IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, April 2026
Impact Factor: 8.2

Traffic Flow Prediction Using Machine Learning and Deep Learning Techniques: A Comprehensive Study

Jayashree Pasalkar[1], Ayush Thorat[2], Vipul Chavan[3]
[1]Professor, Department of Information Technology,
AISSMS Institute of Information Technology, Pune, India
[2],[3]Student, Department of Information Technology,
AISSMS Institute of Information Technology, Pune, India

Abstract: *Urbanization and the rapid growth in vehicle usage have intensified traffic congestion, posing significant challenges for modern transportation systems. Traffic flow prediction is essential for improving the efficiency of Intelligent Transportation Systems (ITS) through better traffic management and route planning. This study analyzes and compares traffic prediction models based on machine learning and deep learning techniques, with a focus on a hybrid LSTM–Random Forest approach for improved accuracy under dynamic conditions. Traditional models such as ARIMA are evaluated alongside machine learning methods and advanced deep learning architectures for real-time prediction. The results show that hybrid models outperform individual approaches in terms of accuracy and adaptability. The study also highlights current challenges and explores future directions, including the use of AutoML for developing scalable and efficient traffic prediction systems..*

Keywords: Traffic Prediction, Machine Learning, Deep Learning, LSTM, Intelligent Transportation System (ITS), Traffic Flow

1. INTRODUCTION

The effects of traffic congestion have been widely observed in urban areas where there is a continued increase in population, urbanization, and the number of vehicles. Such a situation has been attributed to many problems, including long journey times, increased fuel consumption, and pollution of the environment. In this regard, traffic control and management have become a necessary requirement for most urban areas.

Traffic flow prediction becomes an essential aspect within the intelligent transport system (ITS) as it facilitates effective traffic control and makes decisions concerning traffic flow easier. By combining historical patterns with real-time inputs, traffic prediction models can estimate future conditions and support more efficient route planning.

Traffic prediction methods based on statistical modeling have been common in the past. Examples include ARIMA and linear regression methods. These traditional approaches have a disadvantage of failing to capture the complexity and nonlinearity of traffic flow data. Traffic prediction is complicated in the sense that it is determined by many variables, such as weather, road situations, accidents, and special occasions.

As a result of the development of AI technology, ML methods have been frequently used for predicting traffic data. ML methods like Support Vector Machines (SVMs), K-Nearest Neighbor (KNN), Decision Trees, and Random Forest (RF) can learn from large volumes of information and offer enhanced results in terms of traffic prediction. These ML models are able to deal with non-linear relations and can be adjusted to any traffic situations.

Recently, DL methods have received considerable attention in traffic flow forecasting because of their capability to automatically learn complicated characteristics from large-scale data. For example, the Deep Learning models of LSTM, RNN, CNN, and GNN exhibit excellent performance when it comes to analyzing temporal and spatial

DOI: 10.48175/IJARSCT-33635

dependencies of traffic information. Specifically, the LSTM model performs exceptionally well while working with time-series data.

Additionally, hybrid algorithms based on machine learning and deep learning have been developed to increase the accuracy and effectiveness of predictions. Such algorithms incorporate the benefits of both methods to achieve better results in practice. Besides, modern developments such as AutoML and IoT data collection systems can be used to create more intelligent and adaptable traffic flow prediction algorithms.

This paper aims to analyze traffic flow prediction using machine learning and deep learning algorithms. It provides an analysis of existing models, assesses their effectiveness, and identifies the strengths and weaknesses. Moreover, a hybrid algorithm that improves prediction accuracy and highlights potential areas for future development is discussed.

II. LITERATURE SURVEY

Traffic flow prediction has attracted significant research attention due to its potential to improve transportation efficiency and reduce congestion. Over time, a wide range of approaches has been developed, spanning from traditional statistical methods to advanced machine learning and deep learning techniques.

A. Traditional Statistical Models

Early research in traffic flow forecasting primarily relied on statistical methods such as ARIMA, linear regression, and Kalman filtering. These approaches were effective for simple and structured datasets due to their mathematical simplicity and ease of implementation. However, real-world traffic data is highly dynamic and influenced by multiple external factors, including weather conditions, road incidents, and varying traffic patterns.

