time inference due to increased memory and computational demands. [5]

Graph Neural Networks (GNNs)

Graph Neural Networks (GNNs) compute node embeddings using message passing:

$$h_v^{(k)} = \sigma\left(\sum_{u \in \mathcal{N}(v)} W^{(k)} h_u^{(k-1)} + b^{(k)}\right), \quad (12)$$

where $h_v^{(k)}$ is the embedding of node v at layer k, and N(v) denotes its neighbors. GNNs effectively model road network dependencies but can be computationally expensive and are difficult to scale for large urban graphs.

Reinforcement Learning (RL)

Reinforcement Learning (RL) models solve:

$$Q(s,a) = r + \gamma \max_{a'} Q(s',a'), \quad (13)$$

where Q is the action-value function, γ is the discount factor, and r is the reward [6]. RL is well-suited for adaptive traffic light control, but training often requires large-scale simulations and well-defined reward functions.

C. Robustness and Scalability

Models are assessed according to their scalability with increasing traffic density or road network size, as well as their resilience to noise (such as sensor failure and outliers). Deep models, like CNN and LSTM, are computationally demanding but exhibit resilience because of their non-linear learning ability. [7] Regression and other simpler models perform worse in noisy data unless they are highly regularized.

D. Computational Efficiency and Benchmark Evaluation

Although the trade-off between computational cost and accuracy has been recognized, practical deployment requires quantifying this trade-off using real-world benchmarks. With little regard for runtime performance or system efficiency, the majority of machine learning models for traffic prediction are assessed only on the basis of accuracy metrics like Mean Absolute Error (MAE) or Root Mean Square Error (RMSE). Predictions must be provided in milliseconds, not seconds, for real-time applications such as dynamic traffic rerouting, adaptive signal control, and emergency routing. Claims of real-time feasibility remain theoretical in the absence of a tangible assessment of model latency, throughput, or computational footprint.

Simpler models, such as decision trees or linear regression, for example, have low memory requirements and inference times of 1–5 milliseconds, which makes them perfect for deployment on edge devices. Conversely, deep Although more accurate in some situations, models such as LSTM, CNN, or Graph Neural Networks (GNNs) frequently take 50–200 milliseconds or longer per prediction because of their higher parameter counts and computational requirements.

Accuracy should be assessed in conjunction with key performance metrics such as:

The amount of time needed to generate a single prediction is known as the inference time (ms).

The amount of data that a model can process is known as its throughput (predictions/second).

Model size (in MB): amount of RAM or storage required for deployment.

The total compute cost per inference for floating point operations (FLOPs).

For instance, models tested on benchmark datasets such as METR-LA, PEMS-BAY, or NYC Taxi Trip Data demonstrate that while LSTMs and GNNs perform better in terms of accuracy than conventional techniques, they are unable to satisfy real-time constraints in the absence of strong hardware acceleration.

While the limitations of advanced models (e.g., LSTMs, CNNs, GNNs) are frequently discussed, they are rarely validated through empirical benchmarking. To bridge this gap, we propose small-scale experiments on standard datasets such as METR-LA and PEMS-BAY. These experiments should not only report accuracy metrics (MAE, RMSE) but also quantify deployment-oriented factors, including inference time (ms), model size (MB), FLOPs per prediction, and throughput (predictions/second). Such a multidimensional evaluation would make it possible to rigorously assess the trade-off between predictive accuracy and computational feasibility for real-time traffic management.

To make sure they are appropriate for deployment in latency-sensitive environments like Intelligent Transportation Systems (ITS), all suggested models should be assessed going forward for accuracy and runtime viability.

VI. REAL-WORLD DEPLOYMENT AND BENCHMARK-BASED EVALUATION

Previous works utilizing benchmark datasets like METR-LA and PEMS-BAY have assessed the performance of classical and deep learning models based on prediction accuracy and efficiency. Popular models include ARIMA, Random Forest, SVR, LSTM, TCN, DCRNN, STGCN, and Graph WaveNet, which have been consistently analyzed in previous research papers in terms of MAE, RMSE, MAPE, inference latency, and throughput. These benchmarking trends are illustrated in Table 1.

