OPEN ACCESS
International Journal of Innovations in Science & Technology

Predictive Analytics for Smart Cities: Traffic Flow Forecasting Using Ensemble Algorithms

Ahmad Mustafa[1], Khurram Shehzad Khattak[1] , Zawar Hussain Khan[2]
[1] Department. of Computer Systems Engineering, UET Peshawar, Peshawar, Pakistan
[2] College of Computer Science and Engineering, University of Hail, Hail, Saudi Arabia
*Correspondence: cadet.mustafa2011@gamil.com, khurram.s.khattak@gmail.com, z.hussain@uoh.edu.sa

Citation | Mustafa A. Khattak. K. S Khan. Z. H, "Predictive Analytics for Smart Cities: Traffic Flow Forecasting Using Ensemble Algorithms", IJIST, Vol. 07 Special Issue. pp 257-267, May 2025

Received | April 10, 2025 **Revised |** May 12, 2025 **Accepted |** May 14, 2025 **Published |** May 16, 2025.

Traffic flow prediction is crucial for smart transportation systems, as it plays a key role in improving traffic management and planning infrastructure. While many machine learning techniques have been used for this purpose, ensemble methods have proven to be especially effective because they enhance prediction accuracy by combining the strengths of multiple models. This paper offers a detailed overview of how ensemble methods are applied to traffic flow prediction. We start by exploring the basics of traffic flow prediction, including common data sources, types, and performance metrics. Then, we categorize ensemble methods into bagging, boosting, and hybrid approaches, reviewing important studies that show how these methods work, the datasets they use, and their performance results. Real-world examples and case studies are included to highlight the practical effectiveness of these methods in various traffic situations. Finally, we discuss the current challenges and suggest future research directions, aiming to provide a valuable resource for researchers and practitioners interested in improving traffic flow prediction with ensemble techniques.

Keywords: Traffic Flow Prediction, Ensemble, Machine learning, Hybrid Models

Introduction:

Traffic flow prediction (TFP) is a crucial part of modern intelligent transportation systems (ITS), which are vital for improving traffic management, urban planning, and navigation. Accurate predictions help traffic managers optimize signals, inform drivers of delays, and plan infrastructure improvements, reducing congestion and enhancing road safety. The task involves estimating how many cars will pass through a specific road segment within a set time frame, considering factors like the time of day, weather, accidents, and special events [1]. This makes the task challenging due to the complex, dynamic, and often non-linear patterns of traffic flow. Traffic flow prediction is usually divided into two types: Short-Term Prediction (forecasting traffic flow minutes to hours ahead, essential for real-time management) and Long-Term Prediction (predicting traffic flow days, weeks, or even months in advance for planning infrastructure and policies) [2].

Research Context: Accurate traffic flow forecasting requires high-quality data from various technologies, such as fixed sensors, cameras, GPS devices, crowdsourced data, social media, and weather information [3][4]. These diverse sources offer a comprehensive view of traffic conditions and are typically used in two forms. The first is time series data, which consists of data points showing traffic flow at regular intervals. The second is spatial-temporal data, which combines time series with spatial information, capturing traffic flow across different locations over time [5].

Traditional prediction methods, like statistical models, often struggle to capture the complex, non-linear patterns in traffic data. This has led to the growing use of machine learning, known for its strong modeling capabilities. Among these, ensemble methods are particularly effective because they combine the strengths of several models to improve accuracy. These methods overcome the weaknesses of individual models by using different algorithms, resulting in more reliable and precise traffic flow predictions. They are especially suited for handling large and complex traffic datasets. However, there is a lack of comprehensive surveys focusing on ensemble methods in traffic flow prediction, which is why these paper aims to review and analyze recent research in this field.

Scope and Objective: This paper contributes to the field of TFP by exploring ensemble methods. It provides an in-depth analysis of these techniques and categorizes them into three main types: bagging, boosting, and hybrid approaches. The survey serves as a foundation for understanding the current methods used in the literature. The study also compares different ensemble approaches, highlighting their advantages and disadvantages to help identify the most suitable methods for specific scenarios and datasets. This provides valuable insights for both researchers and professionals.

The paper includes case studies and real-world applications, showing how ensemble methods can be applied to solve TFP problems. It demonstrates how theoretical advances can lead to practical solutions in real-world scenarios. To encourage further research and innovation in this area, the paper also discusses current challenges and proposes future research directions.