B. Machine Learning-based Methods

To address the limitations of traditional statistical models, machine learning techniques have been widely adopted for traffic flow prediction. Models such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Decision Trees, and Random Forest (RF) are capable of learning complex patterns from large datasets.

In traffic prediction tasks, SVM performs well with high-dimensional data; however, its effectiveness depends heavily on proper parameter tuning, which can limit its practical application. KNN is simple and effective for smaller datasets, but its computational cost increases significantly as the dataset size grows. Random Forest improves prediction accuracy through ensemble learning by combining multiple decision trees, although it still struggles to capture temporal dependencies in sequential data.

Overall, while machine learning models can effectively model nonlinear relationships, they have limitations in handling time-dependent patterns in traffic data.

C. Deep Learning Techniques

Deep learning approaches have significantly improved traffic prediction by enabling effective pattern recognition in large-scale and complex traffic data. Techniques such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and Graph Neural Networks (GNN) are widely used in this domain.

Among these, LSTM networks are particularly well-suited for traffic forecasting due to their ability to capture long-term temporal dependencies in sequential data. CNN models are useful for extracting spatial features when traffic data is represented in grid-based formats. In addition, Spatio-Temporal Graph Convolutional Networks (STGCN) further enhance prediction performance by simultaneously modeling spatial and temporal relationships within traffic networks.

Overall, deep learning methods demonstrate superior performance compared to traditional and basic machine learning approaches, especially in handling complex traffic patterns.

D. Hybrid Models

Hybrid approaches combine multiple algorithms to achieve better prediction performance. For example, integrating Random Forest with LSTM enables the model to capture both feature-based relationships and temporal dependencies simultaneously. Similarly, CNN–LSTM architectures leverage both spatial and temporal features, making them well-suited for complex traffic data.

Such approaches are generally more accurate, robust, and adaptable compared to single-model techniques, especially in dynamic traffic conditions.

E. AutoML and Innovative Methods

The emergence of Automated Machine Learning (AutoML) has simplified the development of predictive models by automating tasks such as model selection, hyperparameter tuning, and feature engineering. This reduces manual effort while improving overall model performance and efficiency.

In addition, the integration of Internet of Things (IoT) devices and smart sensors enables real-time data collection, which enhances the accuracy and reliability of traffic prediction systems.

TABLE I: COMPARATIVE ANALYSIS OF TRAFFIC PREDICTION TECHNIQUES

Sr No.	Author	Technique Used	Dataset	Accuracy	Key Findings
1	Deekshetha et al.	Regression ML	Historical traffic data	Moderate	Predicts hourly traffic using past data
2	Poonia et al.	LSTM	Real-time dataset	High	Captures temporal dependencies effectively
3	Khatiriolyaee et al.	AutoML + DL	Multi-source data	High	Automates model optimization
4	Yu et al.	STGCN	Traffic network graph	Very High	Captures spatial + temporal relations
5	Abduljabbar et al.	BiLSTM	Sensor-based data	Very High	Improves prediction using spatial inputs
6	Axcellin et al.	RF + LSTM	Traffic dataset	High	Hybrid model improves performance
7	Razali et al.	CNN + LSTM	IoT-based data	High	Effective for smart city applications
8	Navarro et al.	MLP, RNN	Intersection data	High	Useful for smart traffic signals
9	Kashyap et al.	DL models	Multiple datasets	High	DL outperforms traditional models
10	Jankovic et al.	Supervised ML	Big data traffic	Moderate	Feature selection improves prediction

F. Literature Review Summary

Based on existing studies, traditional statistical models are simple and easy to implement, but they are not well-suited for capturing the complex and dynamic nature of traffic data. Machine learning techniques improve prediction performance by modeling nonlinear relationships; however, they still face limitations in handling temporal dependencies effectively.

In contrast, deep learning approaches provide better results as they can capture both spatial and temporal patterns in traffic data. Furthermore, hybrid models, which combine multiple techniques, offer improved accuracy and robustness by leveraging the strengths of different methods.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, April 2026
Impact Factor: 8.2

Recent advancements, such as AutoML and IoT-based data collection, are expected to play a significant role in the development of more efficient and scalable traffic prediction systems.