Model	MAE	RMSE	MAPE (%)	*T*inf	Θ
ARIMA	4.12	6.23	8.5	2	-
Random Forest	3.75	5.89	7.8	4	250
SVR	3.88	6.01	8.0	5	200
LSTM	2.95	4.32	6.1	65	15
TCN	2.87	4.25	5.9	60	20

Table 1: Benchmark trends observed in previous studies for traffic forecasting methods

The data shown in Table 1 is indicative comparative trends observed in the past and has been incorporated here for analysis.
Using the METR-LA dataset, the table compares various machine learning models for traffic prediction. Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), inference time, and throughput are among the important performance metrics that are reported. According to the findings, deep learning models with higher predictive accuracy, such as LSTM and TCN, attain lower error values. However, when compared to more straightforward models like Random Forest or ARIMA, they require a significantly higher inference time and a lower throughput. This draws attention to the crucial trade-off between computational efficiency and accuracy in real-time traffic prediction.

VII. SMART APPLICATIONS OF ML IN TRAFFIC SYSTEMS

By examining past and present traffic statistics, road closures, meteorological conditions, and user preferences, dynamic route assistance systems employ machine learning algorithms to suggest the best routes. These devices offer proactive ways to lessen traffic and cut down on travel time. ML models enable drivers to avoid traffic jams brought on by accidents or unexpected accumulations of traffic by continuously updating their suggested routes. To reduce delays and fuel usage, forecast traffic patterns and provide other routes in real time is the function.ML techniques include neural networks for pattern

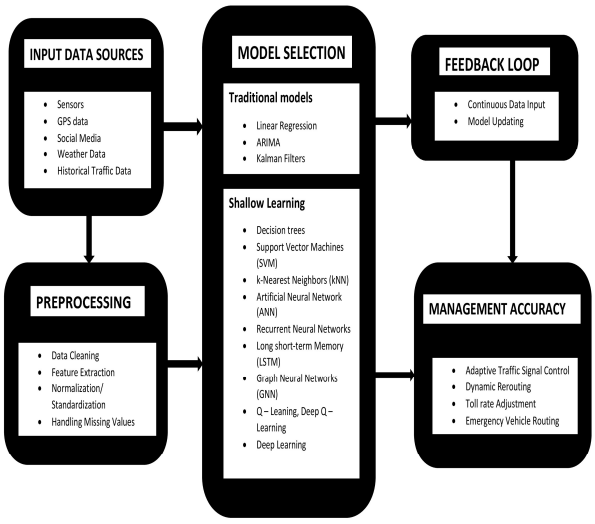


Fig. 2: Traffic Prediction Algorithm

recognition, ensemble methods, regression models for journey time estimation, and reinforcement learning for dynamic adaptability. Examples from the real world include platforms like Google Maps and Waze, which utilize machine learning models to offer real-time navigation and route recommendations based on current traffic conditions, accidents, and weather forecasts