The structure of the remaining document is as follows: Section 2 provides an overview of ensemble methods, Section 3 reviews their applications in TFP, Section 4 discusses the challenges and prospects, and Section 5 presents the key findings and recommendations.

Ensemble Methods in Traffic Flow Prediction:

The flowchart in Figure. 1 outlines the logical structure of the survey paper. It begins with defining the objectives and scope, followed by a comprehensive literature review. The reviewed studies are then categorized into key methodological areas, including ensemble approaches, hybrid models, LSTM-based methods, data sources, spatiotemporal features, decomposition techniques, model comparisons, and real-world applications in urban planning.

Ensemble techniques are a powerful machine learning domain that combines multiple weak learners to form a single, strong, and reliable prediction model. The core idea is that by aggregating the predictions of several weak learners, the ensemble can achieve better performance and generalization than any individual model. This approach takes advantage of the diversity among models to reduce errors and improve accuracy. Ensemble methods are typically categorized into three main types: bagging, boosting, and hybrid approaches.

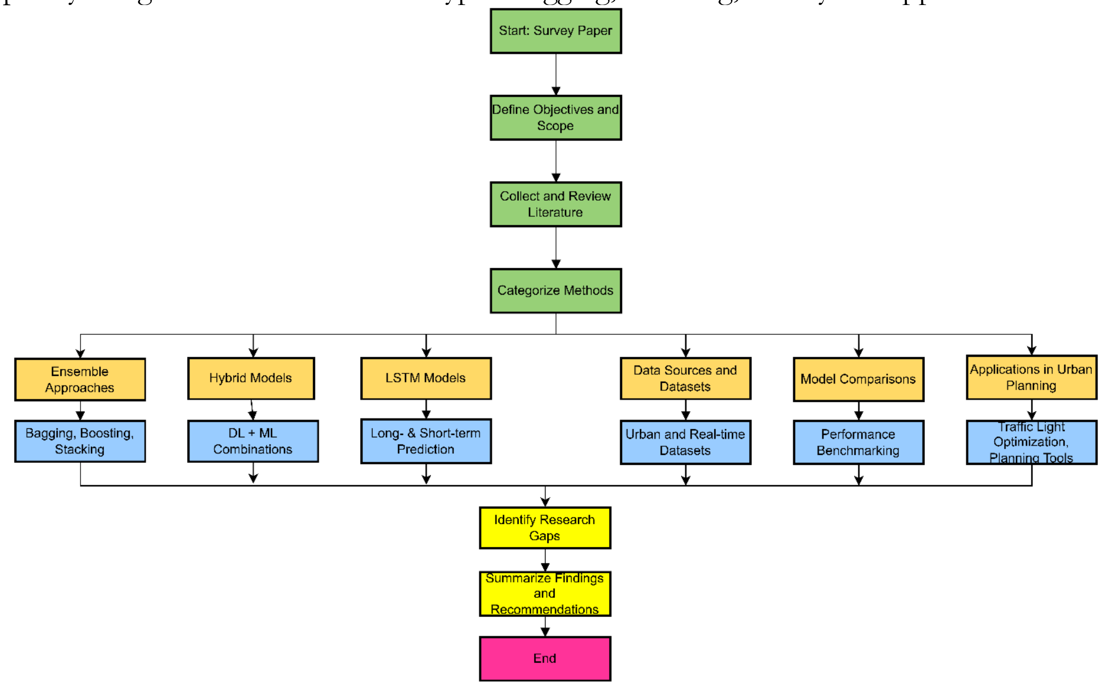

Figure 1: Methodology process used to identify papers reviewed

Bagging (Bootstrap Aggregating):

The core concept of bagging involves training multiple instances of the same model on different subsets of the training data and then averaging their predictions (for regression) or using a majority vote (for classification). A prominent example of this method is Random Forest (RF), which is an ensemble of Decision Trees (DT). In RF, each tree is trained on a bootstrapped sample of the data, and at each node, a random selection of features is considered for splitting.

Boosting:

Boosting train models sequentially, where each new model attempts to correct the errors made by the previous ones. The final prediction is a weighted sum of the predictions from all models. Key boosting techniques include: Gradient Boosting Machines (GBM) are a class of models that optimize a loss function by building models sequentially. Extreme Gradient Boosting (XGBoost) is an enhanced version of GBM that introduces regularization to prevent overfitting, making it more robust. LightGBM offers a faster implementation of GBM by employing a histogram-based approach for training, which improves efficiency and scalability. CatBoost is another gradient boosting library specifically designed to automatically handle categorical features, simplifying the preprocessing steps and enhancing performance.