III. PROBLEM STATEMENT

Traffic flow prediction remains a challenging task due to the dynamic, nonlinear, and time-varying nature of traffic data. Traditional statistical methods are often unable to handle unexpected events such as weather changes, road accidents, and peak-hour congestion. Although machine learning techniques have improved prediction accuracy, they still face limitations in capturing temporal dependencies and real-time variations.

Furthermore, many existing models struggle to effectively integrate both spatial and temporal features, which reduces their overall performance and increases computational complexity. Therefore, there is a need for a robust and efficient traffic prediction system capable of processing complex data and delivering accurate real-time forecasts.

IV. METHODOLOGY

The proposed methodology for traffic flow prediction is designed to accurately model complex traffic patterns by integrating machine learning and deep learning techniques. The system follows a structured pipeline that includes data collection, preprocessing, model development, training, and evaluation. This approach ensures efficient handling of large-scale traffic data and improves prediction accuracy.

Initially, traffic data is collected from multiple sources, including historical traffic datasets, real-time sensor data, and external factors such as weather conditions and road status. The integration of diverse data sources enhances the model's ability to capture real-world traffic variations and improves prediction reliability.

In the preprocessing stage, the collected data is cleaned and transformed to ensure consistency and quality. Missing values are handled using appropriate techniques such as interpolation or mean substitution, while noise and outliers are removed to prevent model distortion. The data is then normalized to bring all features within a similar scale, which improves the efficiency of model training. Feature selection is performed to identify the most relevant parameters, such as traffic volume, speed, time of day, and weather conditions, which significantly influence traffic flow.

Following preprocessing, multiple predictive models are developed using both machine learning and deep learning approaches. Machine learning models such as Random Forest (RF), K-Nearest Neighbors (KNN), and Support Vector Machines (SVM) are used to capture feature-based relationships in the data. Deep learning models, particularly Long Short-Term Memory (LSTM) networks, are employed to capture temporal dependencies in time-series traffic data. Additionally, Convolutional Neural Networks (CNN) are used to extract spatial features from traffic patterns. A hybrid model combining Random Forest and LSTM is proposed to leverage the strengths of both approaches, resulting in improved prediction accuracy and robustness. The proposed model can be implemented using Python with libraries such as TensorFlow and Scikit-learn on publicly available traffic datasets.

The model training process involves splitting the dataset into training and testing sets, typically in a 70:30 ratio. The training data is used to train the models, while the testing data is used to evaluate their performance. During training, hyperparameters are tuned to optimize model performance and avoid overfitting. Cross-validation techniques are applied to ensure model generalization and reliability.

To evaluate the effectiveness of the proposed models, several performance metrics are used, including Mean Squared Error (MSE), Root Mean Square Error (RMSE), and prediction accuracy. These metrics provide a quantitative measure of how closely the predicted values match the actual traffic flow data. Lower error values indicate better model performance.

The overall methodology integrates data-driven techniques with advanced learning models to provide an efficient and scalable solution for traffic flow prediction. The combination of machine learning, deep learning, and hybrid modeling enables the system to handle complex traffic patterns and deliver accurate real-time predictions, making it suitable for intelligent transportation systems and smart city applications.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, April 2026
Impact Factor: 8.2

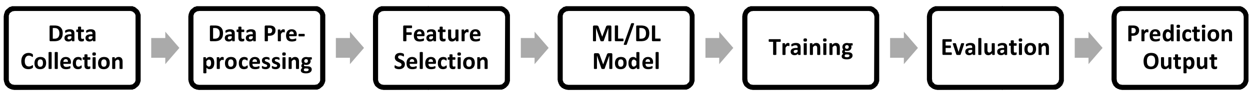

Fig. 1. Methodology Flow Diagram

V. SYSTEM ARCHITECTURE

The proposed traffic prediction system architecture consists of multiple interconnected components designed to efficiently process and analyze traffic data. The system begins with the data input layer, which collects data from various sources such as historical traffic datasets, real-time sensors, and external factors including weather conditions and road status. This data is then passed to the preprocessing module, where it undergoes cleaning, normalization, and transformation to ensure consistency and accuracy.