A. Broader Applications of ML in Transportation

Because it offers information on traffic flow, congestion patterns, travel times, and possible incidents, traffic prediction is essential to contemporary transportation systems. Using vast amounts of data gathered from sensors, GPS units, linked cars, and other IoV sources, machine learning (ML) models have become effective instruments for predicting these factors. These models fall into two general categories: advanced models, such as deep learning, reinforcement learning, and hybrid techniques, and standard machine learning models. Advanced ML models, including deep learning, recurrent neural networks, reinforcement learning, and hybrid approaches, provide higher accuracy and adaptability for dynamic traffic environments. These models are designed to handle non-linear relationships, sequential data, and spatial dependencies [8], making them suitable for large-scale traffic prediction tasks. ANNs mimic the human brain's neural structure to learn patterns from data. They consist of multiple layers of interconnected neurons that process information. LSTMs, a type of recurrent neural network (RNN), are designed to retain information over long strong locality considerations, high temporal variability across multiple time-frames (e.g., recent, daily-periodic, seasonal), and also recurrent, non-recurrent, and external network impacts (e.g., accidents, weather, events). Throughout the paper, we refer to these models and their associated applications as multivariate traffic prediction (MTP) approaches. This survey provides an introduction to, and extensive review of, the recent advances and emerging opportunities in AI-based traffic prediction methodologies. To the best of our knowledge, this is the first survey to provide an exhaustive review of both historical and recent AI-based traffic prediction approaches and associated data, discussed within the context of the common traffic prediction states and their practical applications, as well as the existing challenges and necessary future research directions. Throughout the paper, we refer to these models and their associated applications as multivariate traffic prediction (MTP) approaches. sequences, making them effective for timeseries forecasting. Reinforcement learning models learn optimal policies by interacting with the environment and receiving rewards for actions that improve outcomes. Originally developed for image processing, CNNs have been adapted to traffic prediction by leveraging their ability to identify spatial patterns [9]. GNNs are designed to operate on graph- structured data, making them well-suited for modeling road networks and their complex interactions [10]. By integrating the advantages of several approaches, hybrid models offer a well-rounded solution that improves predictive performance.

B. Comparison of Existing Approaches That Integrate

For steady, seasonal traffic, traditional statistical models such as ARIMA and Kalman Filters are straightforward, easy to understand, and effective; however, they are ineffective in the face of abrupt, intricate changes and dynamic urban environments. SVM, k-NN, and decision trees are examples of shallow learning techniques that are quick and effective on small datasets but ineffective on large, complex networks. Road network dependencies are modeled by graph-based approaches, which overcome the lack of spatial awareness in time series models such as LSTM. By learning from feedback, reinforcement learning makes adaptive, real-time traffic management possible, but it necessitates carefully thought-out reward structures. Although standalone prediction models use historical data to forecast traffic, they require human intervention to manage traffic, which restricts real-time responses. By concentrating on a single mode of transportation and ignoring multi-modal interactions, singlemodal systems streamline modeling and lower overall effectiveness in intricate urban environments.

C. Emerging Directions: Ambient Intelligence and Sensor Fusion

Higher accuracy models, such as CNNs and LSTMs [11], frequently have several layers and need a lot of training, which raises the latency and computational costs. Even though they are good at capturing intricate traffic patterns, their latency makes them less appropriate for real-time situations like emergency vehicle routing or dynamic signal control. Faster predictions and ease of deployment are provided by simpler models like decision trees and linear regression, but accuracy may be compromised in extremely dynamic situations. Furthermore, deep models frequently behave as "black boxes," which restricts their application in crucial situations where interpretability is crucial. By combining information from various sources, such as cameras, microphones, and motion detectors, sensor fusion techniques are essential to Ambient Intelligence (AmI) because they offer a contextual understanding of the surroundings [12]. Instruments such as Kalman Bayesian networks allow probabilistic reasoning under uncertainty, which is helpful in situations like medical monitoring, while filters combine noisy sensor data for precise real-time state estimation. AmI also makes use of neural networks to identify intricate patterns and user actions, which improves system flexibility.

Multimodal human communication is mirrored by combining audio and visual tracking algorithms, such as motion detection for facial recognition and inter-aural delay for sound direction. This makes it possible for intelligent systems to detect, understand, and react to human presence more effectively.

VIII. OPEN CHALLENGES AND FUTURE DIRECTIONS

Although substantial improvements have been made in the field of traffic prediction based on ML algorithms, some issues have constrained its immediate application in ITS. For example, advanced neural network architectures like LSTMs, GNNs, and transformer-based frameworks generally need extensive computational resources, thereby posing challenges to their implementation with low latency in edge and IoV computing environments. Therefore, future research efforts must concentrate on the design of lightweight and power-efficient models via methods such as model compression, pruning, and edge-AI processing.