Hybrid and Advanced Ensemble Methods:

Hybrid and advanced ensemble methods combine various ensemble and machine learning techniques to enhance model performance. Key approaches include stacking and blending, where multiple base models are combined using a meta-model. Stacking typically employs cross-validation to train the meta-model, while blending often relies on holdout validation sets. Deep learning ensembles merge multiple neural networks to capitalize on their strengths and are especially effective in capturing complex spatial and temporal dependencies, such as those found in traffic flow data. Additionally, hybrid models integrate different types of algorithms by combining classical machine learning techniques with deep learning models to further optimize predictive performance.

Review of Recent Studies:

This comprehensive review synthesizes recent research on TFP using ensemble methods, focusing on machine learning and deep learning models applied to diverse datasets and methods. Ensemble techniques have proven to significantly enhance prediction accuracy by combining the strengths of different base models.

Ensemble Approaches:

Ensemble approaches, particularly bagging and boosting techniques, have been extensively utilized in TFP due to their robustness and enhanced predictive capabilities. Methods such as RF and Gradient Boosting Machines (GBM) show the effectiveness of these strategies. For example, the application of combined with boosting [1] has shown superior performance compared to traditional models, including Long Short-Term Memory (LSTM) networks and linear regression.

Hybrid models that integrate deep learning techniques, such as LSTM, Convolutional Neural Networks (CNN), and Gated Recurrent Units (GRU), with traditional machine learning algorithms have gained considerable attention in the field of traffic flow prediction. For instance, studies conducted by Cini et al. [2] have shown that deep ensemble models, which combine multiple deep learning architectures, significantly outperform individual models. This suggests that hybrid approaches offer a promising direction for capturing the complex spatial and temporal dynamics inherent in traffic data.

LSTM Models:

LSTM models have been widely adopted in TFP due to their ability to capture long-term temporal dependencies in sequential data. Studies like [6] have demonstrated that combining LSTM with other models, such as Deep Belief Networks (DBN), within an ensemble framework can significantly enhance predictive performance. This integration allows the model to leverage the strengths of both deep learning architectures, resulting in more accurate and reliable traffic flow forecasts.

For short-term traffic flow prediction, LSTM models combined with boosting methods have demonstrated superior performance compared to traditional baseline approaches such as the Autoregressive Integrated Moving Average (ARIMA) model. This improvement underscores the effectiveness of sequential model building, where the LSTM captures temporal dependencies while boosting enhances the model's generalization and accuracy.

Data Sources and Datasets:

The reviewed studies utilized a wide range of datasets, encompassing both historical and real-time traffic data collected from diverse urban environments such as New York and Sydney. Prominent sources include standardized and widely used datasets like the Caltrans Performance Measurement System (PeMS), which provide high-resolution traffic flow information. The geographic and temporal diversity of these datasets underscores the adaptability and robustness of ensemble methods across varying traffic conditions and infrastructural contexts.

Feature Engineering and Preprocessing:

Data decomposition techniques, such as Ensemble Empirical Mode Decomposition (EEMD), have been employed to address the challenges posed by noisy traffic data. EEMD effectively decomposes complex traffic signals into simpler, more interpretable components, allowing for improved feature extraction and noise reduction. Following decomposition, base learners such as XGBoost or LSTM networks are utilized to perform traffic flow predictions. Studies [7], [8] have demonstrated that this two-step approach enhances predictive accuracy by enabling models to focus on the underlying patterns in the data rather than being influenced by noise and irregularities.

Several studies have emphasized the incorporation of spatiotemporal features to enhance the accuracy of TFP models. These features capture spatial dependencies across different lanes, road segments, or regions, as well as temporal dependencies over time. By modeling both dimensions, spatiotemporal approaches provide a more comprehensive understanding of traffic dynamics, which is essential for accurately predicting fluctuations in traffic flow, particularly in complex and rapidly changing urban environments [9], [10].

Comparison of Ensemble Models:

Ensemble models, including stacking, blending, and bagging, have been systematically compared with traditional machine learning algorithms, such as linear regression, decision trees, and k-Nearest Neighbors, as well as with standalone deep learning techniques. Stacking ensembles, which integrate predictions from multiple base models using a meta-learner, have shown marked improvements in accuracy and reductions in error rates. Studies [11], [12] highlight that such ensemble configurations outperform individual models by effectively leveraging the complementary strengths of diverse learning algorithms.