After preprocessing, the refined data is fed into the prediction engine, which utilizes machine learning and deep learning models such as Random Forest, Support Vector Machines, and Long Short-Term Memory networks to analyze patterns and generate accurate traffic flow predictions. A hybrid approach combining multiple models is also used to enhance prediction performance and reliability.

Finally, the predicted results are presented through the output visualization module, where the data is displayed in the form of graphs, charts, or dashboards. This allows users and traffic authorities to easily interpret the results and make informed decisions for traffic management and route optimization.

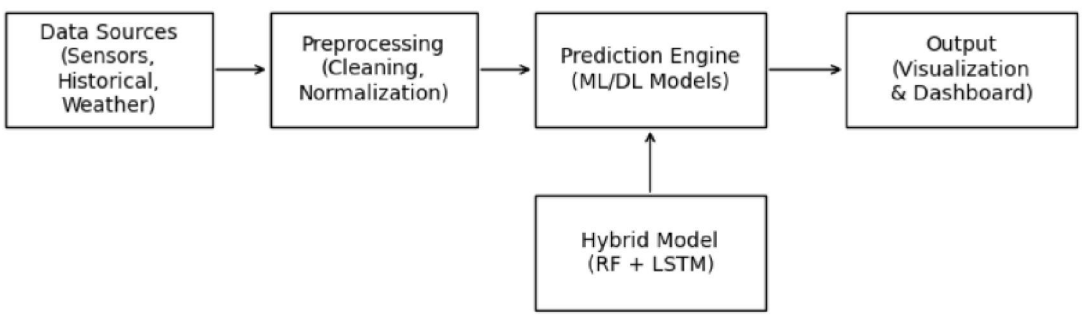

Fig. 2. System Architecture of Traffic Prediction Model

VI. RESULTS AND DISCUSSION

The performance of the proposed traffic flow prediction system was evaluated using various machine learning and deep learning models on a preprocessed traffic dataset. The dataset was divided into training and testing sets in a 70:30 ratio to ensure proper model validation. The models were assessed using standard evaluation metrics such as Mean Squared Error (MSE), Root Mean Square Error (RMSE), and prediction accuracy.

The experimental results indicate that traditional machine learning models such as K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) provide moderate accuracy but fail to capture complex temporal dependencies in traffic data. Random Forest (RF) performs better due to its ensemble learning capability; however, it still lacks the ability to model sequential patterns effectively.

Deep learning models, particularly Long Short-Term Memory (LSTM), demonstrate significantly improved performance due to their ability to capture temporal dependencies in time-series data. LSTM models effectively learn traffic patterns over time and provide more accurate predictions compared to traditional approaches. Convolutional Neural Networks (CNN) further enhance prediction by capturing spatial features when traffic data is represented in grid format.

The proposed hybrid model, which combines Random Forest and LSTM, shows the best overall performance among all models. By leveraging the strengths of both machine learning and deep learning techniques, the hybrid model achieves higher prediction accuracy and lower error rates. It effectively handles both feature-based relationships and temporal dependencies, making it suitable for real-world traffic prediction scenarios.

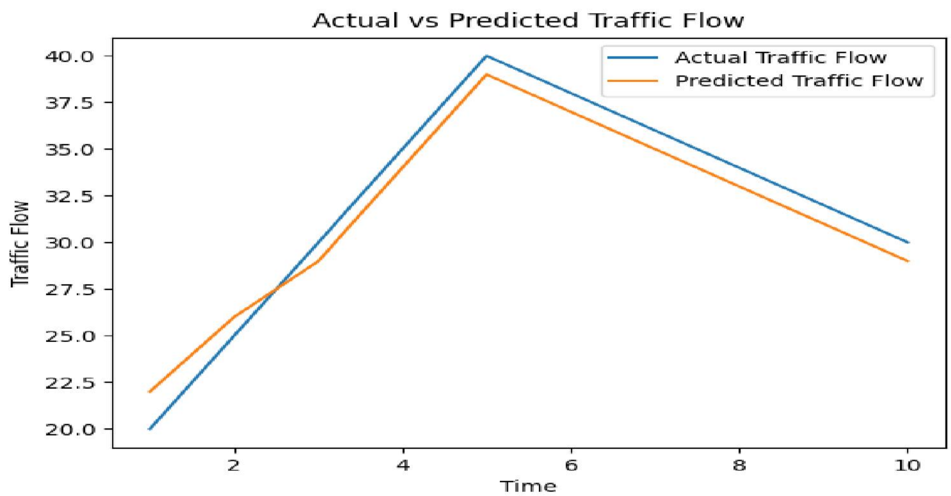

Fig. 3. Actual vs Predicted Traffic Flow

Graph Explanation

The graph illustrates the comparison between actual traffic flow values and predicted values generated by the proposed model. It can be observed that the predicted values closely follow the actual traffic trends, indicating high prediction accuracy. Minor deviations occur during peak traffic hours due to sudden fluctuations in traffic conditions; however, the overall prediction performance remains consistent and reliable.

TABLE II: PERFORMANCE COMPARISON

Model	Accuracy	RMSE	Performance
KNN	Medium	High	Low
SVM	Medium	Moderate	Moderate
RF	High	Moderate	Good
LSTM	Very High	Low	Excellent
Hybrid (RF + LSTM)	Highest	Lowest	Best

Discussion

From the results, it is evident that deep learning models outperform traditional machine learning approaches in traffic prediction tasks. The ability of LSTM to capture long-term dependencies makes it highly suitable for time-series forecasting. The hybrid model further enhances performance by combining the strengths of different algorithms.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, April 2026
Impact Factor: 8.2

The results also highlight the importance of proper data preprocessing and feature selection in improving model accuracy. The integration of multiple data sources, including real-time and historical data, contributes to better prediction performance.

Overall, the proposed system demonstrates strong potential for real-world implementation in Intelligent Transportation Systems (ITS), where accurate traffic prediction can significantly improve traffic management, reduce congestion, and enhance user experience.

VII. ADVANTAGES

The proposed model offers several advantages in terms of accuracy, efficiency, and practical applicability. By integrating machine learning and deep learning techniques, the system achieves higher prediction accuracy compared to traditional approaches. In particular, LSTM-based models effectively capture temporal patterns, while hybrid approaches further enhance overall performance.

The model is capable of handling complex and dynamic traffic data, making it suitable for real-world urban environments. The use of multiple data sources, including real-time inputs, improves the reliability of predictions and supports better decision-making for traffic management.

Another key advantage is scalability, as the system can be adapted to different cities and varying traffic conditions. Additionally, it contributes to environmental sustainability by reducing congestion, fuel consumption, and overall transportation inefficiencies.

VIII. LIMITATIONS

Despite its advantages, the proposed system has several limitations. One major challenge is the requirement for large volumes of high-quality data, as inaccurate or incomplete data can significantly affect model performance. In addition, deep learning models such as LSTM and hybrid approaches require substantial computational resources and longer training times.

The system may also struggle to handle unexpected traffic events, such as accidents or road blockages, since these situations are not always represented in historical data. Furthermore, the need for periodic retraining to adapt to changing traffic patterns increases system complexity.

Another limitation is the integration of multiple data sources, which introduces challenges related to data synchronization and consistency. Moreover, real-time processing of large-scale traffic data remains a significant challenge that requires further improvement.

IX. FUTURE SCOPE

The future of traffic flow prediction lies in the integration of advanced technologies and intelligent systems. One key direction is the use of Internet of Things (IoT) devices and smart sensors, which enable real-time data collection and improve prediction accuracy.

Another promising area is the adoption of Automated Machine Learning (AutoML), which can automate tasks such as model selection, hyperparameter tuning, and feature engineering, thereby reducing manual effort and improving efficiency.

In addition, advanced deep learning models such as Graph Neural Networks (GNN) and spatio-temporal architectures can be utilized to better capture complex relationships within traffic networks, leading to more accurate predictions.