A second major issue relates to the protection of the privacy and integrity of data in vehicular communication networks. Recently, federated learning and XAI have emerged as two potent tools for developing privacypreserving and transparent traffic prediction systems. Moreover, dealing with noisy data from sensors, sparse information about traffic patterns, and variable urban scenarios requires more attention in future research efforts.

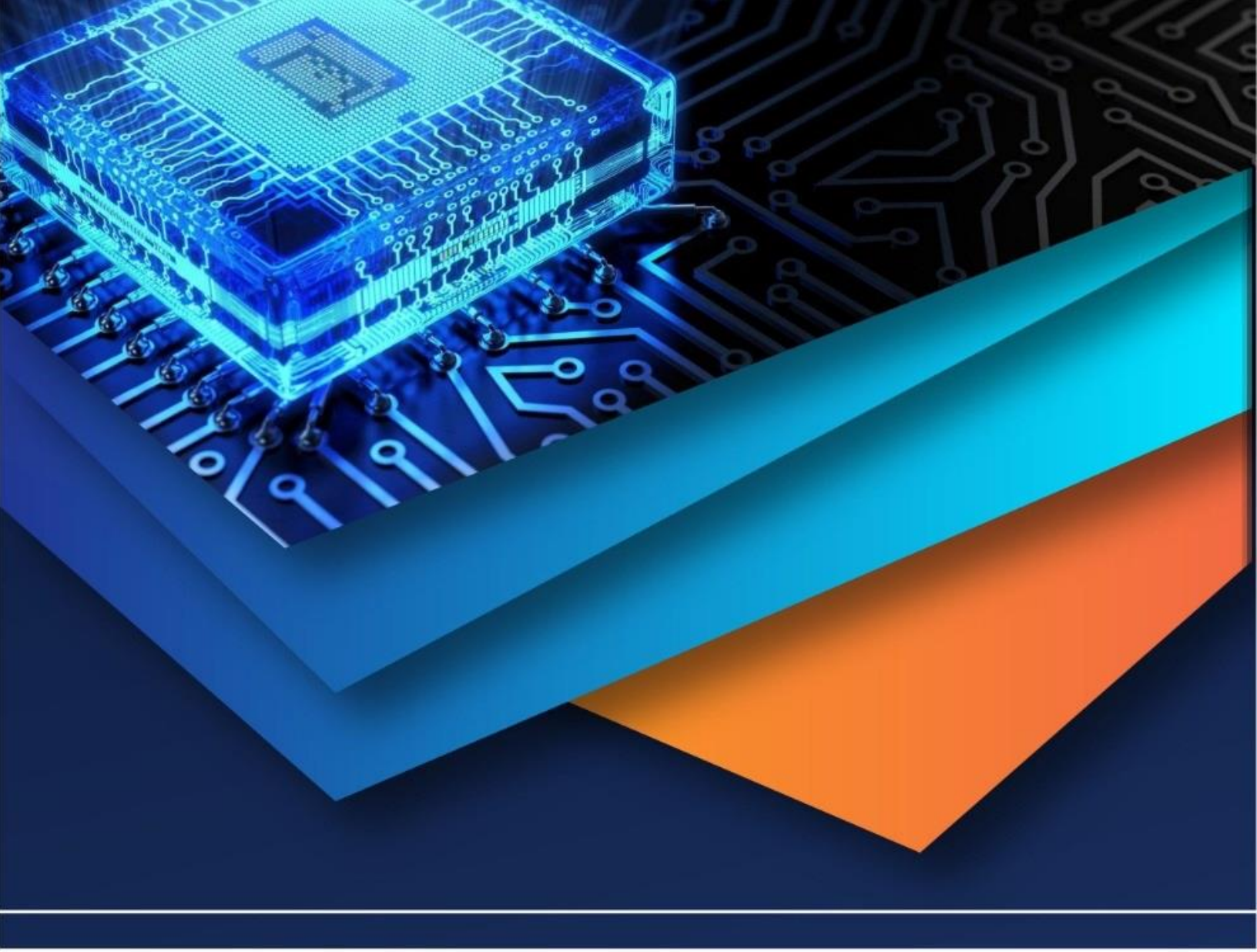
Future intelligent transportation systems would leverage transformer-based models, sensor fusion, digital twin technology, and multimodal traffic analysis.