Several studies have combined gradient boosting frameworks such as LightGBM, XGBoost, and CatBoost to improve predictive accuracy in traffic flow forecasting. These models are recognized for their high computational efficiency and robust performance, particularly in handling large-scale and high-dimensional data. When aggregated within an ensemble framework, these algorithms complement each other's strengths, resulting in enhanced model performance. Empirical evidence from studies [5][13] demonstrates that such ensembles consistently outperform individual models, achieving greater accuracy and reliability in traffic prediction tasks.

TFPin Urban Planning:

TFP extends beyond the goal of accurate forecasting; it plays a pivotal role in improving urban infrastructure and enhancing the efficiency of transportation systems. For instance, studies such as [13] have applied TFP models to optimize traffic light systems, effectively reducing congestion. By adjusting traffic signal timings in real time based on predicted traffic conditions, these models enable more dynamic and responsive traffic management. This approach not only improves traffic flow but also contributes to overall urban mobility, reducing delays and minimizing the environmental impact of traffic congestion.

Multistep traffic forecasting models have been developed in several studies to predict traffic flow over various time horizons, offering significant improvements in long-term traffic management systems. By extending the forecasting window, these models provide valuable insights into future traffic conditions, enabling proactive planning and decision-making. The ability to predict traffic flow at multiple time steps enhances the accuracy of resource allocation, infrastructure management, and congestion control, contributing to more efficient urban transportation systems. Studies such as [14] and [15] highlight the effectiveness of these multistep approaches in improving the robustness and reliability of traffic prediction models for long-term applications.

The use of ensemble methods for TFP has proven to be highly effective in improving the accuracy and robustness of traffic forecasts. By leveraging the strengths of multiple base models, researchers have been able to address complex spatiotemporal dependencies in traffic data. Future work could focus on further optimizing ensemble techniques, integrating real-time traffic data, and refining hybrid models to cater to specific urban traffic management challenges. Additionally, improving model efficiency in terms of computation time and scalability remains an area for enhancement, especially for real-time applications in large cities.

Table 1: Performance of reported ML and Ensemble Models in existing literature.

Ref	ML Models	MAE	RMSE	R^2	MAPE%
[1]	RF	13.76	22.39	0.9341	-
	LSTM	14.74	23.50	0.9275	-
	LR	17.80	27.04	0.9040	-
	RF (Baging)	**13.69**	**22.21**	**0.9352**	**-**
[2]	LSTM	0.1656	-	0.9248	-
	GRU	0.1657	-	0.9213	-
	CNN	0.1675	-	0.9258	-
	Ens1	0.1590	-	0.9301	-
	Ens2	**0.1553**	-	**0.9366**	-
[6]	LR	-	8.190	0.860	-
	DT Regressor	-	6.981	0.898	-
	Sequential (DNN)	-	9.903	0.795	
	Model Stacked	**-**	**5.556**	**0.932**	
[16]	MLP-NN	10.8281	15.4202	0.9304	21.16
	RF	10.8827	15.5481	0.9296	21.84
	GRU	10.8843	15.6191	0.9278	22.85
	LSTM	10.8806	15.6771	0.9267	22.32
	LR	11.2010	15.8545	0.9263	24.32
	Stochastic Gradient	12.8230	18.3727	0.9003	29.01
	GB	**10.8508**	**15.4121**	**0.9305**	**21.99**
[9]	**Proposed**	**3.27**	**1.76**	**-**	**4.24**
[7]	Seasonal ARIMA		93.74	-	13.08
	LSTM		64.79	-	10.96
	E-ARIMA (uniform weight)		60.12	-	10.51
	E-ARIMA (distance-based weights)		**60.11**	**-**	**10.50**
[17]	Single model approach				
	SVR	13.1604	17.7800	-	-
	LSTM	13.3999	18.0299	-	-
	HA	13.9525	19.0658	-	-
	CNN	21.9240	30.940	-	-
	Ensemble model approach				
	WRegression	13.1788	17.7886	-	-
	GBRT	13.2907	17.8656	-	-
	TCAE (LSTM, HA)	13.0387	17.4541	-	-
	TCAE (HA, SVR)	12.9583	17.3950	-	-
	TCAE (SVR, LSTM)	12.9443	17.3748	-	-
	TCAE SVR, LSTM, HA)	**12.9113**	**17.3086**	**-**	**-**
[12]	ARIMA	-	141.5	-	13.35
	BP	-	136.48	-	12.49
	SVM	-	136.1	-	11.94
	DBN	-	135.44	-	11.48
	LSTM	-	135.24	-	10.78
	EEMD-BP	-	135.98	-	11.97
	EEMD-LSTEM	-	132.31	-	10.21
	EEMD-DBN	-	133.45	-	10.86
	EEMD-mRMR-DBN	**-**	**131.58**	**-**	**10.19**

[18]	RF	17.38	-	-	-
	XGBoost	17.27	-	-	-
	GBDT	17.25	-	-	-
	KNN	18.63	-	-	-
	DT	18.45	-	-	-
	GRU	18.54	-	-	-
	Stacking	17.07	-	-	-
	Ba-Stacking	16.99	-	-	-
	DW-Ba-Stacking	**16.87**	**-**	**-**	**-**
[10]	GCN	27.66	41.04	-	22.24
	LSTM	26.82	41.23	-	19.40
	DCRNN	24.70	38.12	-	17.12
	STGCN	22.70	35.55	-	14.49
	ASTGCN	22.93	22.93	-	16.56
	STSGCN	21.19	33.65	-	**13.90**
	LightGGM	21.39	33.71	-	14.92
	Proposed	**20.55**	**32.66**	**-**	**14.21**
[14]	ARIMA	3.041	4.141	-	16.506
	ENN	2.813	3.845	-	15.469
	SVM	3.128	4.413	-	16.480
	ESN	6.352	9.810	-	32.366
	Proposed	**2.730**	**3.752**	-	**15.136**
[19]	SARIMA	-	366.93	-	10.55
	Kalman Filter	-	269.99	-	10.36
	RF	-	266.81	-	7.79
	CWGB-HR	-	**141.36**	-	**5.52**
[13]	LR	318.717	511.908	511.90	-
	DT	63.209	134.135	0.971	-
	GB	63.203	125.709	0.975	-
	KNN	71.713	147.039	0.966	-
	RF	**54.826**	**111.844**	**0.980**	**-**
[20]	ARIMA	4.21	5.78	**-**	21.34
	MLP	4.27	5.84	**-**	23.06
	BPNN	4.22	5.78	**-**	22.27
	LSTM	4.24	5.81	**-**	21.30
	GRU	4.27	5.83	**-**	21.45
	XGBOOST	4.26	3.82	-	22.77
	Wavelet-XGBoost	2.50	3.85	-	14.66
	EMD-XGBoost	2.56	3.54	-	14.37
	EEMD-XGBoost	2.04	2.86	-	11.43
	CEEMDAN-XGBoost	**1.79**	**2.54**	-	**9.88**
[21]	LSTM	**-**	**-**	-	9.88
	EnLSSVR	**-**	**-**	-	8.44
	E-ELM	**-**	**-**	-	7.79
	MLP	-	-	-	7.69
	NCAE-ELM-EA	-	-	-	**7.58**
[22]	MLP	0.75	1.344	-	0.6030
	RF	0.81	1.37	-	0.59

	KNN	0.51	1.04	-	0.382
[8]	LSTM	16.13	21.53	8.31	-
	LSTM+WL (haar)	5.22	9.12	3.35	-
	LSTM + WL (db)	2.36	3.75	1.37	-
	LSTM + WL (sym)	2.95	4.85	1.58	-
	LSTM + WL (coif)	2.64	4.03	1.31	-
	LSTM + EMD	2.23	2.88	1.46	-
	LSTM + EEMD	**1.21**	**1.58**	**0.91**	**-**

Comparative Analysis:

The analysis reveals several critical insights regarding traffic forecasting and model performance. Firstly, long-term traffic forecasting generally exhibits lower accuracy compared to short-term forecasting, primarily due to the heightened complexity and variability inherent in extended prediction horizons. Additionally, standard K-fold cross-validation proves inadequate for time-series data, as it risks data leakage by incorporating future data into the training set. Instead, a more appropriate validation strategy is the blending method, which preserves temporal integrity by reserving the last 20% of the data for validation. In terms of predictive accuracy, boosting methods such as XGBoost and LightGBM consistently outperform other models, demonstrating their efficacy in capturing complex data relationships, particularly in TFP tasks. Hybrid models and deep learning ensembles also exhibit superior performance by effectively identifying intricate patterns in the data. From a computational standpoint, LightGBM emerges as the most efficient algorithm, significantly reducing training time compared to other boosting techniques, while bagging methods like RF remain relatively effective for large datasets despite being less efficient. Furthermore, ensemble methods overall demonstrate greater robustness and resilience to overfitting, enabling consistent performance across varying data conditions. Notably, boosting algorithms such as XGBoost and CatBoost maintain strong predictive capabilities across diverse datasets, reinforcing their widespread adoption in forecasting applications.

The evaluation of different ensemble methods reveals distinct strengths and limitations across various approaches. Bagging methods, such as RF, are recognized for their robustness and ease of implementation; however, they can become computationally expensive, particularly for large datasets, due to high memory demands. In contrast, boosting methods like XGBoost and LightGBM demonstrate superior accuracy, excelling in modeling non-linear and complex relationships. Nevertheless, their performance is highly sensitive to hyperparameter configurations, necessitating meticulous tuning to prevent overfitting and ensure optimal results. Hybrid and advanced ensemble techniques, which integrate multiple modeling approaches, offer enhanced accuracy and robustness by leveraging the complementary strengths of different algorithms. However, these methods often introduce greater implementation complexity and demand significant computational resources, especially when incorporating deep learning-based ensembles.

In summary, boosting methods such as XGBoost and LightGBM are highly effective for traffic flow prediction, offering a strong balance between accuracy and computational efficiency. However, their performance heavily depends on proper hyperparameter tuning to ensure optimal results. Hybrid models and deep learning-based ensembles further enhance predictive accuracy and robustness, making them particularly suitable for complex traffic forecasting tasks, albeit at the cost of increased model complexity and higher computational demands. While simpler methods may be adequate for short-term predictions, long-term forecasting necessitates more sophisticated ensemble approaches to effectively capture the inherent complexities and variability in the data. These findings underscore the importance of selecting appropriate modeling techniques based on prediction horizon, data characteristics, and available computational resources.

Conclusion:

Boosting methods, particularly XGBoost, consistently demonstrate superior performance due to several key characteristics that enhance model accuracy and generalization. First, boosting constructs models in a sequential manner, where each new model is trained to correct the errors made by its predecessor. This iterative error-correcting process enables the ensemble to progressively improve. Second, boosting assigns greater weight to data samples that are difficult to predict, thereby directing more learning capacity toward complex or misclassified instances. This targeted focus enhances the model's overall predictive capability. Third, unlike bagging methods, which primarily reduce variance, boosting effectively reduces both bias and variance, leading to improved generalization across diverse datasets. Furthermore, XGBoost incorporates built-in L1 (Lasso) and L2 (Ridge) regularization techniques, which help mitigate overfitting and ensure robust performance on unseen data. Lastly, XGBoost employs advanced optimization strategies, including the use of second-order gradients and shrinkage (learning rate reduction), which contribute to faster convergence and higher model accuracy.

Ensemble methods, including bagging, boosting, and hybrid approaches, have demonstrated significant potential in improving TFP accuracy while minimizing errors compared to standalone models. However, several challenges remain that hinder their widespread adoption. Data quality and accessibility issues continue to pose limitations, as the performance of these models heavily depends on the availability of reliable and comprehensive traffic datasets. Additionally, while ensemble models often achieve high predictive accuracy, their inherent complexity reduces interpretability, creating barriers for real-world implementation where model transparency is crucial. Furthermore, the computational demands of advanced ensemble techniques, particularly deep learning-based approaches, raise concerns regarding scalability and efficiency, especially when applied to large-scale traffic networks.

To address the existing challenges in traffic flow prediction, this paper identifies several promising avenues for future research. First, advancing data fusion methodologies could enhance predictive accuracy by integrating diverse data sources, thereby creating more comprehensive and representative training datasets. Additionally, there is a critical need to develop interpretable models that maintain high performance while offering transparency—a key requirement for real-world deployment in transportation systems. Another crucial area involves establishing stable training methodologies, particularly for handling the variability inherent in large and heterogeneous traffic datasets.

This study identifies several critical research directions to advance TFP capabilities. First, developing sophisticated data fusion methodologies could significantly enhance prediction accuracy by intelligently integrating heterogeneous data sources, including IoT sensors, GPS trajectories, and traffic camera feeds, to construct more comprehensive datasets. Secondly, optimizing scalable algorithms remains essential to efficiently process ever-growing urban mobility datasets while meeting the low-latency requirements of real-time applications.

A particularly pressing research challenge involves improving the stability of ensemble models to guarantee consistent performance across diverse traffic conditions, geographical regions, and temporal scales. Future investigations should focus on developing adaptive ensemble frameworks that maintain reliability despite data distribution shifts, while simultaneously addressing computational efficiency concerns. These advancements would significantly strengthen the practical utility of traffic prediction systems in smart city infrastructures.

The findings of this study carry significant implications for advancing traffic prediction methodologies within ITS. Addressing the identified improvement areas would enable researchers and practitioners to develop more precise, reliable, and computationally efficient TFP models. Such advancements would directly contribute to optimizing traffic management strategies, reducing congestion, and improving overall transportation system performance.

These research directions promise to yield traffic regulation systems that are not only more accurate but also better adapted to real-world operational conditions. The resulting improvements could transform ITS capabilities, enabling more responsive and adaptive traffic management solutions that account for the dynamic nature of urban transportation ecosystems. Ultimately, such progress would support the development of smarter, more efficient cities with improved mobility for all users.

References:

[1] M. R. Anas Saleh Alkarim, Abdullah S. Al-Malaise Al-Ghamdi, "Ensemble Learning-based Algorithms for Traffic Flow Prediction in Smart Traffic Systems," *Eng. Technol. Appl. Sci. Res.*, vol. 14, no. 2, pp. 13090–13094, 2024, doi: https://doi.org/10.48084/etasr.6767.

[2] Nevin Cini & Zafer Aydin, "A Deep Ensemble Approach for Long-Term Traffic Flow Prediction," *Arab. J. Sci. Eng.*, vol. 49, pp. 12377–12392, 2024, doi: https://doi.org/10.1007/s13369-023-08672-1.

[3] K. A. N. Ayaz Shehzad, Khattak Khurram S, Khan Zawar H, Minallah Nasru, Khan Mushtaq A, "Sensing technologies for traffic flow characterization: From heterogeneous traffic perspective," *J. Appl. Eng. Sci.*, vol. 20, no. 1, pp. 29–40, 2022, [Online]. Available: https://scindeks.ceon.rs/Article.aspx?artid=1451-41172201029A

[4] A. N. K. ALI ZEB, KHURRAM S. KHATTAK, AREEB AGHA, , ZAWAR H. KHAN, M. ATHAR JAVED SETH, "ON-BOARD DIAGNOSTIC (OBD-II) BASED CYBER PHYSICAL SYSTEM FOR ROAD BOTTLENECKS DETECTION," *J. Eng. Sci. Technol.*, vol. 17, no. 2, pp. 0906–0922, 2022, [Online]. Available: https://jestec.taylors.edu.my/Vol 17 Issue 2 April 2022/17_2_07.pdf

[5] R. Singh, G. Gaonkar, V. Bandre, N. Sarang, and S. Deshpande, "Gradient Boosting Approach for Traffic Flow Prediction Using CatBoost," *2021 7th IEEE Int. Conf. Adv. Comput. Commun. Control. ICAC3 2021*, 2021, doi: 10.1109/ICAC353642.2021.9697133.

[6] B. Santhosh Narayanan and L. Shyamala, "Adopting Ensemble Learning and Machine Learning Techniques for Predictive Modeling in Traffic Data Analysis," *Proc. 3rd Int. Conf. Appl. Artif. Intell. Comput. ICAAIC 2024*, pp. 528–533, 2024, doi: 10.1109/ICAAIC60222.2024.10575594.

[7] S. Shahriari, M. Ghasri, S. A. Sisson, and T. Rashidi, "Ensemble of ARIMA: combining parametric and bootstrapping technique for traffic flow prediction," *Transp. A Transp. Sci.*, vol. 16, no. 3, pp. 1552–1573, Jan. 2020, doi: 10.1080/23249935.2020.1764662;SUBPAGE:STRING:ACCESS.

[8] Y. X. Xinqiang Chen , Huixing Chen, Yongsheng Yang, Huafeng Wu, Wenhui Zhang, Jiansen Zhao, "Traffic flow prediction by an ensemble framework with data denoising and deep learning model," *Phys. A Stat. Mech. its Appl.*, vol. 565, p. 125574, 2021, doi: https://doi.org/10.1016/j.physa.2020.125574.

[9] G. Yang, Z. Chen, and X. Ma, "A Novel Attention Ensemble Traffic Flow Prediction Model based on Spatio-Temporal big data," *2022 2nd Int. Conf. Comput. Sci. Electron. Inf. Eng. Intell. Control Technol. CEI 2022*, pp. 451–454, 2022, doi: 10.1109/CEI57409.2022.9950121.

[10] D. Zhao and F. Chen, "A Hybrid Ensemble Model for Short-Term Traffic Flow Prediction," *Proc. - 2022 Chinese Autom. Congr. CAC 2022*, vol. 2022-January, pp. 3887–3891, 2022, doi: 10.1109/CAC57257.2022.10054817.

[11] X. Zeng, Y. Wang, X. Deng, and J. Wang, "Short-Term Traffic Flow Prediction Based on

Ensemble Machine Learning Strategies," *Proc. 2021 IEEE 10th Data Driven Control Learn. Syst. Conf. DDCLS 2021*, pp. 333–338, May 2021, doi: 10.1109/DDCLS52934.2021.9455594.
[12] J. Liu, N. Wu, Y. Qiao, and Z. Li, "Short-Term Traffic Flow Forecasting Using Ensemble Approach Based on Deep Belief Networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 1, pp. 404–417, Jan. 2022, doi: 10.1109/TITS.2020.3011700.
[13] N. R. Idriss Moumen, Jaafar Abouchabaka, "Adaptive traffic lights based on traffic flow prediction using machine learning models," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 5, p. 5813, 2023, [Online]. Available: https://ijece.iaescore.com/index.php/IJECE/article/view/31438
[14] S. Yin, H. Liu, Y. Li, J. Tan, and J. Wang, "Multi-step traffic flow prediction using stacking ensemble learning model," *Fifth Int. Conf. Traffic Eng. Transp. Syst. (ICTETS 2021)*, vol. 12058, pp. 638–643, Dec. 2021, doi: https://doi.org/10.1117/12.2620226.
[15] C. Dong, Z. Lv, and J. Li, "A Forecasting Method of Dual Traffic Condition Indicators Based on Ensemble Learning," *Proc. Int. Conf. Parallel Distrib. Syst. - ICPADS*, vol. 2021-December, pp. 332–339, 2021, doi: 10.1109/ICPADS53394.2021.00047.
[16] E. I.-G. Alfonso Navarro-Espinoza, Oscar Roberto López-Bonilla, Enrique Efrén García-Guerrero, Esteban Tlelo-Cuautle, Didier López-Mancilla, Carlos Hernández-Mejía, "Traffic Flow Prediction for Smart Traffic Lights Using Machine Learning Algorithms," *Technologies*, vol. 10, no. 1, p. 5, 2022, doi: https://doi.org/10.3390/technologies10010005.
[17] Y. Chen, H. Chen, P. Ye, Y. Lv, and F. Y. Wang, "Acting as a Decision Maker: Traffic-Condition-Aware Ensemble Learning for Traffic Flow Prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 4, pp. 3190–3200, Apr. 2022, doi: 10.1109/TITS.2020.3032758.
[18] Y. H. Zhaohui Li, Lin Wang, Deyao Wang, Ming Yin, "Short-Term Traffic-Flow Forecasting Based on an Integrated Model Combining Bagging and Stacking Considering Weight Coefficient," *Electronics*, vol. 11, no. 9, p. 1467, 2022, doi: https://doi.org/10.3390/electronics11091467.
[19] Z. Li, Z. Zheng, and S. Washington, "Short-Term Traffic Flow Forecasting: A Component-Wise Gradient Boosting Approach with Hierarchical Reconciliation," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 12, pp. 5060–5072, Dec. 2020, doi: 10.1109/TITS.2019.2948381.
[20] B. R. and Y. G. W. Lu, Y. Rui, Z. Yi, "A Hybrid Model for Lane-Level Traffic Flow Forecasting Based on Complete Ensemble Empirical Mode Decomposition and Extreme Gradient Boosting," *IEEE Access*, vol. 8, pp. 42042–42054, 2020, doi: 10.1109/ACCESS.2020.2977219.
[21] M. L. and Y. L. X. Qi, X. Qin, Z. Jia, "NCAE and ELM Based Enhanced Ensemble Optimized Model for Traffic Flow Forecasting," *IEEE Access*, vol. 8, pp. 200486–200499, 2020, doi: 10.1109/ACCESS.2020.3034763.
[22] N. A. Noor Ullah Khan, Munam Ali Shah, Carsten Maple, Ejaz Ahmed, "Traffic Flow Prediction: An Intelligent Scheme for Forecasting Traffic Flow Using Air Pollution Data in Smart Cities with Bagging Ensemble," *Sustainability*, vol. 14, no. 7, p. 4164, 2022, doi: https://doi.org/10.3390/su14074164.