Future systems may also incorporate smart traffic management solutions, including adaptive traffic signals and real-time route optimization. Integration with autonomous vehicles and vehicle-to-infrastructure (V2I) communication can further enhance system performance.

Finally, improving model robustness through anomaly detection techniques will help handle unexpected events such as accidents or road blockages, making traffic prediction systems more reliable in real-world scenarios.

X. CONCLUSION

This research presents a comprehensive analysis of traffic flow prediction using machine learning and deep learning techniques. The study demonstrates that traditional statistical models are limited in handling the complex, nonlinear, and dynamic nature of traffic data. In contrast, modern approaches, particularly LSTM-based models, provide better performance by effectively capturing temporal dependencies and improving prediction accuracy.

The results also highlight the effectiveness of hybrid models, which combine the strengths of different techniques to achieve more accurate and reliable predictions. Additionally, the importance of data preprocessing, feature selection, and the use of diverse data sources is emphasized, as these factors significantly influence model performance.

Overall, the findings indicate that hybrid models, especially the combination of LSTM and Random Forest, offer strong potential for real-time traffic prediction in urban environments. Future advancements in technologies such as IoT, AutoML, and cloud computing are expected to further enhance the efficiency and scalability of intelligent transportation systems.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to **Prof. Jayashree Pasalkar**, Department of Information Technology, AISSMS Institute of Information Technology, Pune, for her continuous guidance, valuable suggestions, and support throughout the development of this research work titled *"Traffic Flow Prediction Using Machine Learning and Deep Learning Techniques."* The authors also thank the faculty members of the Department of Information Technology for their encouragement and technical assistance during the course of this project. Their insights greatly contributed to the successful completion of this study. The authors are grateful to AISSMS Institute of Information Technology, Pune, for providing the necessary infrastructure and academic environment required to carry out this research work successfully. Finally, the authors acknowledge the contributions of researchers and scholars whose work has significantly supported this study.

REFERENCES

[1] N. A. M. Razali, N. Shamsaimon, K. K. Ishak, S. Ramli, M. F. M. Amran, and S. Sukardi, "Gap, techniques and evaluation: traffic flow prediction using machine learning and deep learning," *Journal of Big Data*, vol. 8, no. 152, 2021, doi: 10.1186/s40537-021-00542-7.

[2] A. A. Kashyap, S. Raviraj, A. Devarakonda, S. R. Nayak, S. K. V., and S. J. Bhat, "Traffic flow prediction models – A review of deep learning techniques," *Cogent Engineering*, vol. 9, no. 1, 2021, doi: 10.1080/23311916.2021.2010510.

[3] A. Navarro-Espinoza *et al.*, "Traffic Flow Prediction for Smart Traffic Lights Using Machine Learning Algorithms," *Technologies*, vol. 10, no. 1, 2022, doi: 10.3390/technologies10010005.

[4] R. Abduljabbar, H. Dia, and S. Liyanage, "Machine Learning Traffic Flow Prediction Models for Smart and Sustainable Traffic Management," *Infrastructures*, vol. 10, no. 7, p. 155, 2025, doi: 10.3390/infrastructures10070155.

[5] M. Khatiriolyaee, L. Yang, and R. W. Pazzi, "A comprehensive review of traffic prediction: From traditional machine learning to AutoML," *Neurocomputing*, vol. 666, 2026.

[6] B. Yu, H. Yin, and Z. Zhu, "Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting," *arXiv preprint arXiv:1709.04875*, 2017.

[7] P. Poonia and V. K. Jain, "Short-Term Traffic Flow Prediction Using LSTM," in *Proc. International Conference on Emerging Trends in Communication, Control and Computing (ICONC3)*, 2020.

[8] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Wang, "Traffic Flow Prediction With Big Data: A Deep Learning Approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865–873, Apr. 2015.

[9] J. Zhang, Y. Zheng, and D. Qi, "Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction," in *Proc. AAAI Conference on Artificial Intelligence*, 2017.

[10] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM Network: A Deep Learning Approach for Short-Term Traffic Forecast," *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017.