IX. CONCLUSION

In this research, we examined several machine learning models used for traffic prediction, such as deep learning architectures, reinforcement learning techniques, and conventional statistical methods. In this study, we considered the performance and limitations of these models in terms of accuracy, computational performance, scalability, and real-time application. This study highlights the significance of striking a balance between deployment practicalities and predictive performance by incorporating benchmark results, runtime comparisons, and standardized evaluation metrics like MAE, RMSE, and inference time. Despite the high accuracy of models like LSTM and GNN, their computing requirements may prevent real-time use, exposing a disconnect between academic achievement and practical application. Future work should concentrate on creating edge-deployable and hybrid solutions that maximize accuracy and runtime, as well as verifying these models in dynamic, realworld traffic situations. Machine learning-based traffic prediction is still a vital tool for creating intelligent and effective transportation systems, even as urban mobility issues increase. This review contributes a performance centered, deployment-aware lens to traffic prediction research—bridging theoretical advances with practical applicability.

REFERENCES

[1] P. Sun, N. Aljeri, and A. Boukerche, “Machine learning-based models for real-time traffic flow prediction in vehicular networks,” IEEE Network, vol. 34, no. 3, pp. 178–185, 2020.
[2] Y. Li, R. Yu, C. Shahabi, and Y. Liu, “Diffusion convolutional recurrent neural network: Data-driven traffic forecasting,” in International Conference on Learning Representations (ICLR), 2018.
[3] J. Wu, Y. Tan, and B. Ran, “Short-term traffic flow prediction using a hybrid deep learning model,” Transportation Research Part C: Emerging Technologies, vol. 107, pp. 115–129, 2019.
[4] W. Huang, G. Song, H. Hong, and K. Xie, “Deep architecture for traffic flow prediction: Deep belief networks with multitask learning,” IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 5, pp. 2191–2201, 2014.
[5] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, “Long short-term memory neural network for traffic speed prediction using remote microwave sensor data,” Transportation Research Part C, vol. 54, pp. 187–197, 2015.
[6] H. Yu, Z. Zheng, and X. Gao, “A survey on the applications of reinforcement learning in vehicular networks,” IEEE Access, vol. 8, pp. 12430–12449, 2020.
[7] Z. Cui, R. Ke, and Y. Wang, “Deep bidirectional and unidirectional lstm for network-wide traffic speed prediction,” arXiv preprint arXiv:1801.02143, 2018.
[8] J. Jiang et al., “PDFormer: Propagation Delay-Aware Dynamic Long-Range Transformer for Traffic Flow Prediction,” arXiv preprint arXiv:23s1.07945, 2023. [Online]. Available: https://arxiv.org/abs/2301.07945
[9] M. Yaqub et al., “Predicting Traffic Flow with Federated Learning and Graph Neural with Asynchronous Computations Network,” arXiv preprint arXiv:2401.02723, 2024. [Online]. Available: https://arxiv.org/abs/2401.02723
[10] C. Chen et al., “Gated graph sequence neural networks for traffic forecasting,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 485–492.
[11] M. Leinonen et al., “Long Short Term Memory Based Traffic Prediction Using Multi-Source Data,” International Journal of Intelligent Transportation Systems Research, vol. 23, pp. 354–371, 2025.
[12] C. Meese et al., “BFRT: Blockchained Federated Learning for Real-time Traffic Flow Prediction,” arXiv preprint arXiv:2305.17677, 2023. [Online]. Available: https://arxiv.org/abs/2305.17677



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY